
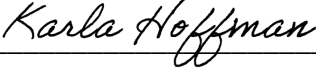


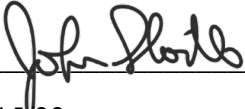


AN APPROXIMATE DYNAMIC PROGRAMMING APPROACH TO FUTURE
NAVY FLEET INVESTMENT ASSESSMENTS

by

Matthew J. Powers
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Systems Engineering and Operations Research

Committee:

	Dr. Rajesh Ganesan, Dissertation Director
	Dr. Karla Hoffman, Committee Member
	Dr. Andrew Loerch, Committee Member
	Dr. Girum Urgessa, Committee Member Dr.
	John Shortle, Department Chair
Date: 4/15/22	Spring Semester 2022 George Mason University Fairfax, VA

An Approximate Dynamic Programming Approach to Future Navy Fleet Investment
Assessments

A Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy at George Mason University

by

Matthew J. Powers
Master of Science
Naval Postgraduate School, 2012
Bachelor of Science
United States Naval Academy, 2001

Director: Rajesh Ganesan, Associate Professor
Department of Systems Engineering and Operations Research

Spring Semester 2022
George Mason University
Fairfax, VA

Copyright © 2022 by Matthew J. Powers.
All Rights Reserved

DEDICATION

I've written over 20,000 words, yet here I am at a loss for what to say. I dedicate this work to my beautiful wife, Shyla, who has loved (or at least, tolerated) me for over 18 years and has been my bride for over 16 years! Of all the challenges we've endured, this PhD is one of them! I also dedicate this work to my three wonderful children – Aidan, Ben, and Lucy. Thank you for sometimes leaving me in peace for the past few years, and for the other times reminding me to put down my notes and just be a dad! I love you all, and I will always owe my successes to you!

ACKNOWLEDGEMENTS

I would like to thank my family and friends whose support has made this journey possible. I did not take the decision to pursue my PhD lightly, nor would I describe this as an “easy” path. I relied on your continuous encouragement to keep me motivated! Thank you to all my George Mason University professors and committee. Dr. Rajesh Ganesan, thank you for your incredible advising efforts and guidance. I hope that I reflect your passion for Approximate Dynamic Programming (ADP), and that I can continue to apply what I have learned throughout my career. I had the honor of working with Dr. Karla Hoffman, Dr. Andrew Loerch, and Dr. Girum Urgessa, all of whom served on my committee. I am truly honored that I have documented proof that I was in the presence of your greatness! Extra thanks to Dr. Loerch for starting and guiding me along this adventure, both on campus and at various Military Operations Research Society (MORS) events. Thank you to Mr. Harrison Schramm and Mr. Brian Morgan for your valuable and ongoing mentorship. Thank you to Ms. Jennifer Ferat and the wonderful MORS staff and fellow members who surround me with OR enthusiasm. Thank you to my colleagues at OPNAV N81 for your professional support during this study, especially to CDR April Bakken for your special brand of encouragement, and to CDR Dan Ciullo for trusting me with this important analysis. Thank you Dr. Lisa Oakley-Bogdewic, Dr. Les Servi, and my MITRE colleagues for your ongoing support of my professional development. Special thanks to Lisa, Harrison, and to Dr. Paul Nicholas for your support and guidance as I began and continued along my academic path. Mary Poirier, thank you for your reliable help and always having an answer for me when I was feeling overwhelmed by the administrative tasks required to graduate.

Once again, I thank my wife, Shyla, and my children, Aidan, Benjamin, and Lucy. You are all the reasons for any success I enjoy. Thank you to my brothers, Andrew and Timothy, for keeping me grounded while my head was in the academia clouds. Thank you to my father, Bernard Powers, and my mother, Diane Powers, for always supporting and challenging me throughout my life. I truly hope that I have made you proud!

TABLE OF CONTENTS

	Page
List Of Tables	viii
List of Figures	ix
List of Abbreviations	x
Abstract	xi
Chapter One - Introduction	1
Section One – Research Objective	1
Section Two – Navy Destroyer (DDG) Program and Future Configurations	2
Section Three – Congressional Issues	4
Section Four – Force Structure Requirement Analysis	6
Section Five – Methodology	9
Section Six – Contribution	10
Chapter Two – Literature Review	13
Section One – Application Domain	13
Section Two – Decision Making	13
Section Three – Dynamic Programming	16
Section Four – Approximate Dynamic Programming	19
Chapter Three – Research Problem and Model Description	21
Section One – The Model	21
Section Two – Data Elicitation	22
State transition probabilities	24
Decision impacts on transition probabilities	25
Decision rewards	27
Configuration impact on decision rewards	28
Correspondence Analysis	29
Section Three – Dynamic Programming	32
DDG Configurations	34

DDG States	35
DDG Actions	36
Action Contributions	37
Transition Probabilities.....	38
Section Four – Approximate Dynamic Programming.....	38
Chapter Four – Solution Methodology	41
Section One – Correspondence Analysis	41
Section Two – Dynamic Programming.....	42
Section Three – Value Iteration Model Inputs	43
Contribution Matrix $C_0(S_t, a_t)$	44
Configuration Impact on $C_0(S_t, a_t)$	45
Configuration 1 (D_I) Transition Probability Matrix $P_I'(S_t, a_t, S_{t+1})$	46
Section Four – Mixed Integer Programming Model	49
Section Five – Approximate Dynamic Programming.....	50
Section Six – Q-Learning Model Inputs	52
Chapter Five – Analysis and Experimentation	53
Section One – Dynamic Programming Results.....	54
Configuration 1	57
Configuration 2.....	58
Configuration 3.....	59
Configuration 4.....	60
Configurations 5, 6, and 7	61
Section Two – Mixed Integer Programming Results	64
Configuration Assignments and R3B Submission	67
Section Three – Approximate Dynamic Programming Results	70
Discount Parameter Impact on Near-Optimal Q-learning Policies	71
Cosine Similarity and Binary Vector Translation.....	71
Configuration 4 Optimal Policy Comparisons.....	72
Configuration 5 Optimal Policy Comparisons.....	74
Section Four – Validation	76
Section Five – Summary	76
Section Six – Additional Research Questions.....	77

How does one measure strategic adherence to DP/ADP – derived optimal/near-optimal policies?.....	77
How do DDG operational timelines map to discount factor (γ) setting with regards to myopic vs. future-seeking policies?	78
Chapter Six – Conclusions.....	79
Section One – Methodology.....	79
Section Two – Application.....	80
Section Three – Broader Contribution	81
Section Four – Future Work.....	82
Appendix A: Sample Decision Impact Calculation Spradsheet.....	84
Appendix B: Sample Contribution Matrix.....	85
Appendix C: Contribution Baseline and Configuration Effect.....	87
Bibliography	88

LIST OF TABLES

Table 1. Response distribution contingency table	31
Table 2. Likert-Numerical mappings	42
Table 3. Configuration-agnostic contribution matrix	44
Table 4. Configuration impact on $C_0(S_t, a_t)$	45
Table 5. $P_I'(S_t, a_t, S_{t+1})$ state-space subset s_k matrices.	46
Table 6. Difference in $P_I'(S_t, a_t, S_{t+1})$ when action a is taken at time t	48
Table 7. R4Q cosine similarity to R_4^*	72
Table 8. R5Q cosine similarity to R_5^*	74

LIST OF FIGURES

Figure 1. Decision-agnostic mock TPM example with notional probabilities	25
Figure 2. Decision impact on transition probabilities with two notional values	26
Figure 3. Configuration-agnostic decision rewards with two notional values.....	28
Figure 4. Configuration impact on rewards with two notional values.....	29
Figure 5. Sample bi-directional 7-point Likert scales for participant reference	30
Figure 6. Scaled utility (g_i) for each configuration (D_i).....	56
Figure 7. Configuration 1 optimal strategy R_1^* distribution.....	57
Figure 8. Configuration 2 optimal strategy R_2^* distribution.....	58
Figure 9. Configuration 3 optimal strategy R_3^* distribution.....	59
Figure 10. Configuration 4 optimal strategy R_4^* distribution.....	60
Figure 11. Configurations 5, 6, 7 optimal strategy R_5^* , R_6^* , R_7^* distribution.	62
Figure 12. Budget vs. Expected Maximum Utility	66
Figure 13. R3B study submission	68
Figure 14. γ vs. configuration 4 cosine similarity.....	73
Figure 15. γ vs. configuration 5 cosine similarity.....	75

LIST OF ABBREVIATIONS

A2/AD	Anti-Access/Area-Denial
AAW	Anti-Air Warfare
ADP.....	Approximate Dynamic Program/Programming
AI	Artificial Intelligence
AMDR	Air and Missile Defense Radar
ASW.....	Anti-Submarine Warfare
ASuW	Anti-Surface Warfare
AWS	Aegis Weapon System
C2	Command and Control
CA	Correspondence Analysis
CAPE	Cost Assessment and Program Evaluation
CNO	Chief of Naval Operations
COA	Course of Action
CONOPS	Concept of Operations
CSG	Carrier Strike Group
DDG	Navy Destroyer
DM	Decision Maker
DoD	Department of Defense
dLNA	Digital Low Noise Amplifier
DP	Dynamic Program/Programming
ESG	Expeditionary Strike Group
GB	Gigabyte
LNA	Low Noise Amplifier
LSC	Large Surface Combatants
MB	Megabyte
MDP	Markov Decision Process
MIP	Mixed Integer Programming
MYP	Multiyear Procurement
OSD	Office of the Secretary of Defense
P _K	Probability of Kill
R3B	Resource and Requirements Review Board
RL	Reinforcement Learning
SAG	Surface Action Group
SEWIP	Surface Electronic Warfare Improvement Program
SME	Subject Matter Expert
STORM	Synthetic Theater Operations Research Model
TACSIT	Tactical Situation
TPM	Transition Probability Matrix
UUV	Unmanned Underwater Vehicle
VLS	Vertical Launch System

ABSTRACT

AN APPROXIMATE DYNAMIC PROGRAMMING APPROACH TO FUTURE NAVY FLEET INVESTMENT ASSESSMENTS

Matthew J. Powers, Ph.D.

George Mason University, 2022

Dissertation Director: Rajesh Ganesan

Navy decision makers and program sponsors must decide on future investments in the context of forecasted threats and operating environments. Investment assessments are difficult in that forecasted costs and utilities are oftentimes based on non-existent resources and technology. Forecasted projection model vectors are informed by current data that reflect similar or “close as possible” technologies, and are limited to scenario scope. That is, the common assessment modeling method of placing representative agents in a scenario-based simulation to assess future investment utilities are limited by scenario and design capabilities.

The research objective is to combat the limitations of specific scenario-based analyses by modeling the operational lifespan of future Navy Destroyer (DDG) fleet configurations as Markov decision processes (MDPs) evaluated with dynamic programming (DP) value iteration to calculate the maximum DDG-configuration utility.

All MDP parameters are informed by existing models and subject matter expert (SME) inputs. The transition probability matrices (TPMs) assess the probabilities that a DDG transitions between states as a chance function of future configuration capabilities and sequential actions that are more representative of the operational lifetime of a configured DDG than that of a single scenario. Likert type values are assigned to each pairwise decision-state so that Bellman's optimality equation solves for maximum expected value via non-discounted value iteration. These maximum expected values become the decision variable coefficients of an integer programming configuration-destroyer assignment model that maximizes the sum of destroyer-configuration values according to budgetary, logistic, and requirement-based constraints. DP value iteration is appropriate for this problem in that the algorithm does not require a time-value discount parameter and the objective is the maximum expected value, and I compare DP results to the approximate dynamic programming (ADP) method of Q-learning. Modeling with ADP removes the need for TPMs for large problem instances, thereby providing a framework for near-optimal decisions, and this research highlights the similarities in the solution between ADP and DP. ADP results align with DP results because the accurate ADP parameter settings enable learning and exploration that guarantees near-optimal ADP solutions, thereby opening the door for computationally scalable algorithms.

This work contributes to SME and DM insight, mitigating bias towards technologically superior configurations by revealing utility values that make the seemingly less capable configurations more competitive in terms of long-term value. This insight is due to DP optimal policies and ADP near-optimal policies that are driven

by values of the states, an insight that would not have been possible without this research. This study demonstrates that the less advanced technologies can be deployed in such a way to maximize their long-term utility so that they are more valuable than expected in future operational environments.

OPNAV desire for modeling methods that complement existing campaign models is evidenced by this method's briefing to incoming OPNAV analysts as a "best practice" for evaluating complex decisions. This study's contribution to the high-visibility R3B study received high-level recognition in a *Presidential Meritorious Service Medal* citation. This research contributes AI-enabled decision-making in a culture that relies on familiar, anecdotal, or experience-based approaches. AI-enabled decision-making is necessary to compete with near-peer adversaries in dynamic decision-making environments.

CHAPTER ONE - INTRODUCTION

The Chief of Naval Operations (CNO) chairs the Resources and Requirements Review Board (R3B) to manage shipbuilding programs that align with the Navy's significant investments to maintain technologically superior warships. CNO staff divisions assess aspects of future programs to calculate return on investment across the fleet. This study represents one such effort to evaluate optimal future destroyer (DDG) configuration assignments to maximize operational benefit while adhering to budgetary, logistic, and requirement-based constraints. Operational benefits are defined as the value added to the fleet if a DDG upgrades to a specific configuration in the context of ever-evolving near-peer adversary capabilities. Ideally, all DDGs would receive configuration upgrades so that they would be deployable and tasked as a blanket-appropriate response to any threat, vulnerability, or operational state. Unfortunately, competing Navy investments require that money and resources be spent judiciously while still meeting operational demand. Furthermore, not all DDGs are physically capable of receiving all available upgrade configurations, but their significant remaining service life deems them eligible for feasible upgrades.

Section One – Research Objective

The objective of this research is to determine maximum future DDG configuration utility while adhering to cost limitations through a method that satisfies

expectations of analytic rigor while remaining consistent with the beliefs held by decision makers. In this spirit, I apply dynamic programming (DP) to model operational decision making under uncertainty in an adversarial environment. As a secondary objective, I compare DP results with approximate dynamic programming (ADP) results that have been informed by the same parameters. My study demonstrates the potential benefits of evaluating long-term utility through methods that, while consistent with decision maker beliefs, require calculations that are beyond human capability.

Section Two – Navy Destroyer (DDG) Program and Future Configurations

The DDG 51 Arleigh Burke Class Destroyer program began in the late 1970s and procured its first DDG 51 in FY1985. DDG 51 class guided missile destroyers are categorized as Large Surface Combatants (LSCs) that provide multi-mission offensive and defensive capabilities. Destroyers can operate independently or as part of Carrier Strike Groups (CSG), Surface Action Groups (SAG), and Expeditionary Strike Groups (ESG), capable of conducting Anti-Air Warfare (AAW), Anti-Submarine Warfare (ASW), and Anti-Surface Warfare (ASuW). The DDG 51 class has an expanded strike warfare role in its utilization of the MK-41 Vertical Launching System (VLS). The DDG 51 class also integrates the Aegis Weapon System (AWS), composed of a multi-function phased array radar, advanced AAW and ASW systems, VLS, and the Tomahawk Weapon System. The DDG 51 class has been, and continues to be, upgraded with advanced sensors, weapons, and support systems. The ability to operate independently or within a group under different warfare postures motivates the states and actions modeled

as part of a Markov decision process (MDP) in this study, which calculates the utility value of potential DDG 51 upgrades in the context of their operational lifespans.

Future upgrades apply to DDG 51 variants known as *Flights*, which are distributed across DDG hull numbers. DDG hull numbers 51-71 are the original class design and are known as Flight I ships; DDGs 72-78 are Flight II ships; DDGs 79-124, and 127, are Flight IIA. The Flight IIIA (Flight III baseline) hull numbers are 125-126, and 128-137. This study also includes future variant Flight IIIB, hull numbers 138-145. Future upgrade designs include combinations of the more capable Air and Missile Defense Radar (AMDR, aka SPY-6), the Surface Electronic Warfare Improvement Program (SEWIP) Block 3, which adds electronic attack to the existing SLQ-32(V) electronic warfare system to meet the urgent operational needs of the fleet, and Low Noise Amplifier (LNA) or Digital LNA (dLNA) technologies to improve sensitivity and capability over legacy (SPY-1D) radars. A DDG, like any ship in the fleet, is part of a networked system of sensors wherein complementary strategic resources relax the demands of a specific ship if the limitations of that specific ship may be balanced by the advantages of another. In this, a value-maximizing solution can include less-than-ideal upgrades for DDGs based on their current configuration and expected assignments.

Overall, DDGs 51-145 are candidates for future upgrades, with each Flight adhering to logistic and physical capability-based limitations. This study refers to these future upgrades as *configurations*, described as follows:

- Configuration 1: SPY-1D, SEWIP Block 2 (Current configuration)
- Configuration 2: SPY-1D, SEWIP Block 3

- Configuration 3: SPY-1D, SEWIP Block 2, dLNA
- Configuration 4: SPY-1D, SEWIP Block 3, dLNA
- Configuration 5: SPY-6, SEWIP Block 2
- Configuration 6: SPY-6, SEWIP Block 3
- Configuration 7: SPY-6, Flight III (A and B), SEWIP Block 3

The future configurations may be introduced to new-or-future construction ships (such as Flight IIIB DDGs) or to in-service ships. This assures increased baseline capabilities of the new ships while providing commonality between new ships and modernized in-service ships according to the multiyear procurement (MYP) contract that Congress approved as part of its action on the Navy's FY2018 budget. The DDG modernization effort intends to maximize warfighting capabilities while reducing total ownership cost to the Navy. This objective motivates this study's approach to achieve maximum future configuration utility subject to budgetary, logistic, and capability-requirement constraints.

Section Three – Congressional Issues

DDG 51s are being procured in FY2018-FY2022 under the MYP contract according to the Navy's FY2018 budget. DDG 51s procured in FY2017 and beyond are being built to the Flight III design (O'Rourke 2020). The Flight III DDG has the space, weight capacity, cooling ability, and power to handle both the SEWIP Block 3 and SPY-6 at full capacity and capability, thereby offering improved utility over Flight II DDGs equipped with SEWIP Block 3 and SPY-6. This study refers to a Flight III DDG with SEWIP Block 3 and SPY-6 as configuration 7. FLT III DDGs 125, 126, and 128-137 are

not equipped with SEWIP Block 3, they instead have SEWIP Block 2, an enhanced version of the legacy passive system, making those DDGs available as candidates for configuration 7 upgrades. DDGs 138-145 will be procured as configuration 7. To scope the budgetary considerations of this study, the total amount of procurement funding requested for the FY2021 DDG 51 program is \$3,079.2M, excluding outfitting and post-delivery costs. The Navy's desire to procure the first ship of a new class of large surface combatants in FY2028 makes FY2027 the final year of DDG 51 procurement. Therefore, for the seven fiscal year period of FY2021–FY2027, the procurement budget for configuration 7 DDGs *alone*, ignoring inflation, is \$21,554.4M. Factoring in scheduling and other technical risk costs increases DDG 51 program total monetary requirements, and costs for configuration upgrades to Flight I and II DDGs have yet to be considered. These budgetary considerations are concerning in the face of projected reduction to the Navy's FY2021 DDG 51 procurement budget. This motivates a modernization and procurement plan that maximizes the lifetime utility of DDGs that considers cost limitations.

Maximized lifetime utility requirements are reflected in the operational necessity of a fleet architecture that is more distributed than that of the 355-ship goal that includes 104 LSCs. This distributed fleet architecture expects a smaller proportion of DDGs that must integrate with smaller ships, lightly manned patrol craft, or unmanned underwater vehicles (UUVs) to respond effectively to the improving maritime anti-access/area-denial (A2/AD) capabilities of near-peer adversaries such as China. This operational necessity demands the type of technical feasibility offered in the configuration upgrades while

adhering to future Navy budgets that are expected to be smaller than the expenses associated with the current fleet architecture. The Hudson Institute, a private think tank that informs the Office of the Secretary of Defense (OSD) Navy force-level analysis, warns that the FY2021-budgeted force structure overemphasizes LSC requirements, and includes too few ships to distribute the fleet and create sufficient complexity to counter A2/AD capabilities (Clark, Walton, Timothy A., and Cropsey, Seth 2020).

