Testing and Evaluation of Military Systems in a High Stakes Environment

by

Raphael Moyer

Submitted to the Department of Mechanical Engineering in Partial Fulfillment of the Requirements for the Degree of

BACHELORS OF SCIENCE IN MECHANICAL ENGINEERING

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010

Signature of Author. _____
Department of Mechanical Engineering
May 17, 2010

Certified by: _____
Ricardo Valerdi
Research Associate
Lean Advancement Initiative
Thesis Supervisor

Certified by: _____
Warren P. Seering
Weber-Shaughness Professor of Mechanical Engineering
Thesis Co-Supervisor

Accepted by: _____
John H. Lienhard V
Mechanical Engineering
Chairman, Undergraduate Thesis Committee

(This page intentionally left blank)

Testing and Evaluation of Military Systems in a High Stakes Environment

by

Raphael Moyer

Submitted to the Department of Mechanical Engineering on May 17, 2010 in
Partial Fulfillment of the Requirements for the Degree of

BACHELORS OF SCIENCE IN MECHANICAL ENGINEERING

ABSTRACT

Testing is a critical element of systems engineering, as it allows engineers to ensure that products meet specifications before they go into production. The testing literature, however, has been largely theoretical, and is difficult to apply to real world decisions that testers and program managers face daily. Nowhere is this problem more present than for military systems, where testing is complicated by of a variety of factors like politics and the complexities of military operations. Because of the uniqueness of military systems, the consequences of failure can be very large and thus require special testing considerations, as program managers need to make absolutely sure that the system will not fail. In short, *because of the high stakes consequences associated with the development and use of military systems, testers must adjust their testing strategies to ensure that high stakes consequences are adequately mitigated.*

The high consequence space is broken down into two types of consequences, programmatic and operational. Programmatic consequences occur while a system is under development, and result when insufficient testing is conducted on a system, leading a program manager to have inadequate certainty that the system works to specification. When the program comes under inevitable scrutiny, a lack of testing data makes the program difficult to defend and can thus result in program termination. To address programmatic consequences, testers must utilize a broad based and adaptive test plan that ensures adequate testing across all system attributes, as a failure in any attribute might lead to program termination. To connect programmatic consequences to the realities of system development, the developments of the Division Air Defense System (DIVAD) and the M-1 Abrams main battle tank are examined in comparative perspective, using testing as an explanation for their dramatically different programmatic outcomes. The DIVAD's testing strategy was not adequate, and the program suffered termination because of public and Congressional criticism; the M-1's strategy, by contrast, was very rigorous, allowing the system to avoid programmatic consequences despite criticism.

Operational consequences result from failures of specific attributes during military operations, after the system has already been fielded. Operational consequences are distinguished by their disproportionate impacts at operational and strategic levels of operations, and require targeted testing based on analysis of critical system attributes. The procedure for this analysis is established through use of two case studies. The first case examines a sensor network designed to stop SCUD launches in austere areas; the second case, designed to analyze one system across

several missions, conducts an analysis of the potential operational consequences of failures in the Predator drone's system attributes.

The following seeks to better define the consequences of system failure with the understanding that the military world is in many ways unique from the civilian world. Implicit in this thesis is a plea for program managers to think carefully before cutting testing time in order to reduce program costs and shorten schedules, because less testing means a higher likelihood of disastrous programmatic consequences and less insurance against operational consequences that can dramatically effect the lives of troops in the field.

Thesis Supervisor: Dr. Ricardo Valerdi
Title: Research Associate at the Lean Advancement Initiative

Thesis Co-Supervisor: Dr. Warren Seering
Title: Weber-Shaughness Professor of Mechanical Engineering

(This page intentionally left blank)

Acknowledgements

First and foremost, I would like to thank Dr. Ricardo Valerdi, the leader of the PATFrame group and my thesis advisor, for being willing to take me on and help me through the process of researching and writing this thesis. He went to great lengths to ensure that I had everything I needed to succeed, and his extensive knowledge helped greatly as I tried to figure out the world of test and evaluation. Bob Kenley, a member of the PATFrame group, was also instrumental in the writing of this document, as he not only gave me very insightful comments but also took a great deal of time to teach me the statistics and theory behind testing, while also giving me background on the PATFrame project and the military testing world on the whole. I could not have written this without his guidance. The rest of the PATFrame team was also very helpful in giving me feedback that helped to guide my work—thanks to each of you.

My thanks also go to Professor Warren Seering of the Mechanical Engineering Department, who took me on despite the fact that he has many other advisees, and was always willing to help when I needed it. My appreciation also goes to Lori Hyke, who helped me navigate the thesis process.

Finally, I would like to thank my family and friends—I couldn't have gotten to where I am without you. My parents have given me unending support as I have gone through MIT, and I will be forever grateful. My brother, also a MIT MechE graduate, has always been there for me to bounce ideas off of, and has always provided me encouragement and advice as I've gone through the program. Thanks also to Anna, for being willing to put up with me over these past few weeks.

(This page intentionally left blank)

Table of Contents

List of Figures

Abbreviations

C4	Command, Control, Communications, and Computer
CEP	Circular Error Probability
DIVAD	Division Air Defense system
DoD	Department of Defense
DT	Developmental Test
DVT	Design Verification Test
ECM	Electronic Countermeasures
EO	Electro-optical
FLIR	Forward Looking Infrared
FY	Fiscal Year
GAO	Government Accountability Office
GD	General Dynamics
HMMWV	High Mobility Multipurpose Wheeled Vehicle
IED	Improvised Explosive Device
JIC	Joint Intelligence Center
KPP	Key Performance Parameter
LT	Limited Test
M-1	M-1 Abrams Main Battle Tank
MIUGS	Micro-Internetted Unattended Ground Sensor
MRAP	Mine Resistant Ambush Protected vehicle
MTTF	Mean Time to Failure
MTI	Moving Target Indicator
NATO	North Atlantic Treaty Organization
NCO	Noncommissioned Officer
OT	Operational Test
P(Failure)	Probability of Failure
R&D	Research and Development
RAM-D	Reliability, Availability, Maintainability, and Durability
SACC	Supporting Arms Component Commander
SAM	Surface to Air Missile
SAR	Synthetic Aperture Radar
SCUD	A missile developed by the Soviet Union
TTP	Tactics, Techniques, and Procedures
UAV	Unmanned Aerial Vehicle
USAF	United State Air Force
v_c	Projected Loss
WMD	Weapon of Mass Destruction

(This page intentionally left blank)

1

Introduction

Testing is a critical element of systems engineering across industries, as it allows engineers to iterate designs and ensure that products will meet specifications before they go into full rate production. As such, there has been an ever-expanding literature on testing within systems engineering circles, amidst a broader set of writings on systems design. While the testing literature has been very valuable in defining the statistics and idealized decisions that underlie testing, it is difficult to apply the literature to the real decisions that testers and program managers must make daily. Nowhere is this problem more present than in the testing of military systems, where the testing landscape is complicated by of a variety of factors, including contractor-government relationships, inter-service and Department of Defense politics, the role of public opinion and Congress, and the operational realities that face systems when they reach troops in the field. Moreover, military systems more than civilian systems are more likely to operate in conjunction with many other systems in austere and difficult conditions, making testing a very complex interaction between testers, evaluators, operators, systems, and the environment, all of which tests not only equipment but also tactics, techniques and procedures.

In this thesis, I hope to help disambiguate the testing of military systems by defining and expanding on the *high stakes environment* in which military systems are developed and operated, an environment that has not seen much discussion in the testing literature. In the high stakes environment, because of the uniqueness of military systems, the consequences of failure can be exponentially higher than those for civilian systems, and thus require special considerations and testing strategies. With very high consequences comes a need for additional testing, as program managers need to make absolutely sure that a system will not fail. The following is dedicated to exploring *high stakes consequences* through real case studies, which connect theory to the realities of military systems development and military operations, in the hopes of helping to close the disjuncture between the literature and real world. In short, *because of the high stakes*

consequences associated with the development and use of military systems, testers must adjust their testing strategies to ensure that high stakes consequences are adequately mitigated.

In Section 2, after discussing the theory behind testing, I divide the high stakes consequence space into two distinct types of consequences, *programmatic consequences* and *operational consequences*. Programmatic consequences occur during a system's development, and result from insufficient testing that allows a system to fall victim to potentially unfounded media, public, and congressional criticism. Programmatic consequences are defined by severe cuts to a program's funding, and, in the worst cases, the program's termination, leaving the service politically damaged, wasting vast amounts of government money, and causing a capabilities gap that cannot be filled until another full development cycle occurs. Programmatic consequences have important implications for systems testing, as they require testers to test rigorously across all systems attributes to ensure that they meet specification, and, when a certain attribute does not meet specification, to adapt the test plan to add in more testing after a fix has been implemented. Armed with extensive test data, program managers can better defend their system when Congressional and public criticism takes hold, as occurs for almost all systems. Section 3 expands and brings programmatic consequences to life through a comparative case study of the Division Air Defense system (DIVAD) and the M-1 tank, both developed by the Army in the 1970's and 1980's. Where the M-1 tank was a stunning success in the face of deep criticism, the DIVAD met failure and program cancellation even though elements within the Army and Congress thought it the right system at the right time. While there were several major differences between the programs, I privilege the testing explanation as the reason for the M-1's success and the DIVAD's failure—where the M-1 had a rigorous and highly adaptive test plan, the DIVAD's plan was purposely short on testing time and was by no means adaptive.