Navy force-level goal recommendations from organizations such as the Cost Assessment and Program Evaluation (CAPE) office within OSD and the Hudson Institute range between 64 to 90 LSCs, a range that sees a reduction from the 355-ship goal of 104 LSCs. Reducing from 104 LSCs to some smaller number while also requesting new Flight III DDG 51 procurements, likely at a decelerated procurement rate, demands a procurement and configuration modernization strategy that yields maximum utility within acceptable cost and time. Arriving at this strategy demands rigorous force structure requirement analysis.

Section Four – Force Structure Requirement Analysis

Navy force structure assessments assemble “bottom up” requirements by building campaign plans that use modeling and simulation to fight and succeed in designed scenarios. The force structure analysis combines these campaign assessment results with combatant commander naval presence requirements. The force structure requirement analysis must consider how to equip, organize, supply, maintain, train, and employ naval forces. Simulation-based campaign analyses measure risk for investment options that will determine naval assets and capabilities for decades to come. However, campaign

modeling to inform force structure requirements relies heavily on attrition-based metrics that do not implement operational concepts that prevent adversary success over the lifetime of the fleet systems. Campaign models, even when combined with naval presence requirements, may be incomplete when considering emerging strategic environments. Adversaries such as China, Russia, North Korea, and Iran have established robust sensor and weapon networks that force US commanders to accept significant risk when deciding on long-term deployment and presence strategies. Furthermore, adversaries are likely to employ mixes of non-violent military and paramilitary actions that do not trigger the kinds of US retaliations modeled in the attrition-centric campaign models. This study proposes a method of analysis that considers risk in the context of long-term naval presence in adversarial environments when evaluating DDG configuration utility (Clark et al 2020).

Campaign model construction is costly in terms of input data, time, and resource requirements. Similarly, campaign model output analyses are extremely demanding and time consuming because of the considerable amount of complex output data. Campaign model complexity grows exponentially with the number of units, facilities, sensors, or weapons. Depending on the model being used, a single run may generate many gigabytes (GBs) of data that demands rigorous analysis to produce useful insights. The analysis must identify trends and relationships across the simulation runs to understand how specific investments (such as DDG configurations) affect the campaign risk assessment. Furthermore, these analyses inform Cabinet-level decisions that are intolerant of falling behind schedule, so the turnaround between model construction and output analysis is

rapid (Loerch and Rainey 2007). One such common, large-scale stochastic simulation model that is commonly used in naval campaign analysis is the Synthetic Theater Operations Research Model (STORM). The US Air Force originally developed STORM to model land, maritime, amphibious, air, space, and logistic campaign elements. STORM represents the classic OR endeavor of enterprise risk assessment. As such, fast modern computers may take hours or days to evaluate campaign results. STORM models may require up to 40 megabytes (MB) across 150 input files worth of data containing scenario details about individuals, groups, entities, and rules. Advancement from the mid-20th century digital computing advent enables rapid model growth in scope, complexity, and realism (Lucas, Kelton, Sánchez, Sanchez, Anderson 2015). A STORM campaign is a complex system with big data challenges. Multi-month campaigns, dozens of ships and battalions, hundreds of aircraft and installations, all encompass tens of thousands of exchanges of combat engagements. Multiple decision cycles within these engagements may yield hundreds of millions of possible actions (Morgan et al. 2018).

Despite the impressive volume of input and output data and complexity, STORM remains a campaign analysis simulation of future possibilities within the context of designed scenarios for which it is impossible to compare to “ground truth.” Instead, SMEs judge the credibility of model designs and findings. This judgement effort is non-trivial. Acquiring, vetting, and verifying the data is a challenging, multi-organizational task conducted across the Department of Defense (DoD). Modelers obtain stakeholder consensus regarding physical aspects such as ranges, speeds, capabilities, and behaviors that accurately reflect a concept of operations (CONOPs) and tactical situations

(TACSITs). For STORM simulations to be accepted by many diverse parties, modeled entities must behave and interact credibly. This credibility is achieved by incorporating performance data from higher-fidelity models, experiments, live exercises, and SME judgement (Morgan et al. 2018). The concept of external model incorporation combined with existing knowledge from exercises and SMEs is the connective tissue between STORM-like campaign models and this study.

This study proposes analytic methodologies that require much fewer resources and computational expense, but enjoy the detailed output analyses and insights from existing campaign models, exercises, and SME judgement. In this, sensitivity analysis and responses to “what if” questions do not require computationally expensive and time-consuming scenario updates and runs. By modeling the campaign analysis outputs as an MDP, it is not necessary to re-run a STORM-like model for every variation of the scenario. The MDP parameters use existing outputs to build the DP and ADP models whose fast computations make them attractive for sensitivity analysis. Moreover, modeling the operational lifetime of future DDG configurations as an MDP divorces the evaluation from scenario-based limitations, while rapid DP/ADP evaluation satisfies the quick-turn demands.

Section Five – Methodology

In response to scenario-based analysis limitation, this study models the life cycle of future DDG configurations as MDPs evaluated with DP value iteration and ADP Q-learning (aka reinforcement learning) to calculate the maximum configuration utilities. SMEs and decision makers (DMs) inform the transition probability matrices (TPMs)

containing the probabilities that a DDG transitions between states as a function of configuration capabilities and actions. From there, Bellman's optimality equation solves for maximum expected value via non-discounted value iteration. These maximum expected values, heretofore referred to as *utility* values, become the decision variable coefficients of a mixed integer programming (MIP) configuration-DDG assignment model that maximizes the sum of configuration-DDG utility values according to budgetary, logistic, and requirement-based constraints. An ADP model with identical parameters as those of the DP model is evaluated for comparison and DP model confirmation.

Section Six – Contribution

This body of work contributes to the desire for the DoD to wield AI-enabled decision-making in an organization that relies on anecdotal, experience-based, or antiquated methods. The military recognizes the potential for AI as a necessary resource to keep pace with near-peer adversaries, yet stubbornly sticks to military decision-making processes that fail to comprehend dynamic environments over large time horizons. (Lettau and Uhlig 1999) nicely ponder the reasons behind this phenomenon:

It is intriguing to speculate about possible resolutions such as instincts, learning from your peers, education, meta-rules for changing rules, or the neuron limits of the brain. We simply take these given rules, as well as the fact that the agent stubbornly sticks to choosing between them throughout his infinite life as primitives of the environment.

Truly understanding the future utility of modernization investments deployed in a dynamic, temporal environment demands that decision makers break free from the constraints of bounded rationality. Decision makers must recognize human cognitive limitations and ensure decision consistency by employing approaches akin to those used in this study. This work also demonstrates alternatives to computationally expensive, scenario-based campaign models that evaluate attrition-based metrics that are unlikely to be observed in adversarial actions that fall below the threshold for hostility.

While the SMEs and DMs display an obvious bias towards the technologically superior configurations 6 and 7, this study results in utility values that make the seemingly less capable configurations more competitive in terms of long-term value than expected by SMEs and DMs. This is due to the insight that DP and ADP offer by way of optimal and near-optimal policy outcomes, an insight that would not have been possible without this research. I show that DP optimal policies align with ADP near-optimal policies because accurate ADP parameter settings enable learning and exploration that guarantee near-optimal solutions. I demonstrate that the less advanced technologies can be deployed in such a way to maximize their long-term utility so that they are more valuable than expected in future operational environments.

OPNAV desire for modeling methods that complement existing campaign models is evidenced by this method having been briefed to incoming OPNAV analysts as a “best practice” for evaluating complex decisions. This study’s contribution to the high-visibility R3B study earned high-level recognition in a *Presidential Meritorious Service Medal* Citation:

Lieutenant Commander Powers' superior professional knowledge and unmatched analytic talent were instrumental in supporting senior leaders' decisions across three budget cycles. His development of a unique analytic tool suite synergized operator inputs, predicted mission outcomes and program costs to better identify and clarify investment options. Use of his tool, in collaboration with stakeholders, resource sponsors and analysts, guided resource and requirement decisions for the combat systems supporting the Flight III DDG, DDG modernization, and Light Amphibious Warship programs. Furthermore, the resulting analytic insights greatly assisted senior leader investment decisions that will shape the future surface Fleet. Lieutenant Commander Powers' in-depth analysis, cross-domain expertise and actionable programmatic recommendations enabled the Chief of Naval Operations and other senior Navy leadership to provide effective guidance and make timely, cost-effective program decisions.

CHAPTER TWO – LITERATURE REVIEW

Section One – Application Domain

This chapter focuses on the domains of decision-making, dynamic programming, and approximate dynamic programming. Decision-making limitations and how DP represents solutions to these limitations segues into the detailed DP discussion. This chapter concludes with a discussion on ADP as a decision-making aid. These domains do not represent the entirety of my body of work. For example, correspondence analysis (CA) methods employed by (Powers 2016) and (Schramm and Powers 2017) are described in detail in Chapter Four. CA is a method for translating ordinal data, like the ones gathered here during data elicitation, into numerical values that are appropriate for model parameterization. This study also applies a measurement known as cosine similarity to compare optimal DP and near-optimal ADP policy vectors, but cosine similarity is not described here. Furthermore, assuming knowledge of the rich MIP optimization history, the parameterization and application of MIP in this study are described in Chapter Four – Solution Methodology.

Section Two – Decision Making

Populating the data necessary to parameterize the DP model for this study requires SME and DM input regarding state transitions, how actions impact those transitions, and how actions in each state affect DDG (or mission) vulnerability. The details of the data elicitation process are described in Chapter Three – Research Problem and Model Description. As described by (Keeney 2004), the participants become very

familiar with the concept of structuring decisions throughout the data elicitation exercise, but they are not required (nor expected) to follow the DP aspect of the analysis. I assure that the participants are presented with a clear list of data elicitation objectives, such as determining transition probabilities within a reasonable degree of precision, or specifying a vulnerability-based value tradeoff to populate the contribution matrix. In this, the complexity and nuance of the DDG operational environment and associated requirements are captured in a form that becomes amenable to DP. The DP model, while capable of calculating optimal decision strategies beyond human capability, behaves consistently with participant decision making patterns, adding to the prescriptive appeal of this modeling approach.

A strong prescriptive appeal is in stark contrast to some of the examples of poor decision analysis methods described by (Keeney 2004), such as failing to clarify objectives, not quantifying probabilities of alternate decisions, relying solely on worst case analysis, assuming a linear relationship between Likert responses and averaging the response values, and avoiding any subjective factors. My methodology avoids these pitfalls. For example, my approach clearly articulates the objective of minimizing vulnerabilities associated with configuration decisions while avoiding singular, worst case attrition metrics such as probability of kill (P_k). To avoid inappropriate averaging, I apply CA to response value distributions to calculate numerical scores for each metric. Subjective factors are captured by adjustable MIP constraints such as minimum system or configuration requirements, themselves defined by factors such as economic or political influences. I present this study's DP application as an evolved decision-making process.

Artificial Intelligence (AI), a core application of DP, should be exploited to continue the military decision-making evolution. (Schramm and Clark 2021) emphasize the adaptability of AI-enabled processes that quickly develop and propose courses of action (COAs) that human planners are incapable of considering. They go on to suggest that successful AI exploitation combines force design with AI-enabled Command and Control (C2), itself a large-scale decision-making process. My method adheres to this suggestion by combining future fleet configuration force design with DP-derived optimal employment in a complex C2 operational environment. Put simply, if these DDGs were to behave optimally according to AI algorithms, then I calculate their maximum utility. Moreover, this utility exists as a provider of prolonged advantage over near-peer adversaries such as China and Russia, in contrast to the myopic and attrition-based utility evaluated by combat models. DP provides the advantage of extending the time horizon because the operational environment is modeled as an MDP. My method exemplifies the emphasis that (Schramm and Clark 2021) put on removing the “buzzword” stigma from AI and using it as an actionable decision-making tool for the Services.

AI-enabled decision-making tools mitigate problems addressed by (Johnson, Kotlikoff, and Samuelson 1987) in an economic consumption experiment where they observed that most of their subjects undervalue future utility due to human inability to make consistent life cycle decisions. Decision patterns are only weakly correlated with intemporal preferences. The authors reject the hypothesis that humans are capable of consistent, temporal decisions based on their evaluations of present and future resources. By applying DP to SME-provided parameters, my method ensures DDG utilization

consistency over a large time horizon based on present and future configuration evaluations.

Decision-making inconsistencies are exacerbated when combined with heavy reliance on “rules of thumb,” as described by (Lettau and Uhlig 1999). The authors define these “rules” as mappings between states and actions that are constructed by averaging past experiences. This can lead to “good state bias,” which myopically favors rules that may make bad decisions because they only apply in good states, and they fail to account for the dynamic nature of the problem. Good state bias chooses a strategy that maximizes historical average payoff vice optimal current state decisions. This bias was observed during data elicitation when the SMEs initially failed to understand why an optimal DDG strategy would not always choose the “best” (or greedy) option from any state, despite their recognition of the dynamic DDG operational environment. (Lettau and Uhlig 1999) explicitly states that DP corrects this bias. The precedence for DP as an OPNAV decision-making improvement method is seen in (Powers 2020), wherein future landing craft investment options are calculated using Bellman’s as a knapsack problem, a study that is also recognized in the *Presidential Meritorious Service Medal* Citation.

Section Three – Dynamic Programming

Dynamic Programming improves decision making because it abandons the infeasible assumption of “unbounded rationality,” a favorite phrase in (Rust 2019) wherein the author recognizes DP as a “powerful tool for solving sequential decision-making problems under uncertainty.” Unbounded rationality suggests that individuals

possess unbounded levels of rational, computational ability. This is simply untrue. Humans are unlikely to comprehend optimal strategies that consider near and long-term consequences in complex, large time horizon situations. Therefore, it is irrational to assume that decision makers in hostile and uncertain environments will make decisions that maximize the utility of their resources. Even in the simplified abstract of the operational environment in my study, the curse of dimensionality would overwhelm a human decision maker in terms of optimal DDG deployment. However, (Rust 2019) suggests that a decomposed representation of an impossibly complex, intractable problem such as DDG deployment strategies can be abstracted and solved with DP. The author reinforces the value of this decision-making strategy by quoting economist Herbert A. Simon's 1978 Nobel Memorial Lecture, "Rational decision-making in business organizations."¹

Decision makers can satisfice either by finding optimum solutions for a simplified world, or by finding satisfactory solutions for a more realistic world. Neither approach, in general, dominates the other, and both have continued to coexist in the world of management science.

My model discovers optimum solutions in a simplified world. However, it is a simplified understanding of the operational environment in which future-configured DDGs will operate as defined by the very decision makers who will determine which

¹ From Simon H. 1978. Rational decision-making in business organizations. Nobel Memorial Lecture, Dec. 8. <https://www.nobelprize.org/uploads/2018/06/simon-lecture.pdf>. Accessed March 18, 2021.

configurations will receive funding. This simplification is an important discussion point when considering how realistic my model is. (Rust 2019) acknowledges that DP solutions may not be useful if the problem structure differs greatly from reality. However, (Watkins 1989) assures that any process can be modelled as an MDP, as long as the state space captures the relevant aspects of the real-world process. In the case of impossibly complex problems, (Rust 2019) says that “direct elicitation of preferences and beliefs” from decision makers is necessary to structure the DP problem. This echoes the approach in (Keneally, Robbins, and Lunday 2016), wherein Monte Carlo methods generate data to parameterize their model. My model is, in fact, a realistic representation of decision makers, with their values provided by existing models, SMEs, stakeholders, and program sponsors. Their knowledge of the decision problem assures the usefulness of the DP model because their knowledge informed the parameters. Decision makers find this very attractive.

Assuming the usefulness of the DP model, my approach calculates the utilities of future DDG configurations by extracting parameter g , the expected value in Bellman’s equation for value iteration and a necessary convergence parameter. This method is seen in (Abdulla et al. 2018) and (Díaz et al. 2018), wherein DP is applied to approximate battery lifetime values and energy grid connections, respectively. Similar to how my model assigns configuration utility values as MIP objective function coefficients, (Xi, Sioshansi, and Marano 2014) uses optimal DP expected values to construct an MIP to optimize distributed energy storage.

Section Four – Approximate Dynamic Programming

The value iteration results satisfy the requirements of the R3B study that motivates this work. For academic and curiosity purposes, I compare DP results to an ADP model of the same problem. It is also noteworthy that ADP is useful when the problem scale exceeds DP tractability. Specifically, I apply the reinforcement learning (RL), model-free algorithm known as Q-learning.² According to (Rust 2019), value iteration algorithms of DP problems inspired stochastic versions of these algorithms, such as Q-learning, that asymptotically converge to near-optimal decision rules and value functions. Q-learning works well for problems with a relatively small state space, as recognized by (Jiang et al. 2014), which discusses the shortcomings of lookup tables in large scale applications. However, modern computing speed and the relatively small scale of my DDG future configuration problem make Q-learning a tractable technique. Furthermore, the Q-learning value function $Q(s, a)$ for state s and action a , while not as accurate as value iteration, is sufficient to model decision maker choices.

This study uses the Q-learning R package, *ReinforcementLearning*, which refers to (Sutton and Barto 2018) for its technical and theoretical details³. The authors recognize that Q-learning convergence is not a point-convergence, but rather a band-convergence that satisfies most common, real-world decision-making requirements. This stems from the assumption stated by (Silver et al. 2017), wherein Q-learning can learn for itself, provided a solid base of first principles that need not include state transition

² Model-free means that transition probabilities are unknown and not required for convergence.

³ *ReinforcementLearning* literature from (Pröllochs and Feuerriegel 2018).

probabilities. In my model, these “first principles” refer to the action-reward and state-state reachability parameters in a tactical environment provided by SMEs and DMs. This resembles (Summers and Robbins 2020) and (Davis 2017), wherein ADP determines near-optimal sequential missile engagement strategies. Both authors apply ADP in anticipation of large-scale dimensionality issues that make DP untenable. My model captures sequential decision making in a tractable operational context, including encounters of various operational states beyond those of tactical attack responses. The similarities exist in the need for near-optimal policies in a complex, dynamic environment.

The uncertainty that exists in dynamic environments motivates (Robbins et al. 2020) and (Rettke, Robbins, and Lunday 2016), where ADP solves for near-optimal military medical evacuation (MEDEVAC) dispatching policies. The authors compare ADP results to existing dispatch policies that are considered “optimal.” Both studies note that, particularly in combat, real-world MEDEVAC policies resemble myopic solutions. (Rettke, et al. 2016) results demonstrate a 31% lifesaving performance improvement between myopic and ADP policies. Since there are no real-world examples of future configuration DDG deployment, I compare DP results to Q-learning results under the same conditions with varying levels of myopic vs. future reward-seeking behavior. (Robbins et al. 2020) goes on to recognize the benefit of ADP as a framework to compute high quality policy approximations with less computational expense than a high-fidelity scenario-based simulation. By demonstrating my work as a decision-making framework, my approach has been adopted as a best practice for OPNAV assessments.

CHAPTER THREE – RESEARCH PROBLEM AND MODEL DESCRIPTION

The objective of my research is to determine maximum future DDG configuration utility within acceptable cost in a manner consistent with the beliefs held by decision makers. I accomplish this by way of DP, specifically evaluating maximum expected value from Bellman's value iteration equation. To parametrize this model, I require data that are elicited from SMEs and DMs. This chapter describes the data elicitation process, including a discussion on correspondence analysis (CA) Likert data transformation to account for inequality or non-linearity between response values. Next, I describe DP and ADP modeling, and how they apply to this problem. However, I begin by describing the overall model.

Section One – The Model

To achieve my research objective, I use SME-elicited data to develop TPMs and contribution matrices for a DP value iteration for average criteria model. From this model, I extract the expected maximum value g , which is the objective function for this equation. I calculate g for each DDG configuration, which become the decision variable objective function coefficients in a maximization MIP purposed to optimize configuration investment decisions constrained by budget, logistics, and requirements. Assuming reader knowledge of MIP processes, I refrain from describing them in this chapter. The output of the DP and MIP models are the results presented to the R3B that motivated this research. For academic and comparison purposes, I compare DP optimum policies R^* to

ADP optimum policies with varying γ , which is the Q-learning discount factor required for convergence.

Section Two – Data Elicitation

Data elicitation, as described in this section, is a non-trivial process that must account for potential misunderstanding regarding problem definition and requirements, inequality between numerical response values, and inconsistencies between responses. I describe my data elicitation process in detail to explain how I deal with these, and other, common problems.

The data required to inform this study were elicited from participants in an OPNAV N81 Pentagon conference room during the months of August – September 2020. All participants are SMEs in the areas of operational assessments, surface warfare, or both, as they are all employed across various OPNAV assessment or surface warfare program sponsor organizations. All participants from OPNAV N96, the surface warfare program sponsor organization, are either decision makers or direct representatives of the decision makers who are active in the R3B study that motivates this work. All participants are comfortable with the 7-point Likert scale responses that are appropriate for the context of each question. Participants are familiar with the DDG program, the mission sets, and requirements. They are familiar with existing literature and assessments related to future configuration investments. Their combined expertise includes familiarity with the systems that comprise the DDG configurations that are being evaluated, specifically the Air and Missile Defense Radar (AMDR, aka SPY-6), the Surface Electronic Warfare Improvement Program (SEWIP) Block 3, the existing SLQ-

32(V) electronic warfare system, and the Low Noise Amplifier (LNA) or Digital LNA (dLNA) to improve (SPY-1D) radar capabilities. Participants from OPNAV N96, while not expected to fully understand the theoretical analytics underlying this study, were familiarized with MDPs and provided the following study methodology explanation:

Using a Markov Decision Process, we will solve for optimal risk decisions based on a DDG's potential states within the Joint Force Operational Scenario. For each state, a reward matrix will be applied using values abstracted from subject matter experts and existing analysis. The optimal Markov Decision values will then be the basis for a mixed integer programming optimization model, constrained by the DDG Mod budget to determine the best allocation of upgrades across the DDG force.⁴

The concepts of transition probability matrices and award matrices were introduced to participants prior to the data elicitation conferences, and participants were provided with the following three statements to define the problem:

- How do we maximize performance at the force level, minimizing risk to both high-value units and self, within a cost constrained budget?
- Ships are part of a system—a network of sensors and weapons—such that the limitations of a single ship or task group may be balanced by the advantages of another.

⁴ From an UNCLASSIFIED brief entitled, “DDG Modernization Optimization Study: Pre-Kick Off Information Brief,” 12 August 2020.

- Each DDG can be in various states of existence within the scenario and has some probability of changing states based on its current state.

The data elicitation conferences are divided so that participants can focus on four distinct data requirements: State transition probabilities, decision impacts on transition probabilities, decision rewards from each state, and configuration impact on each decision reward. The four-part data elicitation structure enables data requirement fulfillment while minimizing the number of questions being asked, which is critical to avoid respondent fatigue and to obey the time-sensitive nature of this study. Decisions play the *action* role in DP/ADP parameterization, and are listed in this section along with states and configurations. After all data are elicited, I apply CA to response values to calculate numerical scores.

State transition probabilities

Participants come to a consensus on decision-agnostic state transition probabilities for each configuration⁵. These TPMs serve as the baseline from which TPMs for each configuration and action emerge. To clarify this requirement, I display a functional spreadsheet (Figure 1) to demonstrate TPM value assignments⁶.

⁵ Decision-agnostic state transition probabilities are the probabilities of transitioning between states, regardless of any action taken, based on friendly and adversary force laydowns, historical and modeled data, literature, and experience.

⁶ Spreadsheets are displayed on a large screen for all participants to observe. I then demonstrate a brief, mock question-answer session to clarify how we will populate each spreadsheet.

	Carrier Escort	LHA Escort	SAG Unit		Row Sum (must equal 1.0)
Carrier Escort	0.7	0.2	0.1		1
LHA Escort	0.9	0	0		0.9
SAG Unit	0.25	0.25	0.5		1
	Reliable OTH Comms	Unreliable OTH Comms			
Reliable OTH Comms	0.7	0.3			1
Unreliable OTH Comms	0.6	0.4			1
	With Advanced SA	Without Advanced SA			
With Advanced SA	0.8	0.2			1
Without Advanced SA	0.25	0.75			1
	Weap. Cap. 1	Weap. Cap. 2	Combined Weap. Cap.	No Weap. Cap.	
Weap. Cap. 1	0.25	0.25	0.25	0.25	1
Weap. Cap. 2	0.3	0.3	0.3	0.1	1
Combined Weap. Cap.	0.4	0.3	0.1	0.2	1
No Weap. Cap.	0.2	0.2	0.3	0.3	1
	ASCM	ASBM	ASCM/ASBM	None	
Surface Threat Vuln.	0.2	0.3	0.2	0.3	1
Missile Threat Vuln.	0.1	0.4	0.3	0.2	1
Combined Vuln.	0.3	0.1	0.3	0.3	1
No Vuln.	0.25	0.25	0.25	0.25	1

Figure 1. Decision-agnostic mock TPM example with notional probabilities. These notional values demonstrate TPM context to participants with the example, “If I am a Configuration 1 DDG as a Carrier Escort, I am 70% likely to remain a Carrier escort, but there is a 20% chance I am assigned as an LHA escort, and a 10% chance that I am detached to a SAG.” Participants assume that transitions occur at the beginning of each assignment cycle (typically one day). Excel’s conditional formatting highlights the row sum values if they do not equal 1.0 (seen in the red-highlighted cell), ensuring proper TPM data requirements.

Upon completion of decision agnostic TPMs for each configuration, participants evaluate decision impacts on transition probabilities.

Decision impacts on transition probabilities

Participants come to a consensus on decision impacts on transition probabilities, as opposed to individually evaluating probabilities for each of nine actions on fifteen sub-states, for seven configurations. Such a process would require 945 individual probability evaluations, leading to participant exhaustion, low-quality data, and would have demanded more time than allowed by R3B requirements. Instead, participants consider whether an action increases or decreases the transition probabilities. Impact consensus

elicitation is based on a seven-point Likert scale ranging from “negligible impact” to “significant impact.”

To clarify the decision impact requirements to participants, I display a functional spreadsheet (Figure 2) to demonstrate decision impact values.

State/Action	Self-defense-Air	Self-defense-Mssl 1	Self-defense-Mssl 2	Area defense-Air	Area defense-Mssl 1	Area defense-Mssl 2	Wide defense-Air	Wide defense-Mssl 1	Wide defense-Mssl 2
Carrier Escort	7				-7				
LHA Escort									
SAG Unit									
Reliable OTH Comms	Deciding to do Self-defense-Air significantly increases the probability of transitioning to a Carrier				Deciding to do Area defense against missile 1 significantly decreases the probability of transitioning to a Carrier Escort role.				
Unreliable OTH Comms									
With Advanced SA									
Without Advanced SA									
Weap. Cap. 1									
Weap. Cap. 2									
Combined Weap. Cap.									
No Weap. Cap.									
Surface Threat Vuln.									
Missile Threat Vuln.									
Combined Vuln.									
No Vuln.									

Figure 2. Decision impact on transition probabilities with two notional values. Excel’s conditional formatting color-codes responses according to values, as seen in the green (7, significantly increases probabilities) and red (-7, significantly decreases probabilities) sample cells. Participants are reminded that these response scores do not evaluate the quality of a decision, rather the impact that decision would have on transitioning to a state.

This part of the data-elicitation process invites an interesting discussion on data inconsistency. For example, if TPM impact consensus is that a decision reduces probabilities of transitioning into all possible states, then we have an infeasible modification. If all possible probabilities reduce, it is impossible for the updated probabilities to sum to 1.0. Another inconsistency is when participants agree that one or more decisions significantly increase transition probabilities, while the remaining

decisions insignificantly decrease probabilities. I take measures to mitigate the effects of these inconsistencies when they occur so that the elicited impacts are satisfied as much as possible within feasible limitations.⁷ One such measure is to define *significant increase* as a large proportion of allowable probability delta, and offset this delta by distributing the decreased probability value among the remaining decisions so that the updated sum of transition probabilities is 1.0. The next part of the process elicits decision reward values.

Decision rewards

Participants evaluate configuration-agnostic decision rewards in the context of vulnerability impact, represented by a seven-point Likert scale like that from the previous section, ranging from “negligible” to “significant” increase/decrease in vulnerability to the DDG or mission⁸. These rewards are the baseline contribution matrix from which contribution matrices for each configuration emerge. To clarify decision reward requirements, I display a functional spreadsheet (Figure 3) to demonstrate reward value assignments.

⁷ A sample decision impact spreadsheet is in Appendix A.

⁸ Participants consider the balance between self-preservation and impact to the mission. For example, if a decision decreases individual DDG vulnerability, but significantly increases mission vulnerability, participants might arrive at the consensus that this poor decision maps to a 1, significant increase in vulnerability.

State/Action	Self-defense-Air	Self-defense-Mssl 1	Self-defense-Mssl 2	Area defense-Air	Area defense-Mssl 1	Area defense-Mssl 2	Wide defense-Air	Wide defense-Mssl 1	Wide defense-Mssl 2
Carrier Escort	7	1							
LHA Escort									
SAG Unit									
Reliable OTH Comms	If I am a Carrier Escort and I choose Self-defense-Air, there is a significant decrease in vulnerability.	If I am a Carrier Escort and I choose Self-defense against missile 1, there is a significant increase in vulnerability.							
Unreliable OTH Comms									
With Advanced SA									
Without Advanced SA									
Weap. Cap. 1									
Weap. Cap. 2									
Combined Weap. Cap.									
No Weap. Cap.									
Surface Threat Vuln.									
Missile Threat Vuln.									
Combined Vuln.									
No Vuln.									

Figure 3. Configuration-agnostic decision rewards with two notional values. Excel's conditional formatting color-codes responses according to values, as seen in the green (7, significantly decreases vulnerability) and red (1, significantly increases vulnerability) sample cells.

The final data requirement is a consensus on configuration impact on decision rewards.

Configuration impact on decision rewards

Participants evaluate the impact that each configuration has on the baseline decision rewards from each state. For example, it is possible that a poor decision may have a less severe adverse impact if made from the most technologically advanced DDG, where the benefits of the weapons system can overcome the effects of the decision⁹. Once again, configuration impact consensus derives from a seven-point Likert scale ranging from “negligible impact” to “significant impact.”

To clarify the configuration impact requirements, I display a functional spreadsheet (Figure 4) to demonstrate configuration impact values.

⁹ An UNCLASSIFIED example of this is seen in configurations 6 and 7, both equipped with SPY-6 and SEWIP Block 3, which enable air-defense postures while maintaining anti-surface threat measures.

Configuration	Self-defense-Air	Self-defense-Missl 1	Self-defense-Missl 2	Area defense-Air	Area defense-Missl 1	Area defense-Missl 2	Wide defense-Air	Wide defense-Missl 1	Wide defense-Missl 2
1: SPY-1D, SEWIP Blk 2	7	-7							
2: SPY-1D, SEWIP Blk 3	Configuration 1 significantly increases the Self-defense-Air reward value.	Configuration 1 significantly decreases the Self-defense-missile 1 reward value.							
3: SPY-1D, SEWIP Blk 2, dLNA									
4: SPY-1D, SEWIP Blk 3, dLNA									
5: SPY-6, SEWIP Blk 2									
6: SPY-6, SEWIP Blk 3									
7: SPY-6 (Flt III), SEWIP Blk 3									

Figure 4. Configuration impact on rewards with two notional values. Excel's conditional formatting color-codes responses according to values, as seen in the green (7, significantly increases reward) and red (-7, significantly decreases reward) sample cells.

In addition to gathering SME and decision maker consensus data, I have gathered response data onto which I apply CA to calculate scores that enable complete TPM and contribution matrix population required for DP.

Correspondence Analysis

The previous four subsections describe data elicitation for state transition probabilities, decision impacts on transition probabilities, decision rewards, and configuration impacts on decision rewards, the last three of which produce matrices populated with 7-point Likert data. Figure 5 is a representative scale made available to participants to clarify their understanding of response values¹⁰.

¹⁰ While the verbiage for Likert scales change according to question context, the value descriptions are consistent with regards to significance.

Configuration Impact Value	Description
-7	Significant Adverse Impact on Vulnerability
-6	
-5	
-4	
-3	
-2	Moderate Adverse Impact On Vulnerability
-1	
0	Slight Adverse Impact on Vulnerability
1	
2	
3	Negligible (or nearly negligible) impact on Vulnerability
4	
5	
6	Slight Beneficial Impact on Vulnerability
7	
	Moderate Beneficial Impact on Vulnerability
	Significant Beneficial Impact on Vulnerability

Figure 5. Sample bi-directional 7-point Likert scales for participant reference. This scale resembles all scales used to populate three of the required data matrices necessary to parametrize the DP model.

According to (Powers 2016), analytic limitations exist in the practice of Likert response analysis when numerical values are treated at face value. A response of “4” is not necessarily twice as much as “2,” since Likert responses are ordinal in nature. Analytic practices such as averaging response values or reporting summary statistics fail to capture the quantitative nuance underlying respondent characteristics or variable relationships. Once again, these approaches assume equal distances between Likert anchor points. One way to compensate for these problems is to include questions during data elicitation that evaluate the relative distances between anchor points¹¹. However, the already demanding data elicitation process of this study would have overwhelmed

¹¹ For example, I could elicit that going from 2 to 3 is twice as impactful as going from 1 to 2, and so on.

participant had we included too many additional questions. This study applies CA to response distribution matrices to approximate quantitative values that parameterize the DP model.

With all SME data elicited, I summarize Likert response value distributions in a contingency table, seen in Table 1, below.

Table 1. Response distribution contingency table. This table is necessary for CA, and summarizes the number of times the absolute value of a response appears in the final data sets. For example, “7” or “-7” appears 14 times in all data associated with State 10.

Value Impact Contingency Table	0	1	2	3	4	5	6	7
Config1	2	4	3	0	0	0	0	0
Config2	2	3	4	0	0	0	0	0
Config3	0	0	3	6	0	0	0	0
Config4	0	0	1	5	3	0	0	0
Config5	0	9	0	0	0	0	0	0
Config6	0	0	0	0	0	9	0	0
Config7	0	0	0	0	0	0	0	9
State1	0	0	7	2	4	0	3	2
State2	0	0	7	2	4	0	3	2
State3	1	2	6	0	4	2	1	2
State4	0	0	3	0	3	6	0	6
State5	0	3	3	3	3	3	0	3
State6	3	0	0	0	6	0	3	6
State7	3	3	3	0	6	0	0	3
State8	3	0	2	0	3	3	1	6
State9	0	0	3	2	3	0	2	8
State10	0	0	1	0	3	0	0	14
State11	0	6	0	1	4	1	1	5
State12	1	3	1	0	3	4	0	6
State13	0	3	1	0	2	3	2	7
State14	0	0	8	0	2	1	0	7
State15	9	0	0	0	9	0	0	0

I apply the CA indicator score process described in (Powers 2016) and (Schramm and Powers 2017) to transform elicited Likert data into numerical TPM and contribution data required for DP. I describe CA results and effects on DP model parameters in Chapter Four – Solution Methodology.

Section Three – Dynamic Programming

The term dynamic programming (DP), introduced by Richard Bellman in 1950, is the recursive, iterative process of discovering optimal strategies for dynamic, sequential, and uncertain decision-making problems (Rust 2019). DP is a very efficient problem-solving method for small Markov Decision Processes (MDPs) with known transition probabilities. The Markov property of these processes states that the transitions and rewards depend only on the current state and current action regardless of previous states, actions, or rewards. Furthermore, the algorithm does not greedily maximize immediate reward, rather it maximizes reward over a time horizon. For clarity, I apply (Powell and Powell 2011) terminology and definitions when discussing Bellman's equations. I refer to DP equations as Bellman's equations, and they take various forms throughout their applications; deterministic, stochastic, policy iteration, and value iteration, to name a few. This study applies value iteration, a very efficient computational technique for calculating optimal expected value and policy.

The value iteration algorithm is simple to implement and naturally lends itself to problems like this study's future DDG configuration selection in that the necessary parameterization components are attainable, and it scales nicely to the problem size (Winston and Goldberg 2004). Given a set of states $(s_1, s_2, \dots, s_I) \in S$, actions $(a_1, a_2, \dots, a_I) \in A$, contributions $C(s_t, a_t)$, and action-based transition probabilities between states i and j $P_{ij}(a)$, value iteration seeks the optimal action in each state such that the objective function is maximized (or minimized) in the long run, thereby learning which action to take from each state. Value iteration resembles backwards DP in that it iterates until

satisfying convergence criterion ε by way of optimizing the average expected value g in the stochastic Bellman's equation for average criteria, Equation 1.

Equation 1. Stochastic Bellman's for Average Criteria

$$V_t(S_t) = \max_a [C(S_t, a_t) - g + \sum_{S(t+1)} P(S_t, a_t, S_{t+1}) V_{t+1}(S_{t+1})]$$

$V_t(S_t)$ is the value of being in state s at time t , which monotonously increases (for maximization problems) with iteration n until ε -convergence, defined by Equation 2:

Equation 2. Epsilon Convergence

$$V_{t+1}^n(S_{t+1}) - V_t^{n-1}(S_t) < \varepsilon$$

Value iteration learns the value of a state $V_t(S_t)$ as the algorithm iterates to convergence. Convergence is made possible by the inclusion of average criteria function g , the objective function of stochastic Bellman's for average criteria (Equation 1) that is maximized in this study. In this, my research provides decision makers with maximum DDG configuration utility in a manner that is consistent with their own beliefs regarding configuration operational contribution and states. In addition to returning g , stochastic Bellman's (Equation 1) also returns optimal policy vector R^* , which defines the action for each state that yields optimal g . Equation 3 is the discounted criteria version of stochastic Bellman's equation.

Equation 3. Stochastic Bellman's for Discounted Criteria

$$V_t(S_t) = \max_a [C(S_t, a_t) + \beta \sum_{S(t+1)} P(S_t, a_t, S_{t+1}) V_{t+1}(S_{t+1})]$$

Equation 3 requires discount parameter $0 < \beta < 1$ for convergence, representing the time value of money. Since we are maximizing the average *utility* value of a configuration in this problem, β is not an appropriate parameter. However, a β -like parameter (known as γ) appears in the Q-learning equation and will be discussed in Section Four – Approximate Dynamic Programming.

DDG Configurations

This model assesses seven candidate configurations, D_i , for future investment consideration. Each D_i possesses its own set of strengths and weaknesses in terms of electronic warfare and situational awareness capabilities. This model ultimately calculates g_i , the utility value for configuration $D_i \forall i$. These configurations are as follows:

- Configuration 1: SPY-1D, SEWIP Block 2 (Current configuration)
- Configuration 2: SPY-1D, SEWIP Block 3
- Configuration 3: SPY-1D, SEWIP Block 2, dLNA
- Configuration 4: SPY-1D, SEWIP Block 3, dLNA
- Configuration 5: SPY-6, SEWIP Block 2
- Configuration 6: SPY-6, SEWIP Block 3

- Configuration 7: SPY-6, Flight III (A and B), SEWIP Block 3

DDG States

DDG state S_t is defined by combinations of state-space subset s_k , $k \in \{1, 2, 3, 4, 5\}$ with $\dim(k) = \{3, 2, 2, 4, 4\}$, initially yielding 192 possible states S_t . Specifically, state S_t is a combination of the following state descriptions:

- DDG role (3)
 - Carrier escort
 - LHA escort
 - SAG unit
- Communication capability (2)
 - Reliable OTH comms
 - Unreliable OTH comms
- Situation awareness (2)
 - With advanced SA
 - Without advanced SA
- Weapon capability (4)
 - Weapon capability 1¹²
 - Weapon capability 2
 - Combined weapon capability
 - No weapon capability
- Threat vulnerability (4)

¹² For security purposes, specific weapon capability descriptions are withheld.

- Surface threat vulnerability¹³
- Missile threat vulnerability
- Combined threat vulnerability
- No threat vulnerability

DDG Actions

The unique concept of $V(S_t)$ in DP motivates the approach to evaluate utility g_i for each configuration D_i over the lifetime of its operational service, providing insight to complement assessments regarding configuration survivability, offensive capability, and other strategic metrics. As such, this model calculates value over a set of nine actions A that broadly represent the range of postures that configuration D_i may employ in an adversarial environment. These actions are as follows:

- Self-defense – Air
 - DDG defends itself against airborne threats.
- Self-defense – Missile 1
 - DDG defends itself against missile threats identified as category 1¹⁴.
- Self-defense – Missile 2
 - DDG defends itself against missile threats identified as category 2.
- Area defense – Air
 - DDG defends ally assets within an assigned radius against airborne threats.