Operational consequences, in contrast to programmatic consequences, occur after a system is already in operation out in the field, and result when a specific failure mode of a specific system attribute causes disproportionately large negative outcomes for operational and strategic level plans and operations. A good example of such a failure is collateral damage in a counterinsurgency campaign, which can turn a once-friendly village against coalition forces, necessitate major shifts in troop deployments, and cause numerous friendly casualties in the ensuing campaign. Where programmatic consequences require a broad based adaptive testing strategy, operational consequences require a very targeted, prescriptive testing strategy based on

a comprehensive analysis of the system's attributes and potential missions, from which a list of critical attributes can be derived for further testing. In Section 4, we explore operational consequences and the means by which to analyze them further, first examining a specific mission for a system of systems sensor network, and then by looking holistically at the Predator drone's potential operational consequences and targeted testing needs.

In general, the following seeks to better define the consequences of system failure with the understanding that the military world is in many ways unique from the civilian world. Implicit in this thesis is a plea for program managers to think carefully before cutting testing time in order to reduce program costs and shorten the schedule, because less testing can mean a higher likelihood of programmatic consequences and less insurance against operational consequences that can dramatically effect the lives of troops in the field.

2

Theoretical Framework

Section 2 sets the stage for the case studies in Sections 3 and 4, by giving the reader an understanding of why testing is important, and also establishes a theoretical framework by which the amount of testing a system should undergo can be decided. I lay out my primary argument, which drives the rest of the paper—that military systems, because of their unique high consequence failure modes, need special testing considerations that might be missed if traditional testing methods are applied. I further lay out two categories of high stakes consequences, programmatic consequences (case study in Section 3) and operational consequences (case studies in Section 4), which are distinct in their impacts (the former affects service funding and politics at home, while the latter affects operations abroad) and in the testing strategies used to mitigate them.

2.1 Why Test? Why Not?

Before discussing the amount of testing to conduct on a system, we must first understand the basic reason to test. In short, "the strategic objective of testing…is to reduce uncertainty about the attributes of the system under test [to decide] whether to field, fix, or flunk the system."[1] Without testing, a developer has very little idea how a system will perform once it gets off the drawing board and becomes operational. Good testing protocols allow a developer to find potential trouble areas in the early stages of developing a system, where it is much cheaper to implement a fix before mass production equipment is purchased and production is started. Performance characteristics generally fall into two categories: quality (defects in the system and reliability under continued use) and performance to specifications (does the system meet the expectations to which it was designed to perform).

[1] Kenley, Bob. "Metrics for Testing." Working Papers. Lean Advancement Initiative, MIT 2010.

Ideally, developers could conduct repeated testing on a system across several scenarios to ensure that they knew its precise performance characteristics, which would allow an exact estimate of the risk associated with a system's production. From the risk calculation, the developer can make a better decision on whether to roll out the system (field), implement changes to the system (fix), or to simply abandon the system entirely (flunk). However, testing has negative consequences for a system's cost, as tests cost money, and for a system's delivery schedule, as tests take time.[2] The cost and schedule impacts of a test can vary widely based on test characteristics—testing a specific maneuverability characteristic of a fighter jet might take an afternoon with one pilot, while testing squadron level operational capability against an opposing force might take several weeks with hundreds of personnel. Testing thus trades program money and time for insurance about system performance, and program managers must make an educated decision about what limited resources can be allocated to testing.

Within the military specifically, there is significant pressure to speed up testing, despite the high stakes involved. Slippages in cost and schedule are rampant in military systems development, and to appease Congress it is easy to cut testing in favor of further research and development (R&D) or lowered program cost. Moreover, there is constant pressure to get equipment to the troops in the field as quickly as possible, meaning that testing shortcuts may be taken.

2.2 How Much Should We Test?

Put into broad statistical terms, *testing is meant to educate a developer on the risk associated with a system by reducing uncertainty about the probability of the system's failure.* Risk is defined as the product of the probability of system failure and the consequence of failure, or

$$\text{Risk} = \text{P(Failure)} * \text{Consequence(Failure)} \quad ,$$

and tells us a system's expected "loss" for each type of failure.[3] Take the following very simplified example of a fictional laptop computer, on which a manufacturer makes a profit of

[2] See National Research Council. *Improved Operational Testing and Evaluation*, 2003, 27.
[3] Kenley, Bob. "Metrics for Testing."

$170 at the time of sale. If the consequence of failure of the laptop is a repair cost of $300, and the manufacturer knows for certain that the probability of failure is 0.5 (or half of the laptops will fail), than the risk of producing the laptop is $150, leading the manufacturer to expect a profit of $20 per unit. What if the probability of failure is uncertain, and lies somewhere between 0.4 and 0.6? Then the risk could be anywhere from $120 to $180, meaning that the manufacturer might expect to *lose* $10 per laptop. If the manufacturer is planning to mass produce the laptop, he will want to reduce the uncertainty on P(Failure) by conducting more testing before entering production, which would educate him on whether to field, fix, or flunk the laptop.

One statistical principle that can underlie testing is Bayes' Theorem, which, from a testing perspective, states that knowledge about the probability of an event (like a system failure) can be updated based on new information from testing.[4] Before testing commences, the tester has some prior knowledge that informs him about what he thinks the value associated with the system attribute[5] in question will be. Some attributes are much better understood than others, based upon experience with the attribute in question. For example, a car company might have a pretty good idea of how a suspension will perform under new tuning, while it might have very little idea how a newly designed and newly prototyped engine will endure over long-term operation. The confidence that an organization places on its prior knowledge will help dictate how much testing to conduct.

The role of testing, theoretically, is to both to *update estimates* about the actual value of the attribute[6] in question and also to *reduce uncertainty* about that estimate. In other terms, it adjusts the mean of the distribution of the attribute in question and also narrows the confidence interval associated with that mean. Here is a theoretical example of a test designed to better determine the mean time to failure (MTTF) of an unmanned aerial vehicle's (UAV) engine at a set cruising speed. The design group thinks that the MTTF is 103 hours (that is, they think the mean of the MTTF for all the UAVs produced will be 103 hours), and their specification says

---

[4] Clemen. *Making Hard Decisions*. Belmont, CA: Duxbury Press, 1996.

[5] A system attribute is a system characteristic that is to be tested.

[6] An attribute is simply an underlying system characteristic. Attributes can range from the easily measured to the nearly immeasurable. For example, they can be the number miles between failure of a vehicle's drivetrain, the number of casualties an infantry battalion takes while attacked with a certain size enemy force, or the improved "situational awareness" given a military unit by an improved communications system. As these examples imply, finding concrete numerical methods to measure attributes can be very difficult, and the confidence intervals associated with those attributes can be large.

that the MTTF must be 100 hours tested to 95% confidence.[7] Without testing they are not completely confident that the MTTF is above 100 hours (they in fact are 95% confident that it lies between 98 and 108 hours). The design group thus sends the UAV to the testers to better determine the MTTF value. The testers put ten UAVs into the sky one after the other, and find that their engines go out after a mean 110 hours of testing, with a 95% confidence interval of 107 to 113 miles. With this new information, the group can predict that the MTTF of the vehicle will be 108.1 hours, and that the 95% confidence interval will be from 105.6 to 110.7 hours.[8] As we might expect, not only is the 95% interval tightened (from ±5 for prior knowledge and ±3 for testing to ±2.6 once the testers have the benefit of both), the mean for the MTTF has shifted. With more testing, the confidence interval will continue to shift towards the actual mean MTTF and will continue to tighten.

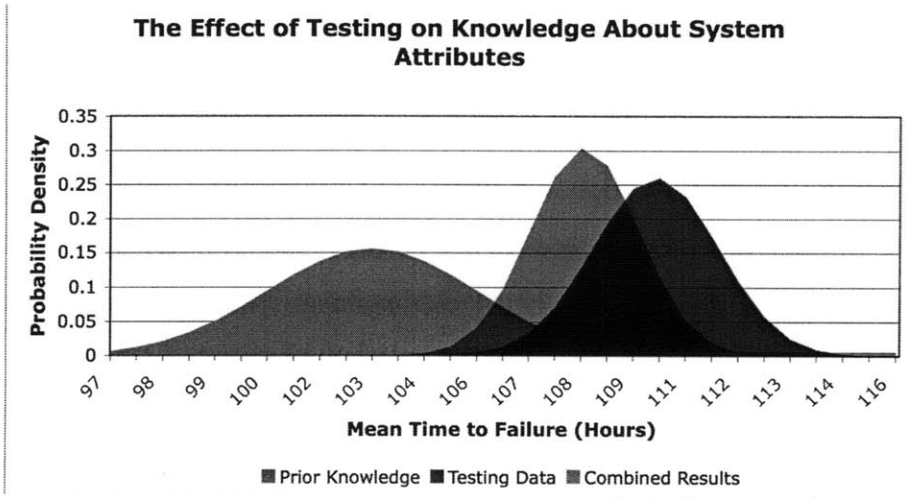

Figure 2-1: Prior Knowledge, Testing, and Final Confidence.