¹³ For security purposes, specific threat vulnerability descriptions are withheld.

¹⁴ For security purposes, specific missile type descriptions are withheld.

- Area defense – Missile 1
 - DDG defends ally assets within an assigned radius against category 1 missile threats.
- Area defense – Missile 2
 - DDG defends ally assets within an assigned radius against category 2 missile threats.
- Wide defense – Air
 - DDG defends assets, locations within assigned geographic area against airborne threats.
- Wide defense – Missile 1
 - DDG defends assets, locations within assigned geographic area against category 1 missile threats.
- Wide defense – Missile 2
 - DDG defends assets, locations within assigned geographic area against category 2 missile threats.

Action Contributions

Action contribution, captured by contribution matrix $C_i(S_t, a_t)$, is the benefit to DDG counter-vulnerability if a DDG with configuration D_i takes action a_t when in state S_t at time t . $C_i(S_t, a_t)$ is populated with campaign modeling data, published studies, and SME input as described in Section Two – Data Elicitation.

Transition Probabilities

DDG transition probabilities are captured in the form of transition probability matrices (TPMs), where a TPM exists for each action a_t taken in state S_t at time t , yielding $P_i(S_t, a_t, S_{t+1})$. TPMs are populated by SME input and campaign model outputs, and each D_i has its own TPM set. Once again, TPM data elicitation is described in Section Two – Data Elicitation.

Section Four – Approximate Dynamic Programming

Reinforcement learning (RL), also known as Q-learning, is an AI method where agents learn through trial and error and continuous engagement with its environment. That is, starting from a specific state and performing an action, the agent then transitions to a new state while being rewarded. In this, the Q-learning learning version of Bellman's equation asynchronously interacts with a dynamic environment through randomly structured observation and reward with the objective of maximizing the reward. The same state-space S , actions A , and contribution (or reward) function C from Equation 1 parameterize the Q-learning model. The algorithm iteratively observes some $s_k \in S$ and selects action $a \in A$, from which it receives reward $c \in C$. Algorithm behavior and performance is stored in Q-matrix $Q(s,a)$, which holds the expected reward for each possible action a taken from state s . As $Q(s,a)$ achieves band-convergence, the learning version of Bellman's approximates the maximum expected reward and the near-optimal policy R^* that yields the approximate maximum reward.

$Q(s,a)$ acts as a dynamic lookup table to replace TPMs by seeking Q-factors q for each action taken from each state. Instead of TPMs, Q-learning requires knowledge of

inter-state reachability so that it knows which states can be visited with each iteration. In this, the learning version of Bellman's equation defines the value of each state, $V(s)$, according to Equation 4:

Equation 4. Q-learning $V(s)$

$$V(s) = \max_a Q(s, a)$$

If $Q(s, a)$ initializes to 0.0, it is clear that $V(s) = 0.0$ across all actions upon initialization.

At iteration n , $Q(s, a)$ populates with \widehat{q}^n approximated by Equation 5:

Equation 5. Iteration n q-approximation

$$\widehat{q}^n = \hat{C}(s^n, a^n) + \gamma \max_{a' \in A} \overline{Q}^{n-1}(s^{n+1}, a')$$

Note that in Equation 5, discount factor γ ensures convergence in a manner like β in Equation 3, calculating updates to $Q(s, a)$. These updates enable learning of near-optimal action a from each state s by defining a at iteration n with Equation 6:

Equation 6. Iteration n Action Evaluation

$$a^n = \arg \max_{a \in A} \overline{Q}^{n-1}(s^n, a)$$

Q-learning uses the estimates $\overline{Q}^{n-1}(s^n, a)$ from iteration $n-1$ at iteration n . With \widehat{q}^n , Q-factors in $Q(s, a)$ update via the learning version of Bellman's, Equation 7:

Equation 7. Learning Version of Bellman's Equation

$$\overline{Q}^n(s^n, a^n) = (1 - \alpha_{n-1})\overline{Q}^{n-1}(s^n, a^n) + \alpha_{n-1}\widehat{q}^n$$

From Equation 7, total reward $\Sigma V(s)$, and near-optimal policy vector R^* are derived. I am interested in comparing near-optimal Q-learning and optimal DP policies with varying γ (from Equation 5). Varying γ effectively compares myopic (γ close to 0.0) and future-seeking (γ close to 1.0) reward policies.

CHAPTER FOUR – SOLUTION METHODOLOGY

I apply DP value iteration to the stochastic Bellman's equation for average criteria (Equation 1) for each proposed DDG configuration, thereby evaluating maximum expected utility for each configuration. The DP model is parametrized by SMEs and decision makers, ensuring consistency with their beliefs regarding configuration contribution. I go on to assign the maximum expected utility values as objective function coefficients in a utility-maximizing MIP model constrained by budgetary and logistic requirements. This chapter describes DP model parameters and data structures derived from data elicitation, all in the specific context of the dynamic DDG operating environment. I apply one of the most common DP algorithm, value iteration, to solve the MDPs representing the operating environment so that I can solve for maximum utility as the algorithm converges (Rust 2019). This method's propriety is validated in (Watkins 1989), which argues that any continuous process, such as DDG operations, can be adequately approximated by an MDP. Next, I define the MIP model objective function and constraints. I end with a description of the ADP Q-learning model that I include for comparative and academic purposes.

Section One – Correspondence Analysis

I describe data elicitation and resultant contingency table (Table 1) in Chapter Three – Research Problem and Model Description. Table 1 is necessary to apply CA, a method used in (Powers 2016) and (Schramm and Powers 2017) to derive numerical values from ordinal Likert data. CA accounts for the fact that, despite being represented

with integers, Likert responses are not numerical values. Likert scales are popular methods for data elicitation, as they are common and familiar to respondents. Had more time been available, I would have enriched my questions by eliciting the inequality or non-linearity between Likert anchors, but time is not a luxury in this study. CA measures the distance between anchor points to extrapolate useful Likert-Numerical mappings, shown in Table 2.

Table 2. Likert-Numerical mappings. Note that differences between anchor points are not equal.

Likert	Numerical
0	0.000
1	0.090
2	0.272
3	0.421
4	0.599
5	0.768
6	0.940
7	1.000

The numerical values in Table 2 are proportions of the entire response scale, making them useful for modifying TPMs and contribution matrices according to respondent data, and generating the data necessary for DP parameterization.

Section Two – Dynamic Programming

Recall Equation 1. Stochastic Bellman's for Average Criteria:

$$V_t(S_t) = \max_a [C(S_t, a_t) - g + \sum_{S_{t+1}} P(S_t, a_t, S_{t+1}) V_{t+1}(S_{t+1})]$$

I evaluate the Equation 1 objective function g_i for each DDG configuration D_i , providing me with expected utility for each configuration according to the underlying beliefs captured during the elicitation process. This model manifests in the MDP toolbox package (*MDPtoolbox*) in R via the MDP relative value iteration function (*mdp_relative_value_iteration*). This function solves Equation 1 with function inputs $P_i(S_t, a_t, S_{t+1})$, $C_i(S_t, a_t)$, (optional) ϵ , and (optional) maximum iterations should the function fail to converge to ϵ . Function outputs are the optimal policy vector R_i^* , and optimal utility value g_i . I compare Equation 1 R_i^* to the R_i^* approximated by Equation 7, the learning version of Bellman's equation, which I discuss in Section Five – Approximate Dynamic Programming. To solve the R3B problem of maximizing DDG configuration utility, I treat each g_i as objective function coefficients in a utility-maximizing MIP model constrained by budgetary and logistic requirements. The next section describes $P_i(S_t, a_t, S_{t+1})$ and $C_i(S_t, a_t)$ development, as required for functional computation.

Section Three – Value Iteration Model Inputs

This section describes the parametrization and output for Equation 1 for each D_i . This function solves stochastic Bellman's equation for average criteria with inputs $P_i(S_t, a_t, S_{t+1})$, $C_i(S_t, a_t)$, and convergence criterion ϵ . Once again, outputs are the optimal policy vector R_i^* , and optimal utility value g_i . The following sections describe model inputs $C_i(S_t, a_t)$, and $P_i(S_t, a_t, S_{t+1})$, using configuration D_I as a concrete example.

Contribution Matrix $C_0(S_t, a_t)$

Table 3 is the configuration-agnostic contribution matrix $C_0(S_t, a_t)$ that evaluates the impact of decisions when made by a DDG in state S_t at time t , regardless of configuration.

Table 3. Configuration-agnostic contribution matrix. These SME, model, and published study – based values represent impact on DDG threat vulnerability.

State/Action	Self-defense Air	Self-defense Missile 1	Self-defense Missile 2	Area defense Air	Area defense Missile 1	Area defense Missile 2	Wide defense Air	Wide defense Missile 1	Wide defense Missile 2
Carrier Escort	2	2	3	4	4	7	2	2	3
LHA Escort	2	2	3	4	4	7	2	2	3
SAG Unit	4	4	7	5	5	6	1	1	4
Reliable OTH Comms	4	4	4	5	5	5	7	7	7
Unreliable OTH Comms	4	4	4	3	3	3	1	1	1
With Advanced SA	4	4	4	7	7	7	6	6	6
Without Advanced SA	4	4	4	2	2	2	1	1	1
Weapon Capability 1	4	5	6	4	7	7	4	7	7
Weapon Capability 2	4	7	7	4	6	6	4	3	3
Combined Weapon Capability	4	7	7	4	7	7	4	7	7
No Weapon Capability	4	1	1	4	1	1	4	1	1
Surface Threat Vulnerability	7	1	5	7	1	5	7	1	5
Missile Threat Vulnerability	1	7	5	1	7	5	1	7	5
Combined Threat Vulnerability	2	2	7	2	2	7	2	2	7
No Threat Vulnerability	4	4	4	4	4	4	4	4	4

For Table 3 illustration, a DDG in a missile threat vulnerability state that takes self-defense air action has a severe adverse impact on threat vulnerability, earning that

decision the lowest value of 1. Simply put, a self-defense posture against an airborne threat is a poor decision when faced with an incoming missile.

Configuration Impact on $C_0(S_i, a_t)$

Table 4 is the impact that configuration D_i has on $C_0(S_i, a_t)$ (Table 3, above).

Table 4. Configuration impact on $C_0(S_i, a_t)$.

Configuration/Action	Self-defense Air	Self-defense Missile 1	Self- defense Missile 2	Area defense Air	Area defense Missile 1	Area defense Missile 2	Wide defense Air	Wide defense Missile 1	Wide defense Missile 2
D₁	0	-0.09	-0.09	0	-0.09	-0.09	-0.272	-0.272	-0.272
D₂	0.272	0.09	0.09	0.09	0	0	-0.272	-0.272	-0.272
D₃	0.272	0.421	0.421	0.272	0.421	0.421	0.272	0.421	0.421
D₄	0.599	0.599	0.599	0.421	0.421	0.421	0.272	0.421	0.421
D₅	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
D₆	0.768	0.768	0.768	0.768	0.768	0.768	0.768	0.768	0.768
D₇	1	1	1	1	1	1	1	1	1

Table 4 values are CA Likert-Numerical mappings (from Table 2) of SME-elicited responses regarding how different configurations affect the Table 3 contribution matrix. For illustration, configuration D_7 enjoys maximum benefit to threat vulnerability across the action spectrum, with the maximum CA-based impact value of 1 for each action taken. Simply put, if a DDG equipped with configuration 7 were to make a poor decision from a given state, it would not suffer as much vulnerability increase as a lesser-configured DDG. Table 4 combines with Likert-numerical mapped Table 3 values to yield $C_i(S_i, a_t)$ for each D_i to parameterize Equation 1.

Configuration 1 (D_1) Transition Probability Matrix $P_I'(S_t, a_t, S_{t+1})$

In order to calculate a TPM for each action (and for each configuration), I begin by constructing an action-agnostic TPM $P_i'(S_t, a_t, S_{t+1}) \forall i \in \{1, \dots, 7\}$, the TPM for D_i when $a_t = \text{NULL}$; that is, prior to incorporating action impact on transition probabilities.

Table 5 displays the state-space subset s_k matrices that multiply to produce $P_I'(S_t, a_t, S_{t+1})$.¹⁵

Table 5. $P_I'(S_t, a_t, S_{t+1})$ state-space subset s_k matrices.

	Carrier Escort	LHA Escort	SAG Unit	
Carrier Escort	0.7	0.1	0.2	
LHA Escort	0.1	0.7	0.2	
SAG Unit	0.3	0.3	0.4	
	Reliable OTH Comms	Unreliable OTH Comms		
Reliable OTH Comms	0.7	0.3		
Unreliable OTH Comms	0.3	0.7		
	With Advanced SA	Without Advanced SA		
With Advanced SA	0	1		
Without Advanced SA	0	1		
	Weapon Capability 1	Weapon Capability 2	Combined Weapon Capability	No Weapon Capability
Weapon Capability 1	0.8	0	0.2	0
Weapon Capability 2	0	0.8	0.2	0
Combined Weapon Capability	0.1	0.1	0.8	0
No Weapon Capability	0.3	0.3	0.4	0
	Surface Threat Vulnerability	Missile Threat Vulnerability	Combined Threat Vulnerability	No Threat Vulnerability
Surface Threat Vulnerability	0.6	0.1	0.3	0
Missile Threat Vulnerability	0.1	0.6	0.3	0
Combined Threat Vulnerability	0.2	0.2	0.6	0
No Threat Vulnerability	0.4	0.3	0.3	0

¹⁵ TPM subset matrices multiply under the independence assumption to produce TPM $P_i'(S_t, a_t, S_{t+1})$ for each D_i .

For Table 5 illustration, a sample state S_t (*carrier escort; reliable OTH comms; without advanced SA; weapon capability 1; surface threat vulnerability*) transitions to the same state when $a_t = \text{NULL}$ with probability $0.7*0.7*1*0.8*0.6 = 0.2352$.

D_1 and D_2 lack of advanced SA capabilities results in $P_1'(S_t, a_t, S_{t+1}) = P_2'(S_t, a_t, S_{t+1}) = 0 \forall S_{t+1}$ involving $s_3 = \text{with advanced SA}$ entries.

$P_i'(S_t, a_t, S_{t+1}) = 0 \forall i \in \{1, \dots, 7\}, S_{t+1}$ involving $s_4 = \text{no weapon capability}$ entries and $s_5 = \text{no threat vulnerability}$ entries, effectively reducing the state-space from 192 states to 108 states.¹⁶ This state-space reduction makes Q-learning an attractive alternative to value iteration, and it will be discussed in a later section.

$P_i'(S_t, a_t, S_{t+1})$ maps to $P_i(S_t, a_t, S_{t+1}) \forall a \in A$ by combining with SME-elicited action impacts on transition probabilities. Table 6 shows the difference in $P_1'(S_t, a_t, S_{t+1})$ when action a is taken at time t .

¹⁶ Several data elicitation rounds yielded transition probabilities of 0 for these entries because SMEs agreed that no DDG would remain in a state of *no weapon capability* after $t = 1$ days, and that once a DDG enters a state of vulnerability, it will never enter a state of *no threat vulnerability* unless the DDG is transiting or departing the operational environment.

Table 6. Difference in $P_I'(S_t, a_t, S_{t+1})$ when action a is taken at time t .

State/Action	1	2	3	4	5	6	7	8	9
Carrier Escort	-0.094	-0.094	-0.1	-0.0272	-0.0272	0	0	0	0
LHA Escort	-0.094	-0.094	-0.1	-0.0272	-0.0272	0	0	0	0
SAG Unit	0.188	0.188	0.2	0.0544	0.0544	0	0	0	0
Reliable OTH Comms	0.0816	0.0816	0.0816	0.2304	0.2304	0.2304	0.3	0.3	0.3
Unreliable OTH Comms	-0.0816	-0.0816	-0.0816	-0.2304	-0.2304	-0.2304	-0.3	-0.3	-0.3
With Advanced SA	0	0	0	0.599	0.599	0.599	1	1	1
Without Advanced SA	0	0	0	-0.599	-0.599	-0.599	-1	-1	-1
Surface Threat Vulnerability	-0.1	0.05	0	-0.1	0.05	0	-0.1	0.0599	0
Missile Threat Vulnerability	0.05	-0.1	0	0.05	-0.1	0	0.0499	-0.11	-0.0544
Combined Threat Vulnerability	0.05	0.05	0	0.05	0.05	0	0.0501	0.0501	0.0544

For Table 6 illustration, if the probability of transitioning from a *carrier escort* state to a *carrier escort* state is 0.7 when $a_t = \text{NULL}$, but that probability changes to $0.7 - 0.094 = .606$ if $a_t = 1$. In this, I generate a unique TPM $\forall a \in A$. However, since unique TPMs must exist for each D_i , calculating $P_i'(S_t, a_t, S_{t+1})$ differences when action a is taken at time t is a non-trivial effort that combines decision impact SME-elicited data (Figure 2) with each TPM_i , while taking into account how much delta is possible without violating the rules of TPM probabilities. For example, suppose $P_I(S_t', a_t, S_{t+1}) = 0.7$ for some state S_t' , and the SME-elicited a_t impact = 6 \rightarrow 0.94 (from Table 2). This yields a $P_I(S_t', a_t, S_{t+1})$ increase from 0.7 to $0.7 + (1.0 - 0.7) * 0.94 = 0.982$. Furthermore, since $\sum_a P_i'(S_t, a_t, S_{t+1}) = 1 \forall i$, the increase in $P_I(S_t', a_t, S_{t+1})$ must be accompanied by a decrease in $P_I(S_t'', a_t, S_{t+1})$ for some $S_t'' \neq S_t'$. If this condition is not met, the change in $P_I(S_t', a_t, S_{t+1})$ is infeasible, and the value remains the same.

Section Four – Mixed Integer Programming Model

The objective of the MIP model is to maximize utility by assigning the optimal DDG-configuration pairings while obeying budgetary, requirement, and logistic constraints. I accomplish this via the following MIP formulation:

Equation 8. MIP Formulation

$$\max_{h,i} \sum_{h=1}^{95} \sum_{i=1}^7 g_i X_{h,i}$$

S.T.

Constraint 1. Budget

$$\sum_{h=1}^{95} \sum_{i=1}^7 c_{h,i} X_{h,i} \leq Budget$$

Constraint 2. One configuration per DDG

$$\sum_{i=1}^7 X_{h,i} = 1; \forall h$$

Constraint 3. Configuration Requirements

$$\sum_{h=1}^{95} X_{h,i} \geq Req_i; \forall i$$

Constraint 4. System Requirements

$$\sum_{h=1}^{95} \sum_{l \in \{L_k\}} X_{h,l} \geq Sys_k \forall k \in \{1 \dots 6\}$$

Constraint 5. Binary Decision Variable

$$X_{h,i} \in \{0, 1\} \forall h, i$$

Where:

g_i is the utility of configuration i , calculated by Equation 1.

$X_{h,i}$ is the decision to equip DDG h with configuration i .
 $C_{h,i}$ is the cost of equipping DDG h with configuration i .
 $Budget$ is the projected total budget for the DDG modernization program.
 Req_i are minimum DDG fleet requirements for configuration i .
 L_k are which configurations are comprised of system k .
 Sys_k are minimum DDG fleet requirements for system k .

The Equation 1 objective function maximizes the sum of configuration utility (g_i) multiplied by the decision to equip a DDG with that configuration ($X_{h,i}$). Constraint 1 ensures that the total cost of equipping the DDG fleet with decided configurations falls below a projected DDG modernization program budget. Constraint 2 ensures that only one configuration is assigned to each DDG. Constraint 3 ensures that assigned configurations satisfy minimum DDG fleet configuration requirements. Constraint 4 ensures that the systems that comprise assigned configurations satisfy minimum DDG fleet system requirements.¹⁷

Section Five – Approximate Dynamic Programming

I compare DP R^* results from Equation 1 to the model-free APD dialect, Q-learning R^* results. Q-learning learns near-optimal behavior through random interactions within a dynamic environment where the algorithm receives reward-based feedback to evaluate its performance. Q-learning differentiates from supervised methods in that there is no specific instruction on how to improve behavior. Unlike DP methods such as Equation 1, Q-learning deals with unknown TPMs in an unstructured state-space with a *learning* version of Bellman's equation that converge to a near-optimal band objective

¹⁷ My code pre-processes Constraint 3 and Constraint 4 so that if configuration requirements fulfill system requirements, the associated system requirement constraint is removed.

vice a point-value. Q-learning is an attractive alternative to DP for this study because of the relatively small state-space, the availability of sufficient computing power, and the lack of TPM requirement. The availability of increased computing power has enabled a comeback, so to speak, in Q-learning popularity.

Lack of Q-learning TPMs aside, I parameterize my Q-learning model with the same as those from my DP model. $S^n = S_t$, actions A remain, and $C_i(S^n, a^n) = C_i(S_t, a_t)$. As seen in Equation 5, $Q(s,a)$ convergence requires discount parameter γ , which is the value of future rewards. Setting γ close to 1.0 treats a future reward as if it were a current reward, whereas setting γ close to 0.0 prefers immediate reward and a short-sighted near-optimal policy. This motivates discussion regarding γ_i as it relates to expected configuration D_i lifetime or continuous D_i exposure to S^n during D_i lifetime. Should commanders wish to position their configured DDGs in such a way to maximize future requirement utility, γ should be relatively high. With no a priori knowledge of such a parameter, I experiment with model sensitivity to $0.0 < \gamma < 1.0$.

I use the *ReinforcementLearning* package and function in R for Q-learning evaluation. *ReinforcementLearning* attributes its theoretical and technical details to (Sutton and Barto 2018), paying particular attention to practical Q-learning applications. For example, the authors recognize that in the case of common, real-world problems, a constant learning rate α (Equation 7) is desirable for rapid band-convergence. While I parametrize my Q-learning model with the same as Equation 1, *ReinforcementLearning* requires a data frame input in a state/action/reward/next-state format to represent the stochastic environment with inter-state reachability. This is a trivial matter of re-

formatting existing data. With the DP and ADP models parameterized, I compare both DP and ADP R^* vectors to explore the effect of varying future reward values.

Section Six – Q-Learning Model Inputs

I initialize $Q(s, a)$ to 0.0, assuming inter-state reachability between all S^n , which yields 104,976 (S^n, a^n, S^{n+1}) combinations, hereby referred to as the *environment*.¹⁸ I limit Q-learning discussion to configurations D_4 and D_5 , a decision stemming from SME disagreement regarding unexpected g_4 and g_5 DP results¹⁹. As mentioned above, $S^n = S_t$, actions A remain, and $C_i(S^n, a^n) = C_i(S_t, a_t)$.

I set Equation 7 learning rate parameter α to 0.1, avoiding Q-factor *apparent* convergence (since $\alpha > 0.0$) without learning too quickly (since $\alpha < 0.99$).

As seen in Equation 5, $Q(s, a)$ convergence requires discount parameter γ , which is the value of future rewards. Recall that setting γ close to 1.0 values a future reward while setting γ close to 0.0 values immediate reward. As previously discussed, I experiment with model sensitivity to $0.0 \leq \gamma \leq 1.0$.

¹⁸ For clarification, I replace time-subscript t from my DP model with iteration-superscript n for the ADP model.

¹⁹ g_5 was expected to be much greater than g_4 , which it was not, prompting a lengthy discussion among SMEs as to the TPM validity for D_5 and D_4 .

CHAPTER FIVE – ANALYSIS AND EXPERIMENTATION

This chapter describes the g_i and R_i^* results from Equation 1. Stochastic Bellman's for Average Criteria, for each $P_i(S_t, a_t, S_{t+1})$ and $C_i(S_t, a_t)$, as discussed in Chapter Four – Solution Methodology. Recall that g_i is the maximum expected utility for DDG_i ; a DDG that has been equipped with configuration $i \in \{1, \dots, 7\}$, and is the objective of Equation 1. R_i^* is the optimal policy, heretofore referred to as strategy, to achieve g_i .

Next, I discuss the MIP model with $\{g_1, \dots, g_7\}$ as coefficients in a utility-maximizing objective function, with budgetary and logistic constraints as described in Chapter Four – Solution Methodology. The MIP results display how many of each configuration can be equipped to maximize DDG fleet utility. I experiment with various budget constraint values to account for uncertainty in expected DDG program funding, and I demonstrate that eventually, no amount of additional funding can improve DDG fleet expected utility.

I conclude this chapter with a comparison of DP and ADP (Q-Learning) optimal/near-optimal utility and optimal/near-optimal strategy results, while experimenting with the Q-Learning discount factor γ . I include ADP analysis to confirm unexpected DP utility value results between technologically dissimilar DDG configurations. ADP maximal values agree with DP utility scores in that configuration 4 achieves greater utility than configuration 5. Throughout the DP analysis process, SMEs and DMs were surprised by configuration 4 superiority over configuration 5, and

iteratively adjusted TPM probabilities to observe new outcomes. The model-free (TPM-free) ADP analysis demonstrates configuration 4 superiority without knowledge of specific transition probabilities, thereby satisfying R3B SMEs and DMs. I vary the Q-Learning discount factor γ to observe where optimal DP policies fall on the Q-Learning myopic \rightarrow future reward seeking scale.

Section One – Dynamic Programming Results

This section displays the g_i and R_i^* results from Equation 1, for each configuration $i \in \{1, \dots, 7\}$. I display g_i results on a normalized $1 \rightarrow 10$ scale, making them easily explainable to R3B DMs. I display R_i^* results as a histogram of action distributions, as opposed to displaying the entire state-action vector. This presentation method is easily understandable by the DMs in the R3B study presentation, and it demonstrates how different configurations achieve maximum utility through different employment strategies.

Recall the DDG states (and sub-states):

- DDG role (3)
 - Carrier escort
 - LHA escort
 - SAG unit
- Communication capability (2)
 - Reliable OTH comms
 - Unreliable OTH comms
- Situation awareness (2)

- With advanced SA
- Without advanced SA
- Weapon capability (4)
 - Weapon capability 1²⁰
 - Weapon capability 2
 - Combined weapon capability
 - No weapon capability
- Threat vulnerability (4)
 - Surface threat vulnerability²¹
 - Missile threat vulnerability
 - Combined threat vulnerability
 - No threat vulnerability

Recall the DDG actions:

- Self-defense – Air
- Self-defense – Missile 1
- Self-defense – Missile 2
- Area defense – Air
- Area defense – Missile 1
- Area defense – Missile 2
- Wide defense – Air

²⁰ For security purposes, specific weapon capability descriptions are withheld.

²¹ For security purposes, specific threat vulnerability descriptions are withheld.

- Wide defense – Missile 1
- Wide defense – Missile 2

Figure 6 plots normalized $g_1 \dots g_7$ values for their respective configurations.

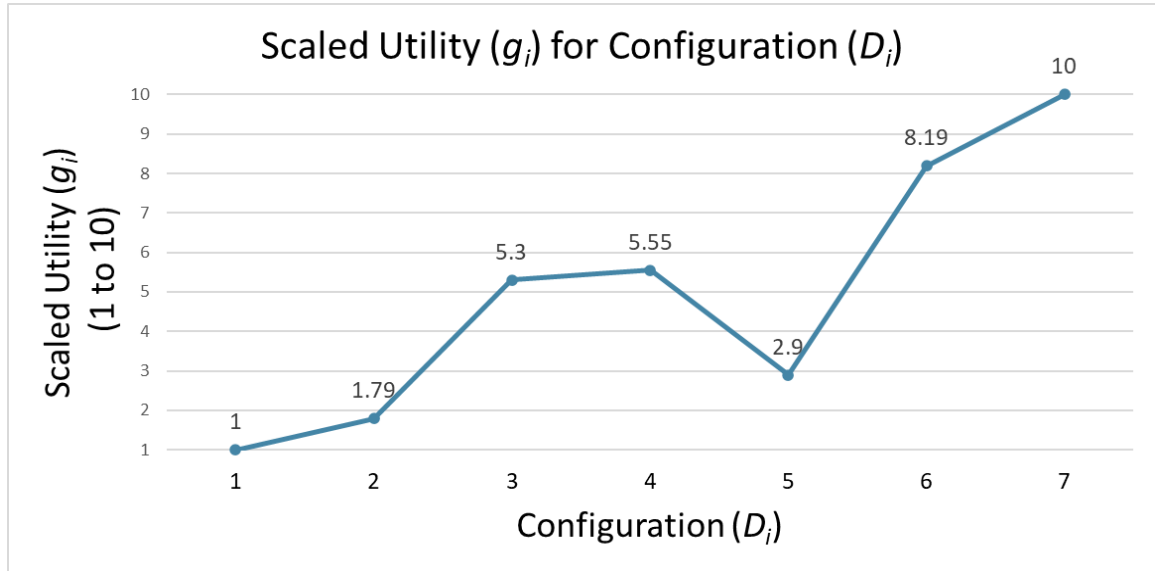


Figure 6. Scaled utility (g_i) for each configuration (D_i).

Note that Figure 6 *does not* confirm the expected rise in utility for each configuration, with a drop between D_4 and D_5 . This drop in utility motivates contextual discussions regarding effective system designs, and how human DMs and SMEs have difficulty understanding the effect of algorithmic decision-making consistency. I discuss this phenomenon in Chapter Two – Literature Review; Section Two – Decision Making. The debate surrounding the validity of these two DP results inspire my ADP evaluation approach wherein I solve Q-learning models for D_4 and D_5 . I recall configuration descriptions in the following configuration sections.

Configuration 1

DDG configuration 1 (D_1) is the SPY-1D, SEWIP Block 2, and is the current configuration. As expected, D_1 achieves the lowest utility value with $g_I = 1.0$. This is not surprising, because D_1 is the configuration upon which the most offensive and defensive improvements apply.

Figure 7 displays R_I^* results; the optimal strategy for D_1 to achieve g_I .

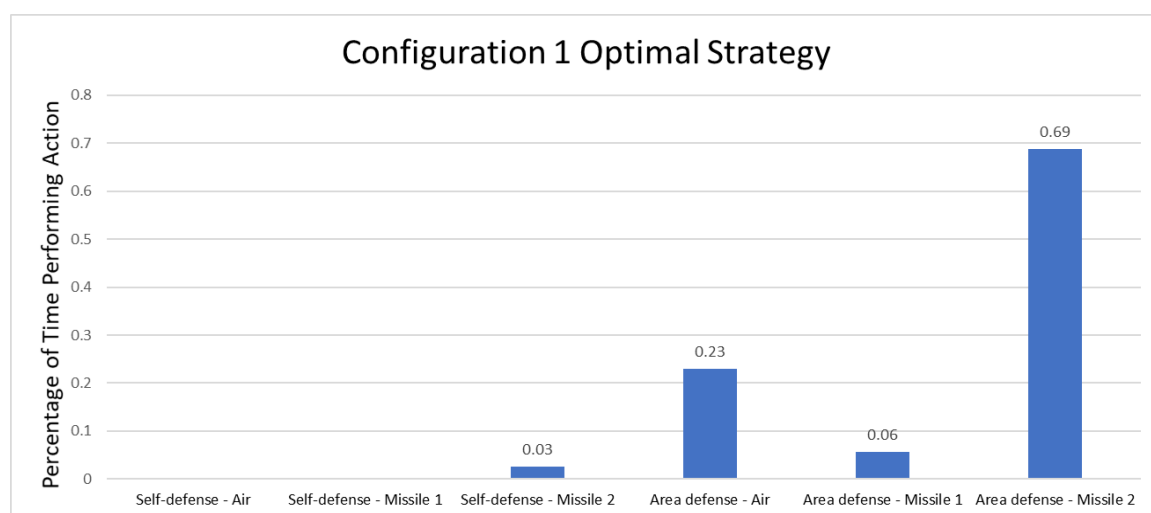


Figure 7. Configuration 1 optimal strategy R_I^* distribution.²²

D_1 DDGs are utilized the most for area defense against category 2 missile threats and for area defense in general among all DDG configurations. Area defense is defined as defending ally assets within an assigned radius against threats. D_1 DDGs maintain relatively high contribution values when area defense actions are taken, particularly

²² Display distribution sums do not equal exactly 1.0 due to rounding error.

against missile 2. SMEs agree that D_1 DDGs are less effective in self-defense and wide defense postures, meaning that area defense actions are most likely to yield maximum utility.

Configuration 2

DDG configuration 2 (D_2) is the SPY-1D, SEWIP Block 3, an improvement over the current SEWIP Block 2, adding electronic attack to the existing SLQ-32(V) electronic warfare system. As such, D_2 achieves an improved utility value of $g_2 = 1.79$, demonstrating benefit to the improved SEWIP system. Figure 8 displays R_2^* results; the optimal strategy for D_2 to achieve g_2 .

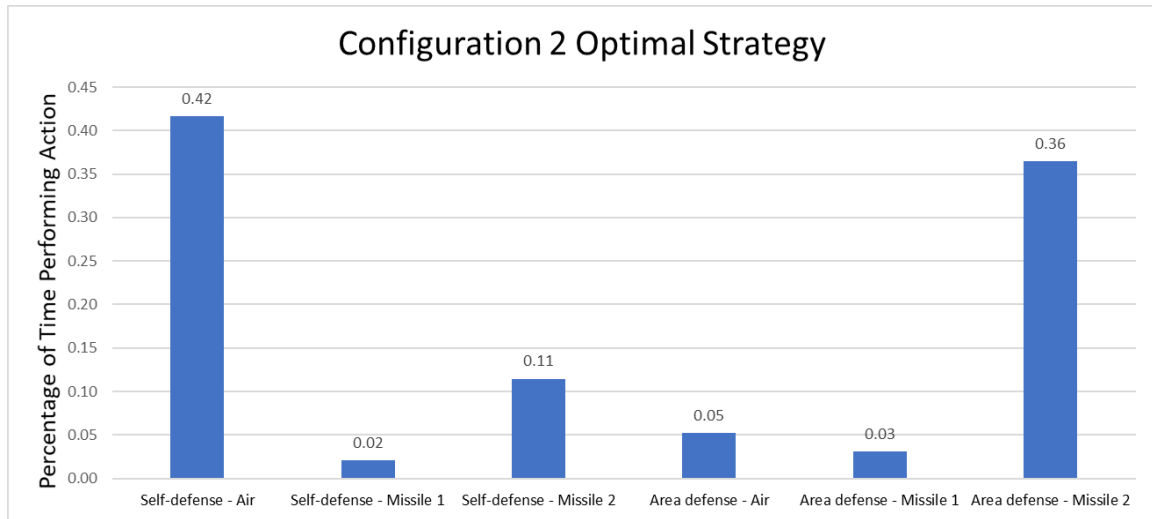


Figure 8. Configuration 2 optimal strategy R_2^* distribution.

Note that the improvement from SEWIP Block 2 to SEWIP Block 3 yields an improved utility score and an action distribution shift that enables more self-defense capability against missile 2 and, to a larger degree, air attacks. From an action

distribution perspective, D_2 DDGs are less likely than D_1 DDGs to be deployed in an area defense posture.

Configuration 3

DDG configuration 3 (D_3) is the SPY-1D, SEWIP Block 2, dLNA. While D_3 regresses back to the SEWIP Block 2, it complements its EW system with the Digital Low Noise Amplifier (dLNA) technology to improve SPY-1D radar sensitivity and capability. The D_3 utility value of $g_3 = 5.30$ indicates dramatic capability contributions from dLNA, even in conjunction with legacy, current EW systems. Figure 9 displays R_3^* results; the optimal strategy for D_3 to achieve g_3 .

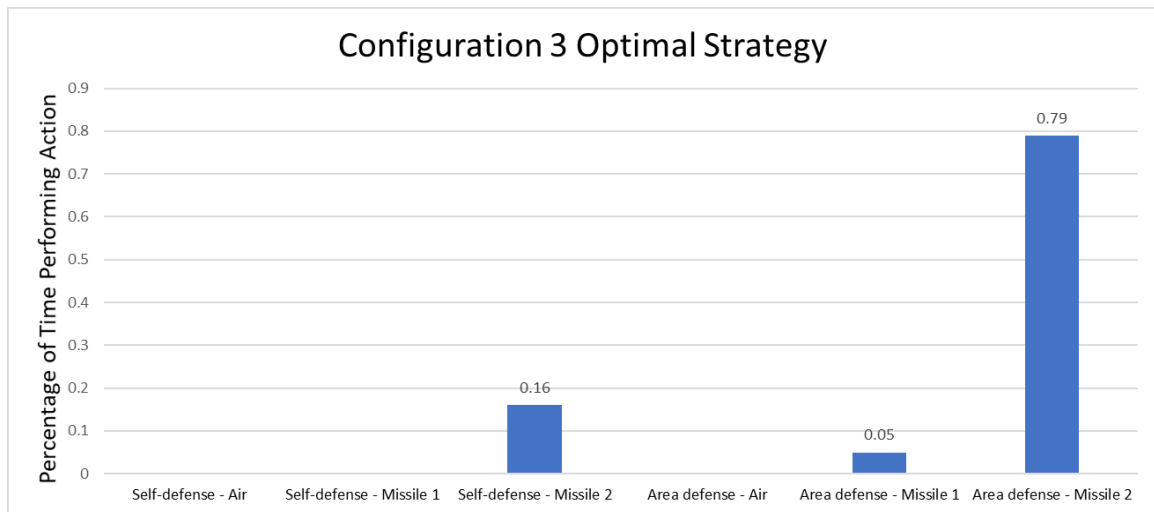


Figure 9. Configuration 3 optimal strategy R_3^* distribution.

Incorporating dLNA improves radar capability and sensitivity so that the optimal strategy includes more self-defense actions against missile 2, without sacrificing area defense postures against missile 2. The dLNA advantages enable self-defense actions

without increasing vulnerability to air threats, which explains why self/area defense actions against air threats do not exist in the optimal strategy.

Configuration 4

DDG configuration 4 (D_4) is the SPY-1D, SEWIP Block 3, dLNA. D_4 improves from the D_3 SEWIP Block 2 to the SEWIP Block 3, adding electronic attack to the existing SLQ-32(V) electronic warfare system. The D_4 utility value of $g_4 = 5.55$ is a slight improvement over $g_3 = 5.30$, suggesting that dLNA accounts for much of the utility gain when compared to D_1 and D_2 , neither of which configurations include dLNA. However, the shift in R_4^* action distributions on display in Figure 10 highlights the effects of SEWIP Block 3 inclusion onto D_4 .

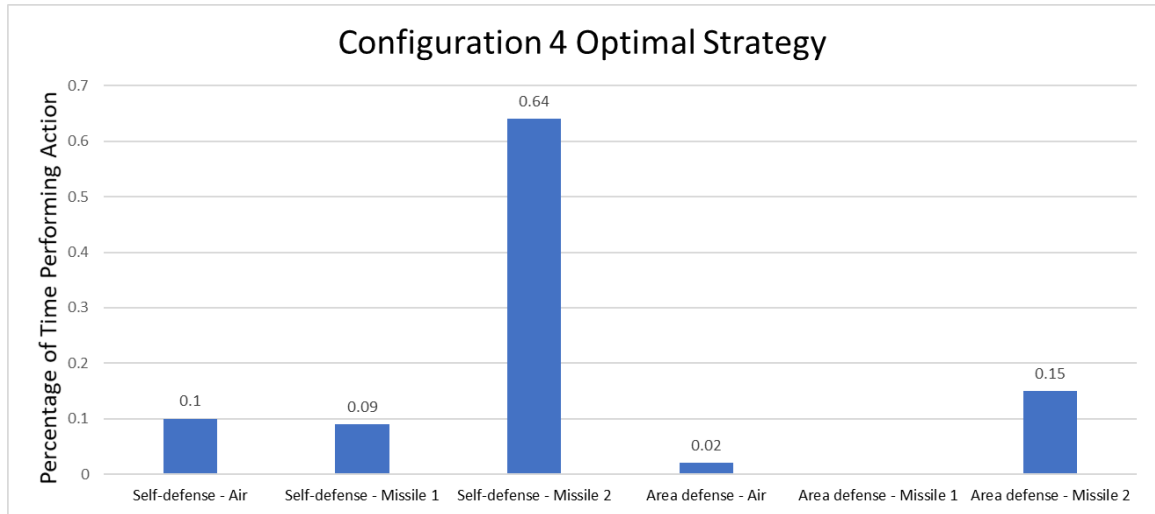


Figure 10. Configuration 4 optimal strategy R_4^* distribution.