It is important to remember that infinite testing does not necessarily mean a very small confidence interval. Testing equipment can only measure to a certain precision, and often the

---

[7] Please note that 95% confidence is used in this paper as the acceptable confidence interval for a system attribute to meet specifications—it is used only to make calculations easy, not to represent the actual confidence interval specifications require.

[8] The value for the new MTTF can be calculated from the equation $\mu_{revised} = \mu_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}(\mu_2 - \mu_1)$ where μ_1 is the mean of the prior knowledge, μ_2 is the mean of the test results, $\mu_{revised}$ is the new knowledge of the MTTF after testing, σ_1 is the standard deviation of the prior knowledge, σ_2 is the standard deviation of the test results, and $\sigma_{revised}$ is the standard deviation after the testing is complete. The value for the new standard deviation is $\frac{1}{\sigma_{revised}^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$. These calculations are equivalent to the Kalman filtering calculations described by Yaakov Bar-Shalom, X. Rong Li, and Thiagalingam Kirubarajan in *Estimation with Applications to Tracking and Navigation,* 2001.

attribute under test has some variance associated with it no matter how many tests are conducted. Eventually, additional testing will only result in marginal gains in information about an attribute. In the UAV case, for example, the engine might have an actual MTTF of 106 hours with a 95% confidence interval of 103 to 109 hours. After 30 tests, the testers might believe the MTTF is 107 hours with 95% confidence from 104 to 110; after 100 tests, the MTTF might seem to be closer to 106.4 hours with 95% confidence from 103.4 to 109.4; after 1000 tests, the MTTF might seem to be 105.9 with an interval from 103.1 to 108.7. After 100 tests, the tester is pretty close to the real value of the attribute; after 1000, the tester is only marginally closer.

The expected value of testing can be calculated through statistics, assuming that the attribute in question can be measured. If an organization has some prior knowledge of the attribute in question, it can develop a probability distribution that indicates what it thinks the attribute will be (see Figure 2-2).[9]

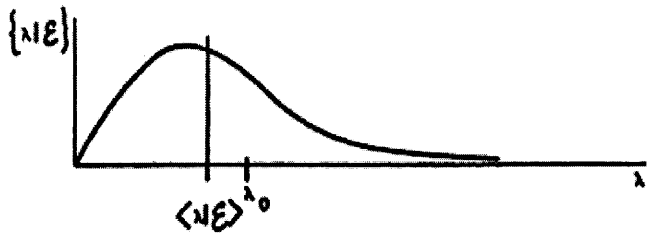

Figure 2-2: Prior Knowledge Distribution ε on Some Attribute λ (Howard, 1965).

If we assume that above some value of the statistical defect rate λ_o the company's product will not be profitable (in the laptop case, for example, λ_o would be the defect rate at which the laptop product can be expected to be unprofitable), the projected loss from the product (v_c) can be calculated and escalates linearly with increasing λ_o (Figure 1-2). If testing gave "clairvoyance" to the tester, or perfect information about the value of λ (this is a very optimistic estimate), the value of that clairvoyance is equal to the projected loss, because testing would have stopped that loss from occurring as the test outcomes would have resulted in a "flunk" decision.[10] Note that the value of clairvoyance is equal to zero below $\lambda = \lambda_o$.

[9] Howard, 1965.
[10] Howard, Ronald A. "Bayesian Decision Models for System Engineering." 1965.

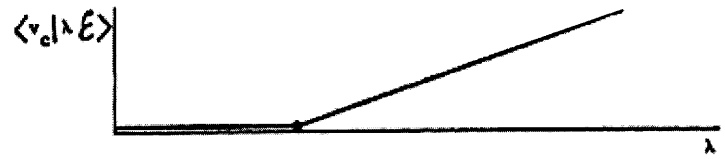

Figure 2-3: Expected Loss (v_c) Based on Expected Probability of Failure λ (Howard, 1965), or the Value of Clairvoyance.

Recall that Risk = P(Failure) * Consequence(Failure). Risk in this case would be the probability distribution for the value of failure, or the area under Figure 2-2, times the consequence of failure (also the value of clairvoyance), or the area under Figure 2-3.[11] The value of risk tells the tester how much testing is actually worth, and informs his decision on whether or not to test. In these idealized situations, the tester accurately knows the probability distribution associated with λ, the expected loss based on various values of λ, and also has testing that gives him complete information about the value of λ. While any of these criteria would be difficult to actually know accurately, they still provide a very good numerical methodology for testing. Moreover, they are not impossible to attain.

Through the process of testing, an organization can thus become more sure about the values of the performance and quality characteristics of the system it has created. After testing, the organization both knows the value of the attribute better, and is more confident that the attribute lies within a certain range. As discussed in the previous section, however, there are constraints to how much a system can be tested based on cost and schedule. This is where organizational preferences come into play—the organization making and responsible for testing the system must decide how much testing is worth to them. Are they willing to forgo additional confidence in the value of the attribute to be tested in order to field the product faster? Is the attribute one whose failure mode could have catastrophic effects, and must be known with very high precision? In short, can their confidence interval about the attribute be large, or must it be small?

[11] Ibid.

2.3 Testing in a High Stakes Environment

In the next two subsections, I seek to outline two types of failure mode largely unique to the military that can have very large negative consequences, and potentially very high risk—hence the "high stakes" label.[12] The key with high stakes events is that their broader impact is disproportionate to their immediate impact. These failure modes, because of their unusually high consequences, need additional testing to ensure that they have both very low probabilities of occurring (and thus low risk), and also that they have tight confidence intervals associated with those low probabilities of failure. Where some system attributes might only need to be tested to a 95% confidence interval that they meet specifications, high consequence attributes might need to be tested to 99% confidence.

In more mathematical terms, high stakes consequences have exponentially higher consequences than regular consequences of system failure. The following is meant to illustrate why we need more testing of attributes that have high stakes consequences, through a loose statistical analysis (meant merely to prove a point, rather than present an analytical method). On an arbitrary scale of the importance of consequences, a tank unexpectedly throwing its tread and being unable to move without maintenance is 1 point, that the death of a soldier because an armor plate failed is 10 points, and the destruction of an Afghani house filled with innocent civilians because of the failure of a guidance system is 50 points (high stakes consequence). The first is not devastating, as there are probably other vehicles to cover for the immobilized one; the second is a tragedy, but the death will probably not radically effect operations; the third is a huge problem, as not only could the village turn against NATO forces and begin to support the insurgents, but several of the members of the village might take up arms, destabilizing a whole area and forcing NATO to commit people and resources.[13]

Let us also assume for discussion's sake that our prior knowledge tells us that each event has an equal probability of 0.2, with a 95% confidence interval from 0.1 to 0.3. Testing costs one point (an assumption to make calculations easier), and will make the 95% confidence interval 0.15 to 0.25. Without further testing, we know that the risk, or P(failure) * Consequence(failure),

[12] High stakes consequences can occur in the commercial world as well. The recent debacle at the Toyota Motor Company, where accelerator pedals got stuck and caused several deaths (Rooney, 2010), stands as an excellent example of this. Though the ensuing recall was expensive, the immediate damage to brand image was far worse.
[13] One of the key tenets of Pashtun culture, for example, is revenge for the death of a relative, which may cause family members to feel obligated to join the insurgency.

associated with the tank throwing its tread is between 0.1 and 0.3 (see Figure 1-4), the risk associated with the armor plate failing is between 1 and 3, and the risk associated with the guidance failure is between 5 and 15. After testing, the numbers would become 0.15 to 0.25, 1.5 to 2.5, and 7.5 to 12.5. We only care about the upper bound, as that is the highest probability of failure and thus has the most risk associated with it. For the tank tread and the armor plate, testing thus only reduces our potential risk by 0.05 and 0.5, respectively. Since testing costs 1 point, it is not as worthwhile to test either. For the high stakes guidance system failure, however, testing reduces our potential risk by 2.5 points, making it worth testing, especially relative to the other attributes. The takeaway is that disproportionately large consequences make testing of associated attributes worthwhile, because the associated risk reduction is much more valuable.

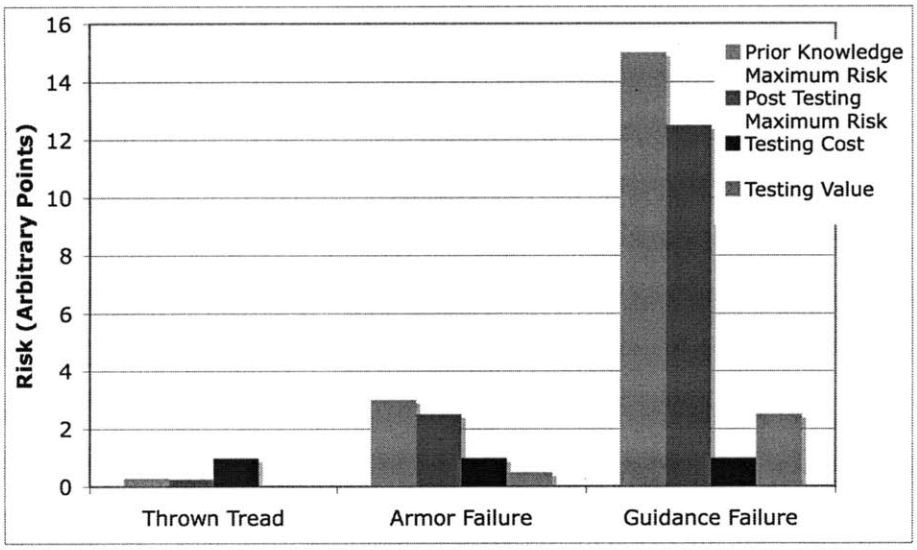

Figure 2-4: High Stakes Failure and Testing.