While D_4 experiences only a slight utility value improvement over D_3 , with $g_4 = 5.55$ and $g_3 = 5.30$, the action distribution in Figure 10 shows that the D_4 SEWIP Block 3

enhancements to the SLQ-32(V) electronic warfare system enable increased self-defense postures while simultaneously withstanding area vulnerabilities. As such, D_4 DDGs need not adhere to strict area defense actions, which is especially evident in the drop in area defense actions against missile 2.

Configurations 5, 6, and 7

I discuss configurations 5, 6, and 7 (D_5 , D_6 , D_7) in the same section, because all three of these configurations upgrade from the legacy SPY-1D radar to the SPY-6 radar, which is a more capable Air and Missile Defense Radar (AMDR). While D_5 , D_6 , D_7 all have different utility vales of $g_5 = 2.90$, $g_6 = 8.19$, $g_7 = 10.00$, all three configurations have nearly identical optimal action distributions: $R_5^* \approx R_6^* = R_7^*$, as seen in Figure 11

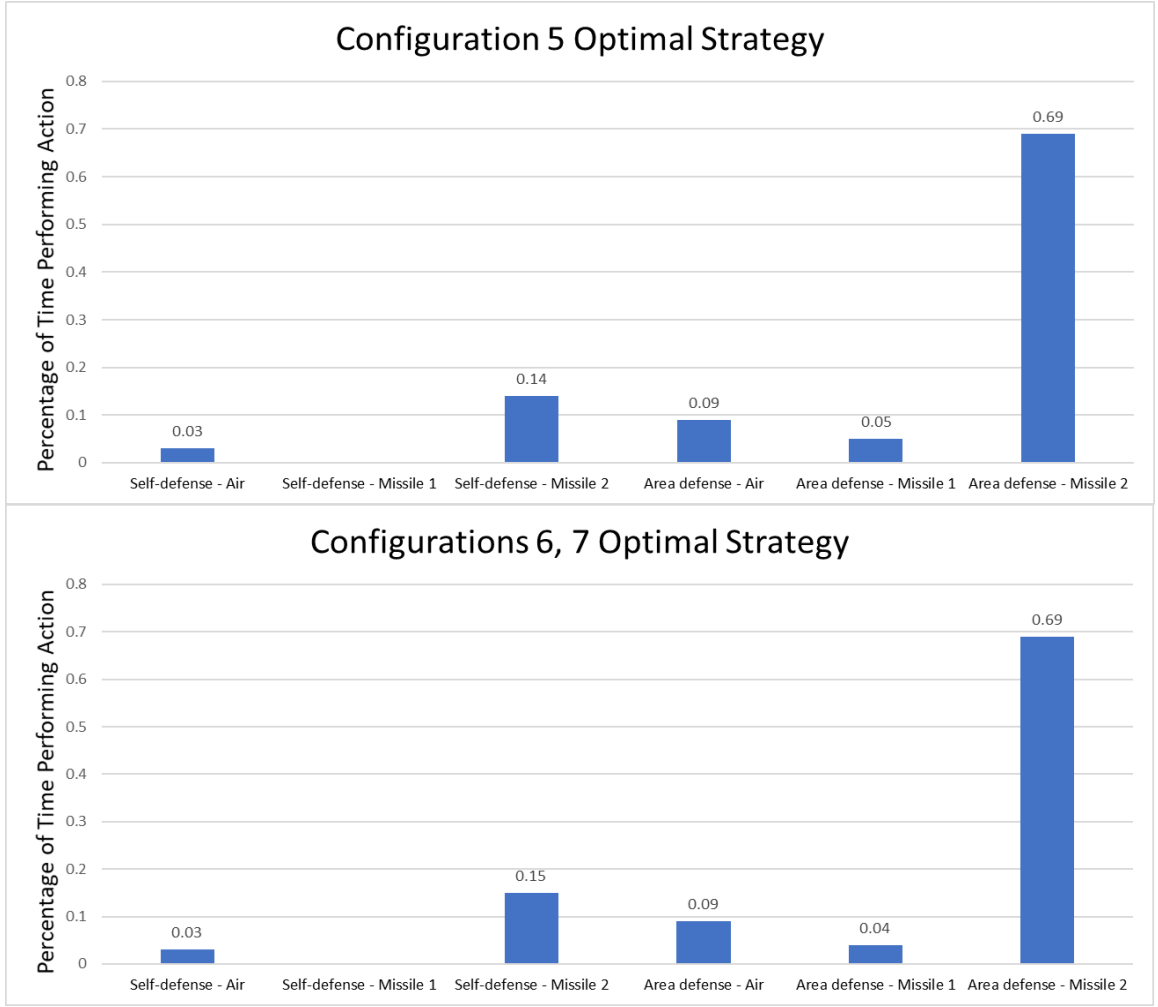


Figure 11. Configurations 5, 6, 7 optimal strategy R_5^* , R_6^* , R_7^* distribution.

The reason for the nearly identical optimal action distributions is relatively simple; all three configurations have nearly identical TPMs, resulting in nearly identical R_j^* , $j \in \{5, 6, 7\}$, but their contribution matrices $C_j(S_t, a_t)$, yield $g_5 < g_6 < g_7$. These results run counter to DM and SME expectations.²³

Recall the descriptions of configurations 5, 6, and 7:

²³ $TPM_6 = TPM_7$, while TPM_5 had only a 0.10 value shift in its *weapon capability 1* sub-state, resulting in a negligible impact on TPM_5^* .

- Configuration 5: SPY-6, SEWIP Block 2
- Configuration 6: SPY-6, SEWIP Block 3
- Configuration 7: SPY-6, Flight III (A and B), SEWIP Block 3

Note that all three configurations introduce the SPY-6 radar into the system, the SEWIP improves from Block 2 to Block 3 between D_5 and D_6 , and the DDG Flight improves to Flight III between D_6 and D_7 . DMs and SMEs considered the SEWIP/Flight changes to be “incremental” improvements compared to the SPY-6 upgrade, and expected g_5 , g_6 , and g_7 to be the “top three” DDG configuration utility values due to their SPY-6 inclusion. However, as seen in Figure 6, g_5 is not only considerably less than g_6 and g_7 , but also less than g_3 and g_4 , both of which include SPY-1D/dLNA combinations. This is evidence that the SPY-6 is not as significant an improvement as expected over the SPY-1D/dLNA combination *from the perspective of the DMs and SMEs*. However, the DMs and SMEs were unaware of how their perspectives would yield optimal strategies and utilities over an operational lifetime that extends beyond human cognitive capability.

The DMs and SMEs involved in this R3B study did not consider the fact that configurations 3 and 4 could achieve a higher utility value than the technically superior configuration 5 if they are strategically deployed to maximize their utility. They believed that the SPY-6-equipped configuration 5 would achieve higher utility than configurations 3 and 4, because they failed to understand that by constructing TPM_5 to (nearly) mirror TPM_6 and TPM_7 (despite $C_5(S_t, a_t) < C_6(S_t, a_t) < C_7(S_t, a_t) \forall S, \forall a$), D_5 would not realize truly maximal g_5 . Simply put, D_5 , because of its SPY-6, is forced to act sub-optimally. The debate over Equation 1 results for D_5 , particularly the fact that $g_5 < g_4$, motivates this

study's application of the ADP Q-Learning model of D_4 vs. D_5 , to be discussed after the following section's MIP results.

Section Two – Mixed Integer Programming Results

Recall Equation 8:

$$\max_{h,i} \sum_{h=1}^{95} \sum_{i=1}^7 g_i X_{h,i}$$

This is the utility-maximizing objective function, where $g_i = \{1.00, 1.79, 5.30, 5.55, 2.90, 8.19, 10.00\}$, the Equation 1 utility value vector for DDG configurations $\{D_1 \dots D_7\}$, and the binary decision variable coefficients for $X_{h,i}$, the decision to equip DDG_h with D_i . Equation 8 is bound by Constraint 1 through Constraint 5, as described in Chapter Four – Solution Methodology. Solving for this MIP yields the maximum expected utility for DDG-configuration assignments, subject to specific budgetary and logistic constraints.

The R3B study DMs and SMEs agree on a minimum requirement of 20 configuration 6 DDGs and 20 configuration 7 DDGs. The 20 configuration 7 requirement ensures that Flight IIIA DDGs 125-126 and 128-137, and future variant Flight IIIB DDGs 138-145, are all equipped with configuration 7. The 20 configuration 6 requirement guarantees at least 40 SPY-6 DDGs in the fleet. The DMs and SMEs understand that the Flight IIIA/B DDGs would be assigned configuration 7 in the optimal MIP solution because they are the least expensive options for upgrade/procurement, as discussed in Chapter One - Introduction. The DMs and SMEs also recognize, after trial and error, that configuration 5, due to its low utility value, would not be in the optimal

solution despite its possession of the SPY-6 radar, so they ensure that the configuration 6 requirement would be such that the minimum of 40 SPY-6 requirement is satisfied²⁴.

While configuration and system requirement consensus are achieved, DMs and SMEs are uncertain of the DDG program budget. To facilitate analysis among this uncertainty, I solve the MIP model with varying budget values assigned to Constraint 1.

Using the FY2021–FY2027 procurement budget amounts discussed in Chapter One - Introduction as budget range baseline estimates, including the configuration 7 procurement budget of \$21,554.4M and the requested FY2021 DDG 51 program budget of \$3,079.2M, R3B DMs and SMEs estimate a minimum expected budget of \$26,000.0M, which is the minimum of the budget range used for Constraint 1.

I demonstrate the effect of various budget amounts on Constraint 1 by plotting budget vs. expected maximum utility, as seen in Figure 12. Expected maximum utility is simply the objective value solution (Equation 8) divided by 95; the number of DDGs in this study.

²⁴ I include the constraints in the MIP model according to DM and SME requirements, but through experimentation I know that the results would have been the same without the constraints.

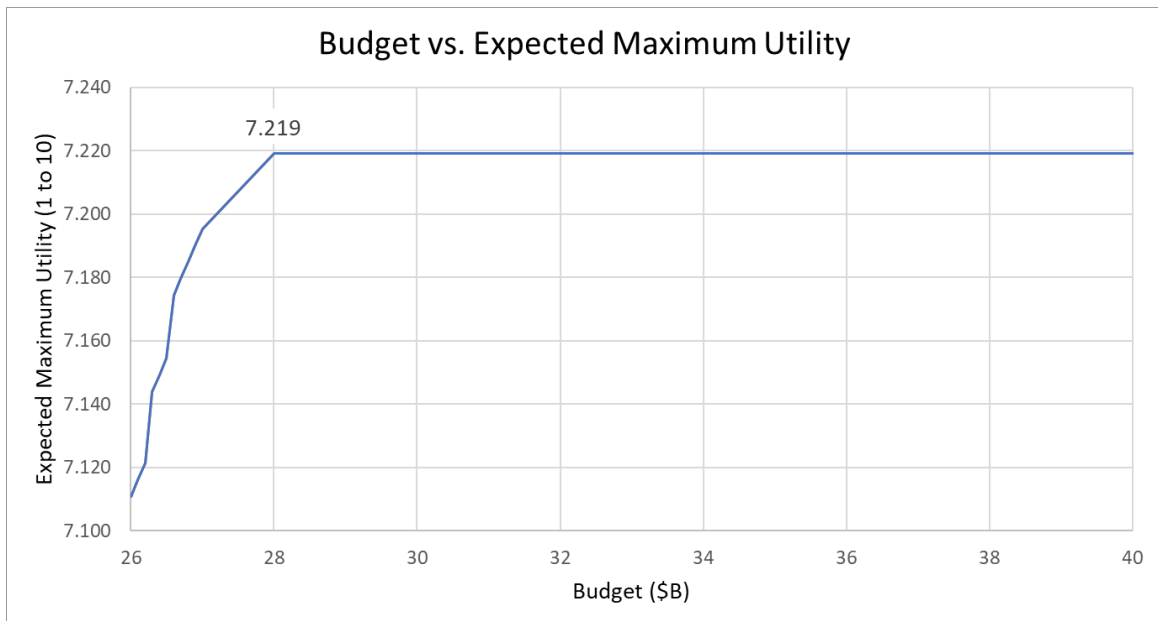


Figure 12. Budget vs. Expected Maximum Utility. The maximum achievable expected utility value (7.219) is labelled for emphasis.

Note that Figure 12 displays a slight expected maximum utility value increase from 7.11 to 7.219 when budget increases from \$26,000.0M to \$28,000.0M, after which the maximum expected utility value remains constant at 7.219. This demonstrates to DMs and SMEs that there is a budget point (here, \$28,000.0M) where higher budget no longer yields higher utility. The Figure 12 results are among the most valuable insights for the R3B DMs and SMEs, since they can argue for a range of DDG program dollars while ensuring senior leaders that there is indeed an upper limit to their budget requirements. This analysis is comforting in the face of projected reduction to the Navy's FY2021 DDG 51 procurement budget.

Configuration Assignments and R3B Submission

Final DDG-configuration assignments are based on the “best-worst case” scenario of a \$28,000.00M budget, which yields the following DDG configuration assignments:

- Configuration 3: 28 DDGs assigned
- Configuration 4: 18 DDGs assigned
- Configuration 6: 29 DDGs assigned
- Configuration 7: 20 DDGs assigned (all Flight III A/B)

These assignments satisfy the requirement that all 20 Flight III A/B DDGs receive the SPY-6 equipped configuration 7, and goes beyond the 20 configuration 6 requirement by assigning configuration 6 to 29 DDGs. Specific hull-number to configuration assignments are beyond the scope of this study, but the results submitted to the R3B represent the *percentage* of eligible DDGs that will receive upgrades, as seen in the histogram on Figure 13, which is the slide submitted to the R3B study.

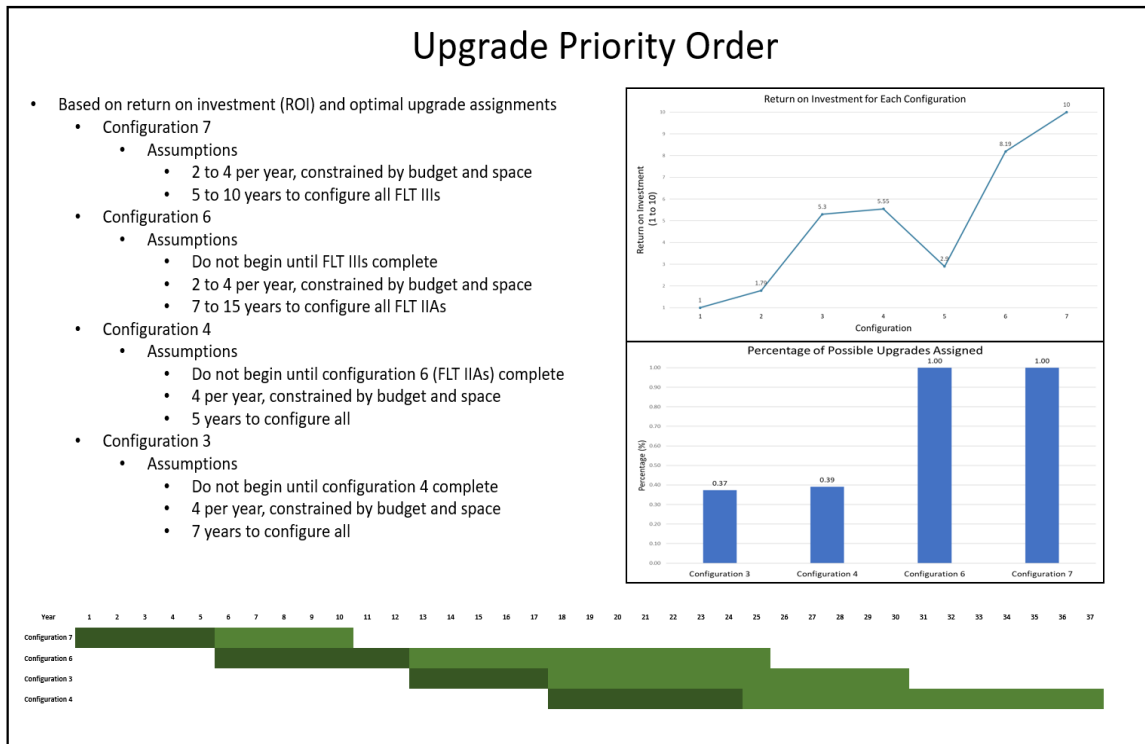


Figure 13. R3B study submission. The top-right image presents g_i as a “return on investment.” The bottom-right image presents the percentage of configuration-eligible DDGs that should receive a particular configuration. The bottom image is a Gantt chart that visualizes the configuration timeline.

Figure 13 displays much information on a single, easily interpretable slide; a necessity when presenting to high-level audiences. The assumptions under each configuration are realistic, but flexible. For example, there is no pushback to the “2 to 4 per year” assumption regarding DDG configuration 7 construction. It is infeasible to remove 20 DDGs from the active fleet and upgrade them all at once. There are also limitations on immediately available budget and space requirements regarding putting the

DDGs in dry dock.²⁵ The assumptions enabled the construction of the Figure 13 Gantt chart. Note that the dark green Gantt chart lines represent the “best-case” timeline, whereas the light green represents the longest case timeline. These timelines are yet another important insight offered to the R3B by this study.

Following the best-case timeline, construction of all SPY-6 equipped DDGs (configurations 6 and 7, *the highest priorities*) takes over 12 years to complete. The longest case for SPY-6 completion extends beyond 25 years, but the true answer is likely somewhere in between. The key point is that completion will almost certainly take more than ten years to complete! Navy program plans and evaluations are always thinking *at least* ten years ahead. Therefore, by the time configurations 6 and 7 DDG upgrades are nearing completion, program evaluations are already being conducted on the next “thing,” thereby reducing the chance of configurations 3 and 4 ever being constructed. In this, the Figure 13 timeline demonstrates that only a fraction of the expected \$28,000.00M will be required to upgrade to the highest priority configurations 6 and 7.

Once again, this is a comforting analysis when faced with projected DDG 51 procurement budget reduction. These results satisfy the need for a modernization and procurement plan that maximizes the lifetime utility of DDGs while adhering to cost limitations, the primary objective of this study.

This study’s secondary objective compares DP results with ADP results that have been informed by the same parameters, which I discuss in the next section.

²⁵ A dry dock is a narrow basin or vessel that can be flooded to allow a load to be floated in, then drained to allow that load to come to rest on a dry platform. Dry docks are used for the construction, maintenance, and repair of ships, boats, and other watercraft (Wikipedia).

Section Three – Approximate Dynamic Programming Results

Recall Equation 4: $V(s) = \max_a Q(s, a)$

Equation 4 calculates the value of state s as the maximum value in the table of Q-values across all actions a . From this, Q-learning evaluates expected utility in a manner that, while numerically different than the DP Equation 1, can confirm Equation 1 results with respect to dominant utility values. Figure 6 shows the unexpected DP utility value result of configuration 4 (D_4) utility value $g_4 > g_5$, the utility value for configuration 5 (D_5). R3B DMs and SMEs inquired about the effect of changing TPM₄ and TPM₅ inputs on DP results. These changes yield different optimal policies, but g_4 remains greater than g_5 . To alleviate concerns over TPM accuracy, I demonstrate Q-learning as a model-free alternative to DP utility evaluation. Instead of TPMs associated with each configuration, Q-learning simply requires state-state reachability, in addition to contribution values for choosing action a from each state s . Q-learning confirms that D_4 utility is greater than D_5 utility, even with unknown transition probabilities:²⁶

- D_4 expected total reward: 112,998.10
- D_5 expected total reward: 73,760.41

These results are consistent across varying levels of Q-learning parameter γ , the discount parameter in Equation 5 that is necessary for $Q(s, a)$ convergence. Varying γ between values close to 0.0 and 1.0 compares myopic and future-seeking reward policies, respectively. While the following discussion regarding varying γ values is outside of the

²⁶ I refer to Q-learning results as “total reward” so as to not confuse these outcomes with the DP utility values.

R3B study scope, the experiment satisfies R3B DMs and SMEs with respect to relative D_4 and D_5 utility while motivating an interesting discussion comparing DP-derived optimal policies and Q-learning near-optimal policies across varying γ values.