The methodologies used to explain testing thus far have been highly idealized, as it is very difficult to measure in precise terms how much testing is worth in military settings (especially high stakes ones). How can we know how much value to place on collateral damage? Defining confidence intervals and specifications for high stakes events is also very difficult. How can we set a standard for fratricide, for example, especially in an organization that has a policy of zero tolerance? Even if we decide that only one out of every 10,000 rounds fired is allowed to hit friendly forces, how can we be 95% confident that a weapons system has met this standard? Especially with resource constraints, how can we decide which high stakes attributes to test, at

the expense of testing others? These questions will be briefly addressed in the conclusion, which proposes future work on metrics for high stakes consequences and their implications for testing.

The two key categories of failure that occupy the bulk of my research are high stakes failures that have *programmatic consequences* and those that have *operational consequences*. The delineation between these types of consequences is how and when they occur—the former sorts of consequences occur before full-scale production and are unpredictable, needing broad spectrum testing of many systems attributes, while the latter occur after the system has been fielded and are predictable through analysis, and must receive targeted in depth testing of a few systems attribute. Programmatic consequences can cause a program to be canceled or severely disrupted, with important impacts for both inter-service politics and troops on the ground. Operational consequences involve failures during operations that have ripple effects far greater than the immediate impact of the failure itself. The following discussions of these two categories set the stage for the comparative case study of programmatic consequences (Section 3) and the case studies that will help us define and explore operational consequences (Section 4).

2.3.1 Programmatic Consequences and High Stakes Failure

The term "programmatic consequences," as used here, defines consequences that have severe negative impacts on the resources of a systems acquisition program.[14] The most severe of these consequences is program cancellation—because of a specific system failure, Congress might decide to cut funding to a program entirely, forcing the service to either start a new program to fill the capability gap or simply abandon the capability entirely. Other programmatic consequences might be significant funding cuts, which would impede program progress, and additional oversight, which can often create delays by forcing the project to meet milestones and preventing it from proceeding without a green light from the Department of Defense. It is very difficult to predict which attributes will suffer from programmatic consequences, and these consequences occur only before the system has been fielded.

Programmatic consequences are important for a variety of reasons, and their implications necessitate that systems receive testing sufficient to ensure that all major specifications are met. One easily recognized implication is that, because the program has been cancelled or delayed,

[14] The words system and program are used interchangeably in the following section.

troops out in the field do not receive a system, causing units to miss a potentially critical capability. A good example of a missed capability is the Army and Marine Corps's ubiquitous Mine Resistant Ambush Protected vehicle (MRAP) program. Because of a variety of factors, the heavily armored MRAPs only made it into the field several years after the need for them was recognized, causing the loss of numerous soldiers' lives as they rode in lightly armored vehicles. Programmatic consequences also mean that a service can lose prestige and power vis-à-vis the other services. If the Army cannot develop an air defense system, for example, then the Air Force may receive money formerly allocated to the Army to beef up fighter forces. Finally, cancellation of a program is very expensive, as the money spent on research and development is forfeited, and cannot be used to work on another program that might be just as important.

So how can testing help with programmatic consequences? Programs can often be canceled regardless of whether rigorous testing was conducted—if the system simply cannot perform to specification, and the program has already had several setbacks, the system might get canceled outright. Even if the program appears to be on track to meet specifications, it still might be cancelled or have funding cut. There might have been cost overruns that lead the service or Congress to decide that the program simply is not worth it; the program might be way behind schedule; once the system has been in development for several years, changing political conditions might make it no longer necessary (the Army's Crusader artillery system, a relic of the Cold War and cancelled in 2002, is an excellent example of this).

Where testing really helps is when, as we will see in Section 3, public and media misperceptions of system performance lead Congress to cancel a program. Especially when aggressive development schedules are pursued, bench level testing, integration testing, and developmental testing can be given short shrift in order to accelerate the development of a prototype.[15] When the under-tested prototype reaches operational testing, where it is put into a full-fledged field environment, it can often perform poorly—there simply has not been time to fix deficiencies in the system. If something goes wrong in testing, especially if it goes wrong in a spectacular fashion (spraying bullets near the audience of a demonstration, for example, as

[15] Bench level testing is the testing of system components individually to ensure that they work. Integration testing is when multiple components that form an assembly are tested together. Developmental testing is when a system is tested to ensure that it meets specifications, but is not put through rigorous testing in a field environment (which occurs during operational testing) (Conversations with MAJ Kris Cowart, USAF).

happened in the Division Air Defense (DIVAD) system's case),[16] it can cause a media firestorm that causes severe doubt about the system both by the public and by Congress. If the service responsible for the system has rigorous operational testing data that shows the mishap to be an anomaly (or has a good fix for the problem prepared), it can properly present its case in Congress and with the rest of the Defense community; if the service has little hard testing data to show, the system may be at risk for schedule delays, budget reductions, or, in the most extreme cases, cancellation. Perception, not reality, often can shape programmatic decisions.

Programmatic consequences cannot be avoided by the rigorous testing of one system attribute, as they can result from the failure of almost any part of the system. If the armor on an armored personnel carrier allows a machine gun round through, the media might proclaim it unsafe for soldiers; if the fire control system of an air defense gun goes haywire, it might be called a friendly fire risk; if a plane crashes because its engine went out, there might be public outcry for the safety of pilots. Regardless of the type of failure, the consequences can be the cancellation of the program, a capabilities gap in military equipment, and the loss of prestige and power for the service involved. Assuming that the failure is an anomaly, the service's only hope to avoid consequences is the presence of extensive testing data with which the service can prove its case.

The key is that the service must establish clear specifications for each primary system attribute, and test them so that they are confident that they meet those specifications—if one attribute is given short shrift in testing, and comes up deficient in an isolated incident, the system could see extensive scrutiny based on anecdotal evidence. Say, for example, that a system has attributes one through five. After the first round of testing, testers are 95% confident (for discussion's sake 95% confidence is considered sufficient) that attributes one, two, and four meet specifications, but are not confident about three and five. They decide to conduct further testing on five to get it to a 95% confidence interval, but find that testing three would be too expensive, not thinking about the potential for high stakes consequences. They may have assessed the risk of failure properly at the small picture tactical level—a tank throwing a tread more often than expected might be a reasonable risk if the testing would be prohibitively expensive—but at a political level an unusual failure on attribute three (a platoon of vehicles, by chance, might throw multiple treads while a congressman was watching), causing congressional oversight that,

[16] Krishnan, Armin. *War as Business: Technological Change and Military Service Contracting,* 2008. 75.

without rigorous testing to refute congressional claims, may result in large negative consequences.

2.3.2 Operational Consequences and High Stakes Failure

"Operational consequences" denote consequences that can result from the failure of a specific system attribute during an operation that has effects disproportionately larger than those immediately seen. A good example of this phenomenon might be collateral damage from a "precision" missile strike in Afghanistan. While the loss of non-combatant life (the immediate consequence) is tragic, what is more worrisome from a military standpoint is that the collateral damage may cause more people to join the insurgency, growing the movement and making it much harder to defeat.[17] Where programmatic consequences affect service politics and also can lead to capability shortages, operational consequences affect current military operations. Operational consequences become much more relevant in the current media-rich, unconventional warfare environment, because incidents in remote areas and the deaths of small numbers of noncombatants or soldiers can quickly have far-reaching effects. Operational consequences need not be limited to a theater of war, as high consequence incidents (like a UAV crashing into an airliner, for example) can have harmful effects on public opinion that can damage civil-military relations and cause operations to become hamstrung by government regulation. Those systems attributes that can have operational consequences need to be tested extensively to ensure that high consequence failure modes do not occur.

Unlike programmatic consequences, which point to a need for rigorous testing across numerous attributes, operational consequences require testing of very specific system attributes to very tight confidence intervals. The best way to find which attributes need especially rigorous testing is through systems analysis: how will the system be used, and which failure modes can have disproportionately large negative effects? Analysis can never uncover all of the attributes that could have high stakes failure modes, because systems are often used in ways that they were never intended to be used, but it gives us the best possible way to decide which attributes to test. If testing for these specific attributes is not conducted to a high enough confidence interval, we cannot be sure that the system has a low probability of failure in that attribute, and thus cannot be

[17] See, for example, Mayer, Jane. "The Predator War." The New Yorker. October 26, 2009.

sure that the system does not have very high risk. In the example above, testing for air-to-ground weapon systems must be especially rigorous on attributes that involve targeting, so that we can be sure there will be little collateral damage during air strikes. Other attributes, like the maneuverability of the aircraft, do not have high-stakes operational consequences, and do not need to be tested to the same very high confidence intervals.

2.3.3 Programmatic and Operational Consequences in Comparative Perspective

There are two primary differences between programmatic and operational consequences: the former occur before the system is fielded and involve testing to specifications on a broad range of attributes, while the latter occur when the system is already operational and require very extensive testing on a few attributes. The consequences also have very different implications, as programmatic consequences negatively affect the development of a system and operational consequences negatively affect military operations. As was noted above, testing for attributes that could have programmatic consequences is only needed to ameliorate negative public, Congressional, and defense community perceptions of a system based on a particular incident, so that the program does not lose funding or slip on schedule; testing for specific failure modes that could have operational consequences is needed to prevent calamities in the field that can have significant impacts on operations.

So why is there such a large difference in how the two sets of consequences need to be accounted for during testing? Programmatic consequences, as discussed above, can result in the failure of almost any system attribute, meaning that all important system attributes must be tested rigorously to specifications. Operational consequences, by contrast, can only result from the failure of specific attributes, and require especially rigorous testing on those attributes. It is easy for a tread or armor failure to cause public outcry when a system is in testing stateside, with heavy extra-service presence (GAO, DoD, and media); when the system is deployed abroad, such operationally low stakes failures do not have dramatic consequences, as they only really affect the unit operating the system with little broader effect. An armor failure causing the death of a soldier, for example, is tragic, but has little greater operational or strategic impact.

The other primary difference between programmatic and operational consequences is the time at which they occur. The timeframe for programmatic consequences is obvious—if the

system has not been fielded yet, it cannot have operational consequences, and the worst that can happen is that the system's program gets cut. But why can't operational factors have programmatic consequences? Presumably, a system's program could still be cut and the system recalled after its deployment. There are two reasons why the distinction between operational and programmatic consequences is reasonable: first, by the time a service has fielded a system, it has invested large sums of money and institutional effort in creating it—the service will be loath to give it up unless the system proves absolutely incapable of performing its mission, making programmatic consequences unlikely. Additionally, once a system is deployed, programmatic consequences become secondary, as operational consequences, which can cost many lives, destabilize security efforts, and cause public opinion disasters may dominate.

2.4 Conclusion

The primary thrust of the argument thus far is that there are certain high stakes consequences, unique to military systems, which require systems to be tested rigorously. These consequences come in two forms: programmatic consequences, which require all main system attributes to be tested rigorously to specifications, and operational consequences, which require especially rigorous testing for certain attributes. The motivation for discussing high stakes consequences lies in a fear that military test planners may not account for certain failure modes adequately during testing. If testers only think in reference to design specs, rather than the big picture consequences of testing, they may not test key features that could end up causing massive consequences in the field. Budget and time constraints may mean that only limited testing is conducted, making it all the more important to target testing to the most critical systems areas. Additionally, the metrics involved in testing high stakes consequences can be difficult to quantify, leading testers to steer away from trying to test them at all. Finally, the literature on testing uses idealized equations to assess the value of testing, and consequences may be seen to all have the same general order of magnitude. These valuations, however, do not take into account "high stakes" consequences that are exponentially more important than other consequences.

The following sections will examine the "high stakes consequences" challenge in more depth. The next section takes a look at programmatic consequences by conducting a case study

on the Division Air Defense system, which did not receive enough testing and was cancelled after the expenditure of $2 Billion. The third section looks at potential operational consequences specific to unmanned systems, to give a flavor for the sort of operational consequences that could result from inadequate testing. The final section, as alluded to earlier, attempts to quantify "high stakes" consequences and the amount of testing that needs to be conducted to prevent them.

3

Programmatic Consequences, the DIVAD, and the M-1 Tank

The following section is meant to show programmatic consequences in action by looking in comparative perspective at the U.S. Army's M247 "Sergeant York" Division Air Defense System (DIVAD) and M-1 Abrams Main Battle Tank (M-1), both developed from the mid-1970's to the mid-1980's. Programmatic consequences, as discussed in Section 2.3.1, are consequences that can result in the termination or delay of a system while it is still in the acquisition phase, before it is ever fielded. These kinds of consequences occur when a system falls under severe criticism from the media, Congress, or organizations like the Government Accountability Office, and the service cannot bring adequate testing data to the fore to exonerate the system's performance. The criticism can range from the extraordinary, where a system goes haywire, to the ordinary, where a system shows normal growing pains that nevertheless become the target of negative media reports and congressional attention. We are only interested in cases where the system is fundamentally sound and insufficient testing makes the system vulnerable to criticism. If the system is a failure, and testing reveals severe deficiencies leading to program cancellation, then testing has worked properly—a decision to discontinue a bad system is a good outcome (at least from the testing perspective).

So how can we ensure that programmatic consequences do not occur? As suggested in Section 2.3.1, the military needs to make sure that each system attribute meets specification to a reasonably high level of confidence. Problems will inevitably arise during testing, especially when programs are accelerated and rushed into operational testing, where the system is put into a full-fledged combat environment. The problems themselves do not mean termination for a system—in fact, they are unavoidable and are a natural part of a system's development. The key is to rapidly find which attributes are problematic, develop fixes, and retest those attributes in a very targeted way to ensure that they meet specifications. Some specifications will be met during the first round of testing; others will take iterations of testing and fixing before the specification

is met. Regardless, the specification must be met such that the tester is confident that results will withstand the inevitable public scrutiny that comes with any weapons system.

Throughout the cases, you will see the interaction of testing and programmatic consequences, and how both systems experienced different levels of testing that led to very different programmatic outcomes. The DIVAD's testing program was limited at best (and often misdirected), and, when the public opinion firestorm came, testing evidence was unable to withstand Congressional and public inquiry, leading to system cancellation. The M-1, on the other hand, also received a great deal of public criticism, which the Army was well equipped to handle through voluminous testing data and a targeted testing program.

3.1 Background on the M-1 and DIVAD

The M-1 is a 70-ton main battle tank, heralded for its overwhelming battlefield dominance during the First Gulf War, and widely known as the world's premiere tank (Figure 3-1).[18] The DIVAD, on the other hand, an anti-aircraft gun meant to be the equivalent of the Soviet ZSU-23-4, is a classic example of abject failure in weapons development, and was perceived to have such a multitude of technical problems that only 50 out of a proposed 618 were produced (Figure 3-2).[19]

Figure 3-1: The DIVAD.[20]

[18] Babcock, Charles R. "M-1 Tanks Quickly Demonstrate Superiority; Iraqis' Soviet-Built T-72s Prove to be Outmatched So Far in Desert Battlefield Engagements." 1991.
[19] Ditton, Major Michael H., *The DIVAD Procurement: A Weapon System Case Study*. 1988, pp. 3-8.
[20] Office of the Secretary of Defense. <http://dodimagery.afis.osd.mil/imagery.html#

Figure 3-2: The M-1 Tank.[21]

What is truly remarkable, considering the outcomes, are the similarities between the systems before they started development. Both were conceived in the early and mid-1970's, and were seen as urgently needed systems. One general, in charge of Army's weapons acquisition programs, stated that "there is no substitute…for a system of the DIVAD type,"[22] as it was crucial to the Army's air defense against the rapidly expanding Soviet force of helicopters and low-flying planes. Even the Secretary of Defense at the time, Caspar Weinberger, felt that the DIVAD was essential to the forces in Europe,[23] and congressmen roundly agreed that the system was a very high acquisitions priority.[24] The M-1 was also seen as a top priority, with Senator Stratton, the chairman of the Senate Armed Services committee, stating that there was an "urgent strategic necessity to get the best possible tank in the largest possible numbers to our forces in the field at the earliest possible date,"[25] and with General Starry, commander of the Army's Training and Doctrine Command, saying that the M-1 was "the single most important thing [the Army] can do to the battlefield to improve our capability."[26]

---

guid=9708eec69f0a774b68e8ccb3bffdecf61a575744>

[21] How Stuff Works. *How Future Combat Systems Will Work*. <http://science.howstuffworks.com/fcs1.htm>.

[22] House of Representatives Appropriations Committee. *Department of Defense Appropriations for 1984*, 1983. 823.

[23] Senate Committee on Armed Services. *Oversight on the Division Air Defense Gun System (DIVAD)*,1984. Page 4.

[24] See for example Senator Stratton's comment that "We have got to replace the quad [.50 caliber machine guns] from World War II before long," Ibid.

[25] House of Representatives Committee on Armed Services. *Oversight Hearings on the Status of the Army XM-1 Tank Program*,1977. 1-6.

[26] House of Representatives Committee on Armed Services. *Hearings on Military Posture and H.R. 10929 Department of Defense Authorization for Appropriations for Fiscal Year 1979, Part 2: Procurement of Aircraft, Missiles, Tracked Combat Vehicles, Torpedoes, and other Weapons*, 1978. 88.