Discount Parameter Impact on Near-Optimal Q-learning Policies

I compare optimal/near-optimal policy vectors from both DP and ADP solutions. For clarity, I refer to DP optimal policy solutions as R_4^* and R_5^* , while ADP near-optimal policy solutions are R_4^Q and R_5^Q . I describe R_4^* and R_5^* action *distributions* earlier in this chapter in Section One – Dynamic Programming Results, but here I emphasize that these are, in fact, categorical vectors populated with possible action values 1 through 9. To properly calculate similarities between R_4^* and R_4^Q , and R_5^* and R_5^Q , I translate these vectors into binary vectors so that I can apply the similarity measure known as cosine similarity.

Cosine Similarity and Binary Vector Translation

Cosine similarity is a measure of vector similarity that is often used for document clustering and text mining, as seen in (Muflikhah and Baharudin 2009) and (Li and Han 2013). Cosine similarity efficiently calculates sparse vector similarities, such as the binary optimal policy vectors, as it only considers the non-zero dimensions.

For illustration, cosine similarity between vectors A and B is the cosine of the angle θ between A and B , represented with dot product and magnitude as seen in Equation 9.

$$\cos \theta = \frac{A \cdot B}{\|A\| \|B\|}$$

While vectors R_4^* , R_4^Q , R_5^* , and R_5^Q are represented numerically with possible values 1 through 9, they are ordinal vectors requiring binary encoding to apply cosine similarity.

Configuration 4 Optimal Policy Comparisons

Table 7. R_4^Q cosine similarity to R_4^* .

γ	Cosine Similarity (R ₄ *)
0.1	1.000
0.2	0.991
0.3	0.981
0.4	0.972
0.5	0.972
0.6	0.963
0.7	0.880
0.8	0.806
0.9	0.620

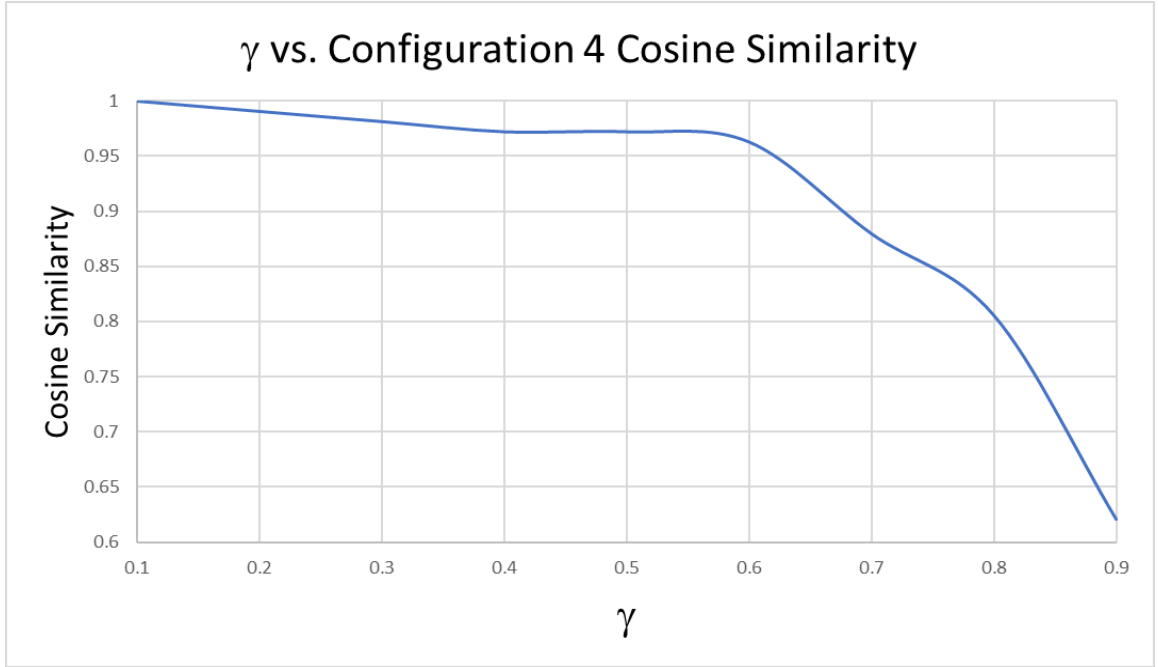


Figure 14. γ vs. configuration 4 cosine similarity.

The cosine similarity between R_4^* and R_4^Q is 1.0 when $\gamma = 0.10$, meaning $R_4^* = R_4^Q$ when γ is at the myopic end of the parameter range. This suggests that unbeknownst to R3B DMs and SMEs, their attitudes regarding configuration 4 are short-sighted. However, note that Figure 14 cosine similarity remains high (greater than 0.96) as γ approaches 0.60, a relatively future-seeking parameter setting. Therefore, I propose that it is more accurate to say that R3B DM and SME attitudes represent a *range* between myopic and *somewhat* future-seeking. After the point $\gamma = 0.60$, R_4^* and R_4^Q similarity drops rapidly, suggesting that R3B DM and SME attitudes do not value far-future rewards as much as near-future and immediate rewards. This makes sense in the force

structure requirement analysis context, as it is appropriate to consider the “lifetime” of a campaign as a timespan that ranges between short-sighted, unlikely combat scenarios and relatively long geographic presence missions that include the non-combat vulnerabilities modeled in this study.

Configuration 5 Optimal Policy Comparisons

Table 8 displays R_5^Q cosine similarity to R_5^* for varying γ , and Figure 15 plots γ vs. configuration 5 cosine similarity.

Table 8. R_5^Q cosine similarity to R_5^* .

γ	Cosine Similarity (R_5^*)
0.1	0.954
0.2	0.954
0.3	0.954
0.4	0.944
0.5	0.944
0.6	0.935
0.7	0.917
0.8	0.793
0.9	0.667

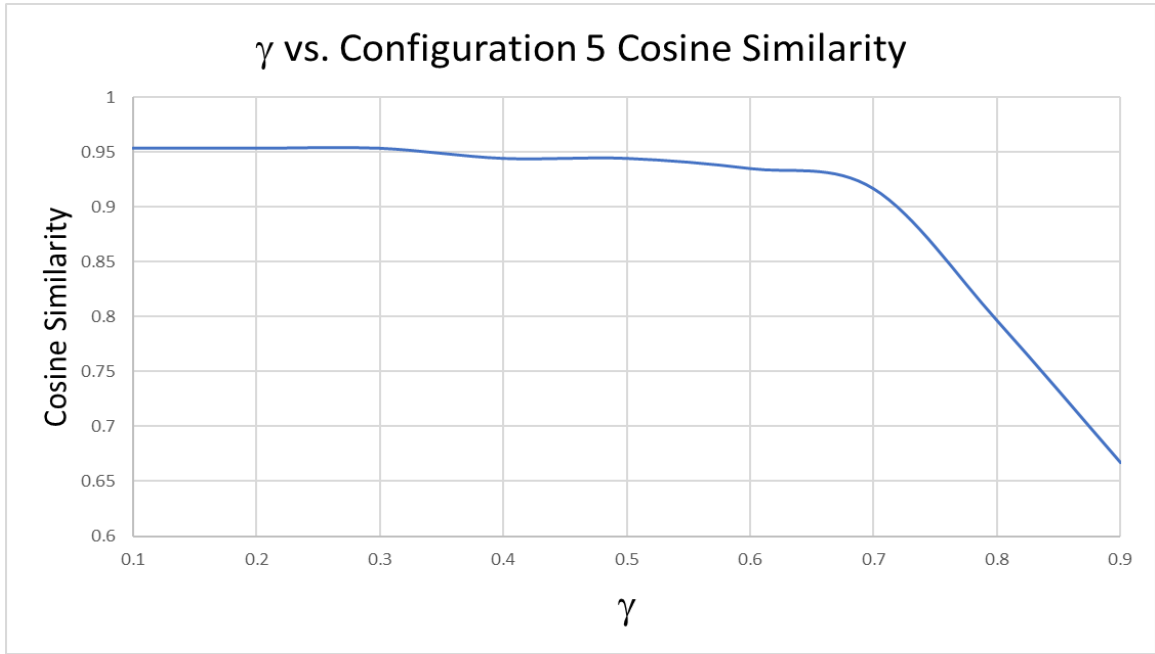


Figure 15. γ vs. configuration 5 cosine similarity.

The cosine similarity between R_5^* and R_5^Q never reaches 1.0 on the myopic end of the plot (γ close to 0.0), but remains high (greater than 0.93) through $\gamma = 0.70$, after which the similarity rapidly decreases. This suggests that R3B DM and SME attitudes towards configuration 5 represent a wider range along the myopic to future-seeking spectrum, while never thinking quite as myopically as they do with configuration 4. Even with the wider range of γ values that are covered with high cosine similarity, these results are consistent with those of configuration 4 regarding force structure requirement analysis and the campaign “lifetime” as a variable timespan. R_5^* remains similar to R_5^Q at higher γ than do R_4^* and R_4^Q , suggesting higher levels of future-reward seeking attitudes among the R3B DMs and SMEs with respect to configuration 5. Recall that configuration 5 includes the SPY-6 radar, and that the DMs and SMEs initially assumed that

configuration 5 utilities are nearly on par with those of configurations 6 and 7. In this, DMs and SMEs likely approached configuration 5 assessment with more far-reaching implications than with configuration 4, which yields a more myopic assessment.

Section Four – Validation

DP and ADP model parameters are validated by SMEs and DMs who combine their own operational experience and expertise with existing simulation model outputs, literature, and related studies. The models do not purport to replace or perform as high-fidelity, realistic models of DDG operational environments, rather they serve as methods to put the collective knowledge of R3B SMEs and DMs to “work.” That is, DP and ADP algorithms remain consistent with SME and DM thoughts and beliefs regarding the operational environments on a temporal scale that is beyond human cognitive capability. Real world validation of this study would require knowledge of how configured DDGs (or, more accurately, DDG Commanding Officers) were to behave in the adversarial environments modeled in this study. This is because the DP and ADP outputs include optimal and near-optimal policies that are in essence DDG deployment strategies designed to maximize configured DDG operational utility. These utilities are not realized if the real-world strategies do not align with optimal policies.

Section Five – Summary

The results of this analysis provide high-level, R3B DMs to make rigorously informed decisions regarding expensive and operationally significant decisions for the future of the US Navy DDG fleet. The unique nature of DP and ADP state-value calculations and optimal/near-optimal policy outputs offer the otherwise difficult-to-

comprehend concept that technologically inferior configurations remain competitive in adversarial, dynamic environments when strategically utilized to maximize their benefit. This study establishes a framework for DP and ADP as decision aids in uncertain environments. The utility-maximizing MIP model results in a defensible roadmap for DDG-configuration assignments and timelines. Here, ADP not only confirms the controversial DP results, but sets a precedent for assessment methodologies that have been adopted as an OPNAV “best practice,” while earning praise by way of a *Presidential Meritorious Service Medal*.

Section Six – Additional Research Questions

Naturally, additional research questions emerge throughout the analysis process that may be outside of the primary objective scope but motivate further research. This study yields two additional questions, one that is operational in nature, and the other theoretical. Operationally, how does one measure strategic adherence to DP/ADP – derived optimal/near-optimal policies? From a theoretical perspective, how do DDG operational timelines map to discount factor (γ) setting with regards to myopic vs. future-seeking policies?

How does one measure strategic adherence to DP/ADP – derived optimal/near-optimal policies?

Maximal DDG-configuration utilities require adherence to the DP/ADP – derived optimal/near-optimal policies. For the purposes of this study and, by extension, the R3B requirements, the DP-evaluated utilities are sufficient measurements of DDG configuration operational value. However, in the context of real-world application, how

does one determine if a DDG Commander is following the optimal strategy, especially considering the low-fidelity, high-level detail that parameterizes this study's models? This question sets up the opportunity for future work that is akin to studies referenced in Chapter Two – Literature Review, specifically those that compare actual MEDEVAC dispatch policies to approximately optimal ADP policies. Such an endeavor would set an impressive precedent for operational-level decision-making aided by AI.

How do DDG operational timelines map to discount factor (γ) setting with regards to myopic vs. future-seeking policies?

Varying γ to observe policy impact is outside of R3B requirement scope, yet the experiment yields an interesting real-world discussion: In terms of DDG operational lifetime, what does it mean to set γ close to 1.0? A twenty-year horizon? An infinite horizon? These are the kinds of questions that I expect would have been asked by R3B DMs and SMEs had the γ sensitivity analysis been within R3B scope. Fortunately, were that the case, Figure 14 and Figure 15 demonstrate that ADP near-optimal policies remain identical / nearly identical to DP optimal policies throughout ranges of γ -settings, beginning on the myopic end of the temporal spectrum. This means that DMs can take comfort in the fact that, despite an uncomfortable grasp on the concept of an infinite time horizon, the DP optimal policies remain optimal within the “near term” (γ close to 0.0) timeframe that exists within operational decision windows.

CHAPTER SIX – CONCLUSIONS

This chapter concludes discussions on methodology, how the study applies to a real-world problem, how the methods and contributions can apply in a broader sense, and future work.

Section One – Methodology

The primary objective is to determine maximum future DDG configuration utility within acceptable cost while satisfying expectations of analytic rigor while remaining consistent with decision maker beliefs. The very nature of this objective invites both technical and theoretical challenges. From a technical perspective, DP and ADP are flexible, interpretable, and practical methods to quantify utility so that the widely accepted MIP method is solvable. These approaches feature immediate sensitivity analysis capabilities, which are attractive to high-level DMs for whom study response time is limited. From a theoretical perspective, this study overcomes challenges regarding the DoD-wide desire to implement AI as a decision-making tool that complements established force structure requirement analysis methods. Military leaders know that they want to incorporate AI into their decision-making processes, but they are not certain how to apply it. This work demonstrates AI-enabled decision-making, but also offers insights into overcoming *bounded rationality* so that DM beliefs remain consistent over a time horizon that overwhelms human cognitive capability.

This study's secondary objective compares DP results with ADP Q-learning results that have been informed by the same parameters. Q-learning output confirms

comparative DP utility results to further satisfy the primary objective, but they also provide academic opportunities to experiment with Q-learning parameters. Specifically, I show potential for estimating where DM and SME attitudes fall on the myopic to future-reward seeking spectrum. I also demonstrate agreeance between DP optimal policies and ADP near-optimal policies due to ADP parameter settings that enable learning and exploration that guarantees near-optimal solutions. This demonstration opens the door for computationally scalable algorithms for larger problem instances. This experiment's practical implications include insight into whether ADP models break free from myopic, combat scenario, attrition-based model limitations.

Section Two – Application

This study's results informed the high-level Resources and Requirements Review Board (R3B) plan to invest in future Navy Destroyer (DDG) weapon systems in order to remain competitive with near-peer adversaries such as China and Russia. Since DP has shown to be a fast, viable technique that can incorporate the very thoughts and beliefs that ultimately drive the decisions, this method achieves practical requirements while gaining DM buy-in. Such buy-in is a non-trivial task, particularly with alternative methods such as campaign modeling and combat simulations that invite platform, scenario, and sensor criticisms.

DP and ADP applications are meant to complement, not compete with, existing models, studies, experience, and overall DM beliefs. DM belief consistency is an attractive modeling feature in and of itself, made more attractive by the ability to instantly re-calculate results when confronted with inevitable "what-if" questions. DP

optimal policies highlight the (perhaps obvious in hindsight) concept that different, technically inferior technologies can achieve superior utility if they are deployed optimally. Specifically, this study's configuration 5, which includes the advanced Air and Missile Defense Radar (AMDR) known as the SPY-6, is technically superior to configuration 4 with its legacy SPY-1D radar. However, R3B DMs and SMEs, believing in configuration 5 superiority, parameterized the model to force it to behave sub-optimally. DMs and SMEs accept this outcome once they understand the optimal/near-optimal policy insight that DP and ADP offer.

This study contributes a rigorously analyzed, well informed DDG configuration construction plan to the R3B. The multi-layered, DP-MIP optimal solution provides construction plans and informs a practical timeline that yields optimal utility at a budget that falls within reduced DDG procurement resources. As per usual for high-level presentations, this study's results are allotted one slide to explain the results, which removes technical details from the presentation. However, R3B study presenters can boast that the results are backed by AI, DM and SME expertise, and multiple studies and simulation models; the level of desired analytic rigor required in this study's primary objective.

Section Three – Broader Contribution

This study's methodology has been adopted as an OPNAV best practice because it represents an all too rare occurrence of OPNAV-stakeholder collaboration. The analytic process begins with DM and SME data elicitation that ensures belief consistency throughout the model evaluations. Furthermore, the software requirements are Excel and

R, two resources that are available to OPNAV analysts who have been trained on the process and underlying theory. OPNAV leadership is actively working on overcoming security and bureaucratic challenges that prohibit the use of advanced analytic capabilities and access to cloud computing, but in the meantime, this study manifests “low power computing” that demonstrates the power of DP and ADP as decision-making enablers. Should this study whet the appetites of DoD leadership so that they overcome aforementioned challenges and leverage available analytic technologies, this study’s methods are scalable to higher fidelity, larger-scale models.

Section Four – Future Work

In the spirit of the idea that any process can be modelled as an MDP, this study motivates the possibility of translating large-scale, scenario driven campaign models to DP/ADP models that enjoy fast computation and break the constraints of specific scenario limitations. When technologically advanced analytic resources become available, DoD analysts can model and quickly evaluate “big-data” (another DoD buzzword) sized problems that are tractable when represented as ADP models. This study, which is by no means a “big data” problem, applies ADP Q-learning to solve model-free versions of the DP algorithm. Q-learning is feasible in this sense because of the relatively small computational requirements. However, now that a strong precedence has been established for ADP application in a military decision-making context, consideration may be given to exploring Q-learning function approximation or post-decision state applications to eliminate the need to evaluate all possible action outcomes.

I recommend (Sutton and Barto 2018) and (Mes and Rivera 2017) as references for practical ADP applications.

From a practical, DoD perspective, the most important aspect of future work is that it breaks free of the confines of “how we have always done things.” This study has been very well received even though it is the first ADP study application in the OPNAV Assessments Division. This study’s adoption as a best practice not only sets a precedent for the approach, but it removes DP/ADP ignorance within the OPNAV analytic community. This study’s recognition in the *Presidential Meritorious Service Medal* Citation reveals the high-level appreciation of innovative, yet practical analytic techniques.

APPENDIX A: SAMPLE DECISION IMPACT CALCULATION SPRAEDSHEET

The spreadsheet sample below displays transition probability deltas given an action is taken by a specifically configured DDG, here being configuration 1. For example, if a configuration 1 equipped DDG takes action 1 (AAW/self-defense), its probability of transitioning to a carrier escort state reduces by -0.094. This spreadsheet ensures that the sum of probability deltas within any sub-state group equals 0.0, so that the sum across all resultant TPM rows is 1.00.

	AAW/self defense	BMD/self defense	IAMD/self defense	AAW/area defense	BMD/area defense	IAMD/area defense	AAW/wide defense	BMD/wide defense	IAMD/wide defense
	1	2	3	4	5	6	7	8	9
Carrier Escort	-0.094	-0.094	-0.1	-0.0272	-0.0272	0	0	0	0
LHA Escort	-0.094	-0.094	-0.1	-0.0272	-0.0272	0	0	0	0
SAG Unit	0.188	0.188	0.2	0.0544	0.0544	0	0	0	0
Sum of Changes	0	0	0	0	0	0	0	0	0
Reliable OTH Comms	0.0816	0.0816	0.0816	0.2304	0.2304	0.2304	0.3	0.3	0.3
Unreliable OTH Comms	-0.0816	-0.0816	-0.0816	-0.2304	-0.2304	-0.2304	-0.3	-0.3	-0.3
Sum of Changes	0	0	0	0	0	0	0	0	0
With advanced SA	0	0	0	0.599	0.599	0.599	1	1	1
Without advanced SA	0	0	0	-0.599	-0.599	-0.599	-1	-1	-1
Sum of Changes	0	0	0	0	0	0	0	0	0
Weapon capability 1	0	0	0	0	0	0	0	0	0
Weapon capability 2	0	0	0	0	0	0	0	0	0
Combined weapon capability	0	0	0	0	0	0	0	0	0
No weapon capability	0	0	0	0	0	0	0	0	0
Sum of Changes	0	0	0	0	0	0	0	0	0
Surface threat vulnerability	-0.1	0.05	0	-0.1	0.05	0	-0.1	0.0599	0
Missile threat vulnerability	0.05	-0.1	0	0.05	-0.1	-3.46945E-18	0.0499	-0.11	-0.0544
Combined threat vulnerability	0.05	0.05	0	0.05	0.05	-6.93889E-18	0.0501	0.0501	0.0544
No threat vulnerability	0	0	0	0	0	0	0	0	0
Sum of Changes	0	0	0	0	0	0	0	0	0

APPENDIX B: SAMPLE CONTRIBUTION MATRIX

This sample spreadsheet displays the first 50 of 192 states and the numeric contribution value of the actions taken from that state, for a specific configuration, in a format amenable for R function calculation. State names are codified as numeric combinations of sub-states.