Moreover, the systems shared several similarities in terms of the operational setting that defined their requirements. The two were meant to operate in tandem, with the M-1 destroying enemy ground forces while the DIVAD provided aerial cover.[27] As such, both were designed as track-based armored vehicles, and were meant to fight the Soviet combined arms forces that would come pouring across the Fulda Gap in Germany if hostilities broke out. The DIVAD's worth in a potential conflict was even measured in the $4 billion in M-1 tanks and M-2 infantry fighting vehicles it would supposedly save through the anti-aircraft umbrella it would provide; ads that Ford Aerospace ran in the papers boasted that the DIVAD would protect ten times its worth in Army equipment.[28]

Finally, the strategies for acquiring the systems were remarkably similar. The acquisitions processes for both systems were designed to compress and reduce costs for the systems' product development cycles by shortening testing and giving maximal freedom to contractors. Each system's program had two contractors produce a prototype with a "skunk works" approach,[29] in which the contractors saw very little government oversight. Both contractors had to produce a prototype for a Development/Operational Test, which was to be a "shoot off" to see which system would undergo full development. Additionally, both systems were created using a "concurrent development" strategy, where production commenced on the systems while the systems were still under development. Fixes were sped along by cost-to-contractor warranty provisions, which dictated that contractors retroactively had to pay the full cost of fixing any system failures.[30]

3.2 Alternative Explanations of Failure

There are many potential explanations beyond testing for why the DIVAD failed while the M-1 succeeded. While I do not discuss them again, they are worth mentioning:

---

[27] House of Representatives Committee on Armed Services. *Status of Army Air Defense Planning*. 1980.
[28] House of Representatives Appropriations Committee. *Department of Defense Appropriations for 1984*. 1983. 825. Also Senate Committee on Governmental Affairs, *Oversight of the Sgt. York (DIVAD) Air Defense Gun and the DSARC Decisionmaking Process*, 1984. 11.
[29] Ibid.
[30] Senator Wilson's comments. Senate Committee on Armed Services. *Oversight on the Division Air Defense Gun System (DIVAD)*. October 2, 1984. 7. See also Senate Committee on Governmental Affairs, *Management of the Department of Defense Hearing: Oversight of the Sgt. York (DIVAD) Air Defense Gun and the DSARC Decisionmaking Process*. 1984. 43.

- The Army gave too little oversight to the DIVAD, allowing the contractor to make too many mistakes without government intervention. The M-1 saw a great deal more government involvement in everything from engineering to manufacturing.[31]
- Ford Aerospace had a lot of company-level problems during the development of the DIVAD, and the company's downfall may have harmed the DIVAD's viability as a program. The quality of the DIVAD could have been reduced because Ford Aerospace may not have had the necessary technical savvy to produce the DIVAD (I discount this hypothesis later because of consistent Army statements to the contrary).[32]
- There may have been very little organizational impetus to develop the DIVAD—despite its words to the contrary the Army may not have felt like it desperately needed the system. This hypothesis is supported by two primary arguments: First, the Army put a Major General in charge of the M-1 program, while the DIVAD project was only assigned a Colonel. Second, the Army doctrinally has always preferred to focus on ground combat, meaning that an air defense gun may have gotten less visibility and funding.
- The concurrent development strategy may have lent itself better to tank development, which had a great deal of organizational and technological precedence, than to air defense development, which is presumably a more difficult task and more prone to system failures.

While I by no means discount these explanations, this thesis privileges a testing and programmatic consequences explanation as the reason for the DIVAD's downfall and the M-1's success. Testing without a doubt had a large effect on the viability of the DIVAD and the M-1 programs, and, as you will see below, appropriate testing strategies are crucial to save a system from inevitable negative attention.

---

[31] Moyer, Raphael. "The Army, Contractors, and Weapons Development: Evidence From Congressional Hearings." Working Papers, 2009.
[32] Conversation with Christopher Clary, MIT PhD candidate, Security Studies Program.

3.3 The Division Air Defense System (DIVAD) and Programmatic Consequences

Before discussing the DIVAD, it is important to note that I assume that the Army did not lie or attempt to mislead Congress during testimony, despite their organizational interests. Even as disastrous media reports were coming out about the DIVAD, senior Army officers involved in the program consistently proclaimed that the DIVAD was on track and operational. General Louis Wagner, deputy chief of staff for the Army's Research, Development, and Acquisition office, even went so far to say that "we see no showstoppers at this time" months before the system was cut.[33] Additionally, an Army Captain on the testing crew, Senator Goldwater, and an Army senior noncommissioned officer (NCO) all made statements in committee indicating that the system was performing very well.[34] If the Army were lying about systems performance, and the DIVAD was in fact an unmitigated disaster as the media reported, than the termination of the DIVAD would actually have been a good outcome, as it would have saved the Army from wasting billions procuring additional systems.

3.3.1 The DIVAD Program and Testing

To start engineering development, the government gave two contractors, Ford Aerospace and General Dynamics, a list of systems specifications and told them each to produce a prototype within 29 months, at which point the Army would conduct a "shoot-off" to determine which system would go into full-scale engineering development to work out the system's kinks.[35] This development strategy was meant to compress the development cycle, getting systems to the field faster and taking the maximum advantage from current technology, all while saving a purported

---

[33] House of Representatives Committee on Armed Services. *Defense Deapartment Authorization for Appropriations for Fiscal Year 1986*, 1985. 405. See also Van Voorst, Bruce, and Amy Wilentz. "No More Time for Sergeant York." *Time* 9 Sept. 1985.

[34] Ibid., 452. See also Senate Committee on Governmental Affairs, *Management of the Department of Defense Hearing: Oversight of the Sgt. York (DIVAD) Air Defense Gun and the DSARC Decisionmaking Process*. 1984. 2.

[35] House of Representatives Committee on Armed Services. *Status of Army Air Defense Planning*. September 30, 1980; House of Representatives Committee on Armed Services. *Hearings on Military Posture and H.R. 10929 Department of Defense Authorization for Appropriations for Fiscal Year 1979*. February-April 1978; Senate Committee on Governmental Affairs. *Management of the Department of Defense Hearing: Oversight of the Sgt. York (DIVAD) Air Defense Gun and the DSARC Decisionmaking Process*. September 28, 1984. 4.

five years and $1 billion in development.[36] To save in costs and development time, the DIVAD was to be created with predominantly off-the-shelf components, like the M-48 tank chassis and drivetrain, a preexisting radar, and a proven gun system.[37] Ford and General Dynamics each produced generally similar systems for the June 1980 competition (officially called Development Test/Operational Test II, or DT/OT II), with large differences only in the radar (Ford used a Westinghouse modified F-16 fighter jet radar, while GD used an in-house radar designed for the Phalanx air defense system), and in the type of gun (Ford used twin 40mm Bofors guns, while GD used twin 35mm Oerlikon guns).[38] In a controversial decision, the Ford prototype was chosen.[39]

Because of the off-the-shelf nature of the components used, the program made a decision to speed through the early testing stages in order to get the prototype into operational testing as quickly as possible. Once the "shoot-off" was complete, the Army immediately bought several prototypes from the winner and begin testing them (as part of the concurrent development "test-fix-test" approach), attempting to create fixes for deficiencies found in testing just in time for the next production round. To this end, the Army planned to procure 12 systems from Ford a short six to eight months after the shoot-off was completed.[40] These systems were produced by Ford as part of "Option I" of the contract, which allotted for production of 50 vehicles, before fixes for deficiencies had been established and tested.[41] The Army fully recognized the risks associated with a limited testing strategy, but thought it best to proceed with the program development to cut costs and shorten the schedule, without thought to the risk of programmatic consequences.[42]

Between November 1981 and January 1982, the DIVAD gun "check test" was conducted on the first production units, to ensure that progress was indeed being made on the system. In the

---

[36] House of Representatives Committee on Armed Services. *Hearings on Military Posture and H.R. 1872 and H.R. 2575 and H.R. 3406*, 1979. 545. Also Senate Committee on Governmental Affairs. *Management of the Department of Defense Hearing: Oversight of the Sgt. York (DIVAD) Air Defense Gun and the DSARC Decisionmaking Process*. September 28, 1984. 29-36.

[37] Ibid., 43.

[38] Senate Committee on Armed Services. *Oversight on the Division Air Defense Gun System (DIVAD)*. October 2, 1984. 6-10.

[39] Senator Wilson's comments. Senate Committee on Armed Services. *Oversight on the Division Air Defense Gun System (DIVAD)*. October 2, 1984. 7. Also Senate Committee on Governmental Affairs. *Management of the Department of Defense Hearing: Oversight of the Sgt. York (DIVAD) Air Defense Gun and the DSARC Decisionmaking Process*. September 28, 1984. 43.

[40] House of Representatives Committee on Armed Services. *Hearings on Military Posture,* 1979. 545.

[41] Senate Committee on Armed Services. *Oversight on the Division Air Defense Gun System (DIVAD)*. October 2, 1984. 5.

[42] House of Representatives Committee on Armed Services. *Defense Deapartment Authorization for Appropriations for Fiscal Year 1986*. 1985. 415.

Fiscal Year 1983 Appropriations Hearing, Major General Maloney, chief of Army weapons system development, stated that in the check test the system achieved 11 out of 12 requirements (it did not properly conduct identification of friendly units) capable of meeting threats from helicopters and low-flying aircraft. While it appeared to the Army that the check test went well, the data appeared at best inconclusive to outside sources—the Government Accountability Office (GAO) reported that the test "indicated that deficiencies...disclosed during [DT/OT II] were not overcome as required," and, perhaps worse, a planned 7-month reliability test was greatly reduced in scope. The Army may have thought that the system's development was going well, but was largely unable to support its claims with hard data.

There were also consistent claims made in 1984 that in a May 1982 meeting the Army was inadequately reporting data to the rest of the Department of Defense, largely because the Army made poor efforts to interpret and integrate results.[43] The Army was certainly making efforts to test the DIVAD, but even in the early stages the test plan was simply not robust enough to silence the system's detractors. Despite the criticism, Option II, which meant the production of an additional 96 systems, was scheduled for May 31, 1983, bringing the total amount allotted for production to 146 systems.[44]

The next major round of testing, conducted on low-rate production units, occurred in the form of the Design Verification Test (DVT) and Limited Test (LT) in mid-1984. The DVT was conducted on the first production unit, and was a combined government and contractor test meant to prove that the system met 26 contractor performance specifications and ten government performance characteristics. While 18 of the contractor specifications were met, three were not, which included the key performance attributes of target classification with the search radar, system reaction time, and hit probability against non-maneuvering aircraft (five results had not been fully analyzed by the time of the GAO report). Additionally, the DIVAD had difficulties in performing in cold chamber testing (at 35° F below zero). As the GAO reported, the Army Materiel Systems Analysis Activity, which evaluated the DVT results, gave an overall impression that the Army was "generally pleased with the results and that they were better than

[43] Senate Committee on Armed Services. *Oversight on the Division Air Defense Gun System (DIVAD)*, 1984. 60.
[44] House of Representatives Committee on Armed Services. *Defense Deapartment Authorization for Appropriations for Fiscal Year 1983*, 1982. 18-45. Also House of Representatives Appropriations Committee. *Department of Defense Appropriations for 1984*, 1983. 818.

or equal to the prototype performance."[45] The DVT did what it was supposed to do by identifying key system attributes that needed further fixes before the next testing round. While this kind of broad based developmental testing, where several attributes were tested in isolation, probably ought to have been conducted on prototypes before low rate production commenced, it was the best possible solution within the concurrent development framework.

The Limited Test in July and August of 1984, which interrupted the DVT for one month, was a completely different story. Where the DVT was meant to test each system specification in a rigorous fashion, the LT sent the DIVAD into an operational testing environment, attempting to test the system in the field before each attribute was proved out independently. The LT was directed at the behest of the Secretary of Defense, who was worried about the DIVAD's progress. According to the GAO, the purpose of the test was to obtain a "quick short look at the system's operational suitability to engage aircraft/ground targets, mobility in accompanying ground forces it should protect and inter/intra system communication."[46] The five-day test assessed system performance in a holistic sense, and the results showed favorable performances on many attributes, including identification of friend or foe (an issue from a previous test), probability of hitting moving targets, and ability to track and engage many targets at once.[47]

However, there were several failures that caused sensational media claims that were to bring about the system's downfall (discussed in the next section). Compounding the problem was the fact that of the 15,000 rounds the DIVAD was supposed to fire and of the 4,000 miles the system was supposed to travel, it only fired 3,600 rounds and traveled 300 miles, hardly enough to convince detractors that the testing data was representative of actual system performance.[48] Indeed, the Army did not set itself up for success, as it used poorly trained crews, and the mock-ups it used to represent helicopter targets could not operate the rotors fast enough to allow the Doppler tracking radar to lock on. Several testing results also became key areas of contention as the Inspector General was called in to investigate a complaint that the Army grossly misrepresented data.[49] The DIVAD simply was not ready for a full operational test, and inadequate data to support Army claims made it impossible for the Army to adequately refute

---

[45] Senate Committee on Armed Services. *Oversight on the Division Air Defense Gun System (DIVAD)*, 1984. 52-55, 13.
[46] Ibid., 55-60.
[47] Ibid.
[48] Ibid.
[49] Ibid., 37-42.

critics—had the Army maintained discipline with the testing plan and refused to throw the DIVAD into operational testing, the program might have had a better chance at avoiding programmatic consequences.

A final round of testing in early 1985, designated the Initial Production Test, had results that showed the fire control system and power and actuation systems to be unreliable, with lower than desired mean hours between failure. As General Wagner was keen to note, the sample sizes within the tests were small at best, as the vehicles only had limited operation time, and the mean hours between failure figures could simply have been a statistical anomaly (in fact, a lot of the failures were directly linked to maintenance errors that had nothing to do with system performance). However, these results still detracted from the Army's case for the DIVAD, as they portrayed it as unreliable and showed that little improvement had been made over the year between testing results, even though the failure might have been a statistical fluke.[50]

3.3.2 The Media, Congress, and Programmatic Consequences

Throughout the media storm that was to engulf the DIVAD, the Army and many in Congress both realized a desperate need for the DIVAD and remained confident that the DIVAD would turn out to be a great system. The system's operators thought the DIVAD commendable, and had excellent results tracking and destroying targets; Senator Goldwater, who personally fired the DIVAD, thought it "an impressive piece of equipment;" General Wagner thought the system ready to proceed with production after the 1984 test results.[51] In fact, many in the Army thought that it was simply the DIVAD's "turn in the barrel," as many other systems, the M-1 Abrams included, had experienced a great deal of media criticism while deficiencies were being corrected.[52]

When the Army did not have sufficient testing data to back up its claims that the system would be successful given more time and that positive progress was being made in correcting deficiencies, it could not stave off the criticism that would lead to the system's demise. As

---

[50] House of Representatives Committee on Armed Services. *Defense Department (sic) Authorization for Appropriations for Fiscal Year 1986*. 1985. 403.
[51] Ibid., 452. Senate Committee on Armed Services. *Oversight on the Division Air Defense Gun System (DIVAD)*, 1984. 1-40.
[52] House of Representatives Committee on Armed Services. *Defense Department (sic) Authorization for Appropriations for Fiscal Year 1986*. 1985. 400-410.

Senator Goldwater said, "I am reminded of a flock of vultures hovering around a dying steer."[53] A sampling of the most poignant criticisms are listed here:

- Media: During the 1984 limited test, the DIVAD mistook half of the 180 decoys sent against it for real targets, and proved incapable of operating against foes using advanced helicopter tactics or electronic countermeasures. As the Washington Post article said, "success was only achieved after several radar reflectors were attached to the target."[54] The Army argued that these test results were misinterpreted and were colored by the availability of only two test vehicles, one of which was sited poorly (the other vehicle had very good hit percentages). Additionally, the crews had only received very limited training.[55]
- Media: The press reported that the DIVAD targeted the fan on an out-house, causing a public sensation. The Army told a very different story to Congress: the system had targeted the fan on the side of a building, as it should have, because it is designed to pick up signatures from helicopters very far away. However, this phenomenon happens often on many systems, and the crew is trained to quickly ignore such signatures once they realize they are not a threat (which the crew did in this instance).[56]
- Media: A rumor developed that the system opened fire at a stand full of military officers and senators who were watching a test.[57] While untrue, the rumor caused a great amount of damage as a result.
- Media: Numerous allegations arose that several Army officers involved in the acquisition of the DIVAD had left the government and taken jobs at Ford Aerospace.[58] While true, these had nothing to do with the performance of the system itself, but the perception of fraud was damaging to the program.
- Congressional: Representative Denny Smith of Oregon, a known detractor of the system from the inception of the program, stated in 1985 that the Army had blown up targets to

---

[53] Senate Committee on Governmental Affairs. *Management of the Department of Defense Hearing: Oversight of the Sgt. York (DIVAD) Air Defense Gun and the DSARC Decisionmaking Process*. September 28, 1984. 2.
[54] Ibid., 15-38.
[55] Ibid.
[56] House of Representatives Committee on Armed Services. *Defense Deapartment Authorization for Appropriations for Fiscal Year 1986*. 1985. 432.
[57] Conversations with Christopher Clary, MIT PhD candidate at the Security Studies Program.
[58] Ditton, Major Michael H., "The DIVAD Procurement: A Weapon System Case Study."

make it appear as though the system was working properly. The Army said that the targets had strayed out of the test area and were blown up for safety concerns.[59]

- Congressional: While Senator Andrews was being shown the system in operation, the main gun jammed, causing him to become pessimistic about the system. Sen. Andrews also noted that the vehicle had problems operating in extreme cold. The gun jamming, the Army purports, was a procedural error by the non-commissioned officer operating the vehicle.[60]

Quickly, the DIVAD became the favored target for a media and Congress eager to cut the Defense budget and reduce corruption. Rumors about the system began to spread, and it became an example to the public of acquisitions gone wrong. The Army had reasonable responses for the allegations, but they lacked credibility, as there was no rigorous testing data to prove the DIVAD's operational capability. There were simply too few rounds fired and too few miles driven to prove that the DIVAD could actually perform on the battlefield. Worse, the Army could not demonstrate that it had properly implemented the test-fix-test approach, as the 1985 testing data appeared to the untrained eye to show that the system was still unreliable several years after the first shoot off (in fact, the 1985 results may very well have been the result of statistical anomaly and the test setup).

The final result was the worst possible programmatic outcome: the cancellation of the DIVAD program by Secretary of Defense Weinberger in the fall of 1985.[61] The impact of the DIVAD failure can be measured in many ways: the DIVAD was supposed to save $4 billion in assets if the Cold War had broken out;[62] the DIVAD program itself was worth $4.2 billion, and its termination meant that all of the money already paid out to the contractor was forfeited;[63] the Army failed to obtain the capability to conduct close in air defense, a capability it lacks to this day; the Army lost a great deal of prestige as a result of the DIVAD's failure. While the goal of speeding the program along and reducing costs was noble, because of testing failures it ended in the worst possible outcome. The final result was much more expensive to the Army than what

---

[59] Keller, Bill. "Weinberger Faces Decision on Fate of $4.5 Billion Antiaircraft Gun." New York Times, 1985.
[60] House of Representatives Committee on Armed Services. *Defense Department Authorization for Appropriations for Fiscal Year 1986*. 1985. 397-420.
[61] Van Voorst, Bruce, and Amy Wilentz. "No More Time for Sergeant York." *Time* 9 Sept. 1985.
[62] House of Representatives Appropriations Committee. *Department of Defense Appropriations for 1984*. 1983. 825.
[63] Senate Committee on Armed Services. *Oversight on the Division Air Defense Gun System (DIVAD)*. 1984. 15-38.

the Army could have hoped to save with the testing-light concurrent development strategy. In short, the Army did very little to mitigate potential programmatic consequences, and paid huge costs that might have been averted with a superior test plan.