	Action	AAW/self defense	BMD/self defense	IAMD/self defense	AAW/area defense	BMD/area defense	IAMD/area defense	AAW/wide defense	BMD/wide defense	IAMD/wide defense
State Name	State	1	2	3	4	5	6	7	8	9
1121314151	1	0.6138	0.3756	0.5754	0.7932	0.6014	0.8172	0.4902	0.3884	0.5538
1121314152	2	0.4318	0.5576	0.5754	0.6112	0.7834	0.8172	0.3082	0.5704	0.5538
1121314153	3	0.4682	0.412	0.6218	0.6476	0.6378	0.8636	0.3446	0.4248	0.6002
1121314154	4	0.5336	0.4774	0.5416	0.713	0.7032	0.7834	0.41	0.4902	0.52
1121314251	5	0.6138	0.422	0.5874	0.7932	0.5894	0.8052	0.4902	0.2726	0.438
1121314252	6	0.4318	0.604	0.5874	0.6112	0.7714	0.8052	0.3082	0.4546	0.438
1121314253	7	0.4682	0.4584	0.6338	0.6476	0.6258	0.8516	0.3446	0.309	0.4844
1121314254	8	0.5336	0.5238	0.5536	0.713	0.6912	0.7714	0.41	0.3744	0.4042
1121314351	9	0.6138	0.422	0.5874	0.7932	0.6014	0.8172	0.4902	0.3884	0.5538
1121314352	10	0.4318	0.604	0.5874	0.6112	0.7834	0.8172	0.3082	0.5704	0.5538
1121314353	11	0.4682	0.4584	0.6338	0.6476	0.6378	0.8636	0.3446	0.4248	0.6002
1121314354	12	0.5336	0.5238	0.5536	0.713	0.7032	0.7834	0.41	0.4902	0.52
1121314451	13	0.6138	0.24	0.4054	0.7932	0.4194	0.6352	0.4902	0.2064	0.3718
1121314452	14	0.4318	0.422	0.4054	0.6112	0.6014	0.6352	0.3082	0.3884	0.3718
1121314453	15	0.4682	0.2764	0.4518	0.6476	0.4558	0.6816	0.3446	0.2428	0.4182
1121314454	16	0.5336	0.3418	0.3716	0.713	0.5212	0.6014	0.41	0.3082	0.338
1121324151	17	0.6138	0.3756	0.5754	0.6476	0.4558	0.6716	0.3202	0.2184	0.3838
1121324152	18	0.4318	0.5576	0.5754	0.4656	0.6378	0.6716	0.1382	0.4004	0.3838
1121324153	19	0.4682	0.412	0.6218	0.502	0.4922	0.718	0.1746	0.2548	0.4302
1121324154	20	0.5336	0.4774	0.5416	0.5674	0.5576	0.6378	0.24	0.3202	0.35
1121324251	21	0.6138	0.422	0.5874	0.6476	0.4438	0.6596	0.3202	0.1026	0.268
1121324252	22	0.4318	0.604	0.5874	0.4656	0.6258	0.6596	0.1382	0.2846	0.268
1121324253	23	0.4682	0.4584	0.6338	0.502	0.4802	0.706	0.1746	0.139	0.3144
1121324254	24	0.5336	0.5238	0.5536	0.5674	0.5456	0.6258	0.24	0.2044	0.2342
1121324351	25	0.6138	0.422	0.5874	0.6476	0.4558	0.6716	0.3202	0.2184	0.3838
1121324352	26	0.4318	0.604	0.5874	0.4656	0.6378	0.6716	0.1382	0.4004	0.3838
1121324353	27	0.4682	0.4584	0.6338	0.502	0.4922	0.718	0.1746	0.2548	0.4302
1121324354	28	0.5336	0.5238	0.5536	0.5674	0.5576	0.6378	0.24	0.3202	0.35
1121324451	29	0.6138	0.24	0.4054	0.6476	0.2738	0.4896	0.3202	0.0364	0.2018
1121324452	30	0.4318	0.422	0.4054	0.4656	0.4558	0.4896	0.1382	0.2184	0.2018
1121324453	31	0.4682	0.2764	0.4518	0.502	0.3102	0.536	0.1746	0.0728	0.2482
1121324454	32	0.5336	0.3418	0.3716	0.5674	0.3756	0.4558	0.24	0.1382	0.168
1122314151	33	0.6138	0.3756	0.5754	0.7238	0.532	0.7478	0.3082	0.2064	0.3718
1122314152	34	0.4318	0.5576	0.5754	0.5418	0.714	0.7478	0.1262	0.3884	0.3718
1122314153	35	0.4682	0.412	0.6218	0.5782	0.5684	0.7942	0.1626	0.2428	0.4182
1122314154	36	0.5336	0.4774	0.5416	0.6436	0.6338	0.714	0.228	0.3082	0.338
1122314251	37	0.6138	0.422	0.5874	0.7238	0.52	0.7358	0.3082	0.0906	0.256
1122314252	38	0.4318	0.604	0.5874	0.5418	0.702	0.7358	0.1262	0.2726	0.256
1122314253	39	0.4682	0.4584	0.6338	0.5782	0.5564	0.7822	0.1626	0.127	0.3024
1122314254	40	0.5336	0.5238	0.5536	0.6436	0.6218	0.702	0.228	0.1924	0.2222
1122314351	41	0.6138	0.422	0.5874	0.7238	0.532	0.7478	0.3082	0.2064	0.3718
1122314352	42	0.4318	0.604	0.5874	0.5418	0.714	0.7478	0.1262	0.3884	0.3718
1122314353	43	0.4682	0.4584	0.6338	0.5782	0.5684	0.7942	0.1626	0.2428	0.4182
1122314354	44	0.5336	0.5238	0.5536	0.6436	0.6338	0.714	0.228	0.3082	0.338
1122314451	45	0.6138	0.24	0.4054	0.7238	0.35	0.5658	0.3082	0.0244	0.1898
1122314452	46	0.4318	0.422	0.4054	0.5418	0.532	0.5658	0.1262	0.2064	0.1898
1122314453	47	0.4682	0.2764	0.4518	0.5782	0.3864	0.6122	0.1626	0.0608	0.2362
1122314454	48	0.5336	0.3418	0.3716	0.6436	0.4518	0.532	0.228	0.1262	0.156
1122324151	49	0.6138	0.3756	0.5754	0.5782	0.3864	0.6022	0.1382	0.0364	0.2018
1122324152	50	0.4318	0.5576	0.5754	0.3962	0.5684	0.6022	0	0.2184	0.2018

APPENDIX C: CONTRIBUTION BASELINE AND CONFIGURATION EFFECT

The spreadsheet below displays baseline (configuration agnostic) contribution values, and the effect that each configuration has on the baseline contribution values. For example, the baseline contribution of choosing action *BMD/self-defense* from a *Carrier Escort* state is 0.272. That value reduces by 0.09 if this occurs from a configuration 1 DDG (labeled *1.0: SPY-1, SEWIP BLK II*).

State	AAW/self defense	BMD/self defense	IAMD/self defense	AAW/area defense	BMD/area defense	IAMD/area defense	AAW/wide defense	BMD/wide defense	IAMD/wide defense
Carrier Escort	0.272	0.272	0.421	0.599	0.599	1	0.272	0.272	0.421
LHA Escort	0.272	0.272	0.421	0.599	0.599	1	0.272	0.272	0.421
SAG Unit	0.599	0.599	1	0.768	0.768	0.94	0.09	0.09	0.599
Reliable OTH Comms	0.599	0.599	0.599	0.768	0.768	0.768	1	1	1
Unreliable OTH Comms	0.599	0.599	0.599	0.421	0.421	0.421	0.09	0.09	0.09
With Advanced SA	0.599	0.599	0.599	1	1	1	0.94	0.94	0.94
Without Advanced SA	0.599	0.599	0.599	0.272	0.272	0.272	0.09	0.09	0.09
Weapon Capability 1	0.599	0.768	0.94	0.599	1	1	0.599	1	1
Weapon Capability 2	0.599	1	1	0.599	0.94	0.94	0.599	0.421	0.421
Combined Weapon Capability	0.599	1	1	0.599	1	1	0.599	1	1
No Weapon Capability	0.599	0.09	0.09	0.599	0.09	0.09	0.599	0.09	0.09
Surface Threat vulnerability	1	0.09	0.768	1	0.09	0.768	1	0.09	0.768
Missile Threat Vulnerability	0.09	1	0.768	0.09	1	0.768	0.09	1	0.768
Combined Threat Vulnerability	0.272	0.272	1	0.272	0.272	1	0.272	0.272	1
No threat vulnerability	0.599	0.599	0.599	0.599	0.599	0.599	0.599	0.599	0.599
Configuration	AAW/self defense	BMD/self defense	IAMD/self defense	AAW/area defense	BMD/area defense	IAMD/area defense	AAW/wide defense	BMD/wide defense	IAMD/wide defense
1.0: SPY-1, SEWIP Blk II	0	-0.09	-0.09	0	-0.09	-0.09	-0.272	-0.272	-0.272
1.0 plus SEWIP Blk III	0.272	0.09	0.09	0.09	0	0	-0.272	-0.272	-0.272
1.0 plus dLNA	0.272	0.421	0.421	0.272	0.421	0.421	0.272	0.421	0.421
1.0 plus dLNA and SEWIP Blk III	0.599	0.599	0.599	0.421	0.421	0.421	0.272	0.421	0.421
1.0 plus SPY-6	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09
1.0 plus SPY-6 and SEWIP Blk III	0.768	0.768	0.768	0.768	0.768	0.768	0.768	0.768	0.768
Flight 3 SPY-6 and SEWIP Blk III	1	1	1	1	1	1	1	1	1

BIBLIOGRAPHY

Abdulla, K., J. de Hoog, V. Muenzel, F. Suits, K. Steer, A. Wirth, and S. Halgamuge. 2018. "Optimal Operation of Energy Storage Systems Considering Forecasts and Battery Degradation." *IEEE Transactions on Smart Grid* 9 (3): 2086–96. <https://doi.org/10.1109/TSG.2016.2606490>.

Clark, Bryan, Walton, Timothy A., and Cropsey, Seth. 2020. "American Sea Power at a Crossroads: A Plan to Restore the US Navy's Maritime Advantage - by Bryan Clark

Timothy A. Walton Seth Cropsey." 2020. <http://www.hudson.org/research/16406-american-sea-power-at-a-crossroads-a-plan-to-restore-the-us-navy-s-maritime-advantage>.

Davis, Michael T. 2017. "Approximate Dynamic Programming for Missile Defense Interceptor Fire Control." *European Journal of Operational Research*, 14.

Díaz, Guzmán, Javier Gómez-Aleixandre, José Coto, and Olga Conejero. 2018. "Maximum Income Resulting from Energy Arbitrage by Battery Systems Subject to Cycle Aging and Price Uncertainty from a Dynamic Programming Perspective." *Energy* 156 (August): 647–60. <https://doi.org/10.1016/j.energy.2018.05.122>.

Jiang, D. R., T. V. Pham, W. B. Powell, D. F. Salas, and W. R. Scott. 2014. "A Comparison of Approximate Dynamic Programming Techniques on Benchmark Energy Storage Problems: Does Anything Work?" In *2014 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)*, 1–8. <https://doi.org/10.1109/ADPRL.2014.7010626>.

Johnson, Stephen, Laurence Kotlikoff, and William Samuelson. 1987. "Can People Compute? An Experimental Test of the Life Cycle Consumption Model." w2183. Cambridge, MA: National Bureau of Economic Research. <https://doi.org/10.3386/w2183>.

Keeney, Ralph L. 2004. "Making Better Decision Makers." *Decision Analysis* 1 (4): 193–204. <https://doi.org/10.1287/deca.1040.0009>.

Keneally, Sean K., Matthew J. Robbins, and Brian J. Lunday. 2016. "A Markov Decision Process Model for the Optimal Dispatch of Military Medical Evacuation Assets." *Health Care Management Science* 19 (2): 111–29. <https://doi.org/10.1007/s10729-014-9297-8>.

Lettau, Martin, and Harald Uhlig. 1999. "Rules of Thumb versus Dynamic Programming." *The American Economic Review* 89 (1): 148–74.

Li, Baoli, and Liping Han. 2013. "Distance Weighted Cosine Similarity Measure for Text Classification." In *Intelligent Data Engineering and Automated Learning – IDEAL 2013*, edited by Hujun Yin, Ke Tang, Yang Gao, Frank Klawonn, Minhoo Lee, Thomas Weise

Li, Bin and Xin Yao, 611–18. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-41278-3_74.

Loerch, Andrew G., and Larry, B. Rainey. Methods for Conducting Military Operational Analysis. Alexandria, Va: Military Operations Research Society, 2007.

Lucas, Thomas W., W. David Kelton, Paul J. Sanchez, Susan M. Sanchez, and Ben L. Anderson. "Changing the paradigm: Simulation, now a method of first resort." *Naval Research Logistics (NRL)* 62, no. 4 (2015): 293-303.

Mes, Martijn R. K., and Arturo Pérez Rivera. 2017. "Approximate Dynamic Programming by Practical Examples." In *Markov Decision Processes in Practice*, edited by Richard J. Boucherie and Nico M. van Dijk, 248:63–101. International Series in Operations Research & Management Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-47766-4_3.

Morgan, Brian L., Harrison C. Schramm, Jerry R. Smith, Thomas W. Lucas, Mary L. McDonald, Paul J. Sánchez, Susan M. Sanchez, and Stephen C. Upton. 2018. "Improving U.S. Navy Campaign Analyses with Big Data." *Interfaces* 48 (2): 130–46. <https://doi.org/10.1287/inte.2017.0900>.

Muflikhah, Lailil, and Baharum Baharudin. 2009. "Document Clustering Using Concept Space and Cosine Similarity Measurement." In *2009 International Conference on Computer Technology and Development*, 1:58–62. <https://doi.org/10.1109/ICCTD.2009.206>.

O'Rourke, Ronald. 2020. "Navy DDG-51 and DDG-1000 Destroyer Programs: Background and Issues for Congress," 36.

Powell, Warren B., and Warren Buckler Powell. 2011. *Approximate Dynamic Programming: Solving the Curses of Dimensionality : Solving the Curses of Dimensionality*. Hoboken, UNITED STATES: John Wiley & Sons, Incorporated. <http://ebookcentral.proquest.com/lib/gmu/detail.action?docID=697550>.

Powers, Matthew. 2016. "Factor Analysis and Correspondence Analysis Composite and Indicator Scores of Likert Scale Survey Data." *Proceedings of the 10th*

Annual NATO OR&A Conference, December, 12 2020. “Dynamic Programming in Support of Decision Making.” *Phalanx* 53 (4): 54–57.

Powers, Matthew. 2020. “Dynamic Programming in Support of Decision Making.” *Phalanx* 53, no. 4 (2020): 54–57. <https://www.jstor.org/stable/26964307>.

Pröllochs, Nicolas, and Stefan Feuerriegel. 2018. “Reinforcement Learning in R.” *ArXiv:1810.00240 [Cs, Stat]*, September. <http://arxiv.org/abs/1810.00240>.

Rettke, Aaron J., Matthew J. Robbins, and Brian J. Lunday. 2016. “Approximate Dynamic Programming for the Dispatch of Military Medical Evacuation Assets.” *European Journal of Operational Research* 254 (3): 824–39. <https://doi.org/10.1016/j.ejor.2016.04.017>.

Robbins, Matthew J., Phillip R. Jenkins, Nathaniel D. Bastian, and Brian J. Lunday. 2020. “Approximate Dynamic Programming for the Aeromedical Evacuation Dispatching Problem: Value Function Approximation Utilizing Multiple Level Aggregation.” *Omega* 91 (March): 102020. <https://doi.org/10.1016/j.omega.2018.12.009>.

Rust, John. 2019. “Has Dynamic Programming Improved Decision Making?” *Annual Review of Economics* 11 (1): 833–58. <https://doi.org/10.1146/annurev-economics-080218-025721>.

Schramm, Harrison, and Bryan Clark. 2021. “Artificial Intelligence and Future Force Design.” In *AI at War*, First, Ch. 13. <https://www.usni.org/press/books/ai-war>.

Schramm, Harrison, and Matthew Powers. 2017. “Five-Minute Analyst: The Force Is Strong with Correspondence Analysis.” *Analytics Magazine*. January 5, 2017. <http://analytics-magazine.org/five-minute-analyst-force-strong-correspondence-analysis/>.

Silver, David, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, et al. 2017. “Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm.” *ArXiv:1712.01815 [Cs]*, December. <http://arxiv.org/abs/1712.01815>.

Summers, Daniel, and Matthew J Robbins. 2020. “An Approximate Dynamic Programming Approach for Comparing Firing Policies in a Networked Air Defense Environment.” *Computers and Operations Research* 117 (May): 15.

Sutton, Richard S., and Andrew Barto. 2018. *Reinforcement Learning: An Introduction*. Second edition. Adaptive Computation and Machine Learning. Cambridge, MA London: The MIT Press.

Watkins, Christopher. 1989. “Learning From Delayed Rewards.” *Ph.D. Dissertation, Psychology Dept., Univ. of Cambridge, UK*, January.

Winston, Wayne L, and Jeffrey B Goldberg. 2004. *Operations Research: Applications and Algorithms*. Australia; London: Thomson Brooks/Cole.

Xi, Xiaomin, Ramteen Sioshansi, and Vincenzo Marano. 2014. “A Stochastic Dynamic Programming Model for Co-Optimization of Distributed Energy Storage.” *Energy Systems* 5 (3): 475–505. <http://dx.doi.org.mutex.gmu.edu/10.1007/s12667-013-0100-6>.

BIOGRAPHY

Matthew J. Powers graduated from Upper Merion High School, King of Prussia, PA, in 1997. He received his Bachelor of Science in English from the United States Naval Academy in 2001. He served as a Naval Flight Officer and Operations Research Analyst in the U.S. Navy from 2001-2021, deploying onboard the USS George Washington and USS John C. Stennis Aircraft Carriers while flying the S-3B Viking and the EA-6B Prowler. Matthew received his Master of Science in Operations Research from the Naval Postgraduate School, Monterey, CA, in 2012, where he received the Chief of Naval Operations Award for Excellence in Operations Research. He served as an Operations Research Analyst at the Joint Center for International Security Force Assistance from 2012-2015 where he authored the *Security Force Assistance Assessments Handbook*, and the Joint Lessons Learned Division from 2015-2018 where he received personal commendation from the Chairman of the Joint Chiefs of Staff, General Joseph Dunford, USMC for his development of the widely impactful natural language processing tool known as CHUPPET. Matthew finished his Navy career as a Senior Operations Research Analyst at OPNAV N81, Navy Assessments Division, at the Pentagon. Matthew is on the Board of Directors of the Military Operations Research Society (MORS), where he serves as the Vice President of Professional Development, and has chaired the 2018 Emerging Techniques Forum (ETF), and the 2020/2021 Education and Professional Development Colloquium (EPD). Matthew's published analyses have appeared in the MORS *Phalanx* magazine and *Military Operations Research* (MOR) Journal where he served as guest editor, US Naval Institute (USNI), the Proceedings of the Spring and Winter Simulation Conferences and the NATO Operations Research and Analysis Symposium, *Analytics Magazine*, and *Modeling Sociocultural Influences on Decision Making*, a textbook on cross-cultural decision making. Matthew retired from the US Navy in June 2021 and began employment with The MITRE Corporation as a Lead Operations Research Analyst, where he continues to work while he completes his Ph.D. at George Mason University. He lives in Vero Beach, FL, with his wife, the former Shyla Winter of Lake City, Iowa, and their three children, Aidan, Ben, and Lucy.