3.4 The M-1 Abrams Tank and Programmatic Consequences

The DIVAD case shows the potential programmatic consequences that can result from a poor test plan. The M-1 case, by contrast, is shows a system with a rigorous test plan that was able to largely avoid programmatic consequences despite criticisms. Instead of going through an in-depth history of the M-1 testing cycle, as we did for the DIVAD, we will instead conduct a more abbreviated analysis that will look through the M-1's testing history to uncover the testing strategy that led the M-1's success. As we go through the case study, two things will become readily apparent that distinguish the M-1 from the DIVAD:

- Where the DIVAD testing program repeatedly did not perform rigorous enough testing to obtain a sufficient amount of data points and tight confidence intervals, the M-1 testers conducted massive amounts of testing above and beyond the test plan to ensure that results could not be contested.
- When problems were found in a specific system attribute, and media and congressional criticisms began to arise, the Army quickly ordered very extensive testing (after a fix was implemented) on that one attribute to gather more data and prove that the attribute did indeed meet specifications. This stands in stark contrast to the DIVAD, where fixes were tested by throwing the system into a full-on operational test for which the system was not prepared.

This section proceeds in three parts. First, I will give a very brief background on the development of the M-1 program to give the reader context for the testing plan. Next, I will discuss the rigor of the M-1 testing, as described above, showing that testing was much more extensive for the M-1 system than for the DIVAD. Finally, I will go over how the Army test planners reacted when system components started failing, to show how the adaptive nature of the test program allowed the Army to rapidly shift testing to problem areas.

3.4.1 The M-1 Abrams Program

In February-May of 1976, two companies, Chrysler and General Motors, entered into a competition (called Development/Operational Test I) with fully operational (but unpolished) prototypes for the M-1 tank. Chrysler's tank, with an under-developed turbine engine, won the competition largely on the basis of its significantly better mobility.[64] After the selection of the Chrysler tank, the Army directed Chrysler to produce 11 pilot vehicles between February and July, 1978, one and a half years after the completion of the competition (as compared to the six to eight months the Army gave Ford Aerospace).[65] Chrysler delivered these vehicles on time, and the Army and Chrysler used the vehicles in Development/Operational Test II (DT/OT II) from May 1978 to February 1979. As with the DIVAD, the M-1 used a concurrent development strategy that started low-rate production before the prototypes were fully tested—even with seven months of testing left, the Department of Defense approved the execution of a production option that meant the creation of an additional 110 tanks in 1979.[66] In FY1982 Congress approved production 665 tanks, and to this date over 9,000 have been produced for the U.S. and several other countries.[67] The M-1 is now widely recognized as one of the premiere battle tanks in the world.

3.4.2 The M-1's Testing Rigor

Like with the DIVAD, DT and OT II happened concurrently. As mentioned above, however, the M-1 distinguished itself from the DIVAD by the incredible rigor of the testing program. As soon as the Army obtained its first order of 11 pilot vehicles between February and July 1978, the tanks were thrown into DT/OT II. The development test occurred in locations all

---

[64] Turbine engines provide a "quantum jump in battlefield mobility," are able to operate better in the cold, and also has no smoke plume that can give away its position. House Committee on Armed Services, *Oversight Hearings on the Status of the Army XM-1 Tank Program*. March 25, 28 1977. 38-41.
[65] Ibid., 44.
[66] Senate Committee on Armed Services. *Department of Defense Authorization for Fiscal Year 1981, Part 5: Research and Development*. 1980. 2767.
[67] Senate Committee on Armed Services. *Department of Defense Authorization for Appropriations for Fiscal Year 1983, Part 4: Tactical Warfare*. February-March 1982 2329. Also see GlobalSecurity.org, "M1 Abrams Main Battle Tank."

across the US, and Chrysler and the Army did extensive testing across nearly all systems attributes:

- Mobility testing (including obstacle crossing, acceleration, braking, fording, cross country mobility, maximum speed, slope speed, and fuel economy)
- Weapons testing (tracking, hitting, and destroying targets both moving and stationary)
- Vulnerability testing (against anti-tank missiles, small and large caliber direct fires, mines, and overhead explosions)
- Adverse conditions testing (ability to withstand electromagnetic pulses, and extreme weather conditions both hot and cold)

The operational testing was also very rigorous. The Army incorporated a five tank M-1 platoon into a company of older M60 tanks, and ran the company through combat exercises that were primarily designed to test survivability, "fightability" (or how easily operable the vehicle is in battle), targeting, mobility, reliability, availability, maintainability, and durability (RAM-D), and human factors. All of these tests went above and beyond the testing plan in terms of number of miles traveled and rounds fired, and are in stark contrast to the scant 3,600 rounds fired and 300 miles traveled (out of the planned 15,000 rounds fired and 4,000 miles traveled) that the DIVAD experienced during 1984 operational testing. The M-1 Operational Test portion, for example, saw 4,689 rounds fired and 19,097 miles driven, compared to the goal of 2,500 rounds and 12,500 miles driven. Contractor testing and the Development test portion added 12,480 rounds fired and 49,000 miles driven—numbers that, in contrast to the DIVAD, are staggering (See Figure 3-3).[68]

[68] House of Representatives Committee on Appropriations. *Department of Defense Appropriations for Fiscal Year 1980*. May 1, 1979. 4-10.

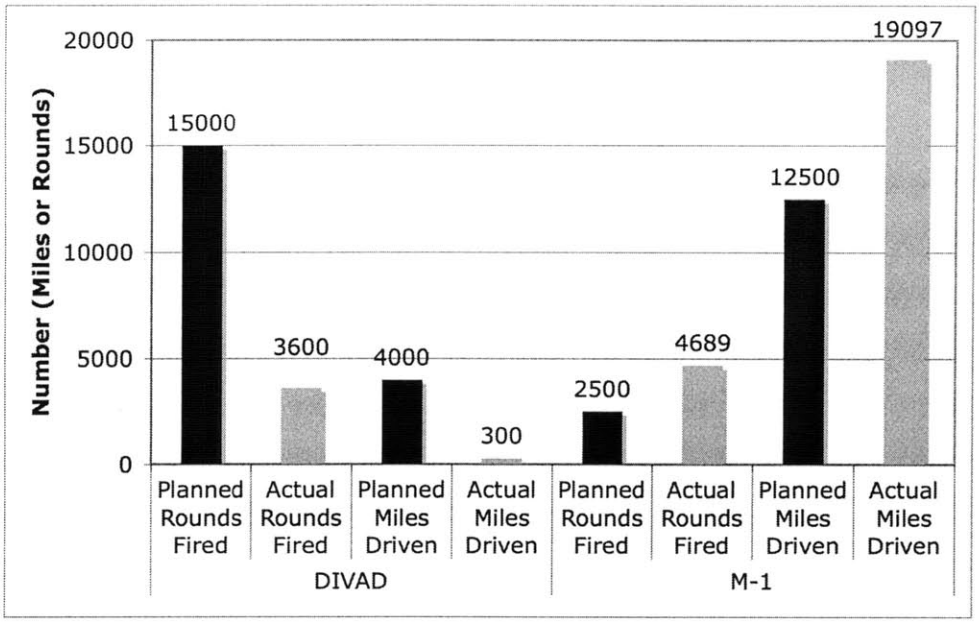

Figure 3-3: Comparison of DIVAD and M-1 Test Plans in Operational Test II.

Development Test/Operational Test III, completed in late May of 1981, was similarly rigorous. Conducted at five facilities across the U.S., including Aberdeen Proving Grounds, White Sands Missile Range, New Mexico, the Cold Region Test Center in Alaska, Ft. Knox, Kentucky, and Ft. Hood, Texas, the test saw 54 vehicles travel an additional 115,000 miles and shoot another 21,000 rounds in a variety of weather and terrain conditions. These tests spanned the same attributes as in the previous DT/OT II round, and concentrated on survivability, firepower (and especially firing on the move), mobility, RAM-D, and troop acceptance (the reception the vehicle received with troops). Because of the vast amount of testing conducted, the Army could confidently show numbers that supported claims validating the M-1's accelerated production schedule. For example, the tank was 50 mean miles between failure above specification for combat reliability and had a higher than required vehicle life. These tests, while expensive, were essential in proving the M-1's operational readiness.

Though the tank proved deficient in some areas, like power train durability, the program management team was able to leverage its performance in executing fixes and validating those fixes with test data in order to convince the public and Congress that future fixes would indeed work. Anecdotal evidence was also very helpful in supporting the claims that the M-1 was on track: Major General Lawrence, commander of the 1st Cavalry Division, for example, reported

that his tank crews were consistently hitting targets over a mile away while traveling at 25 miles per hour; three tanks purportedly had eight large caliber rounds fired at them with no effect on performance or fightability.[69]

When the results of the developmental and operational tests were presented to the public and Congress, there could be little doubt that they were accurate and reliable, as they had the support of voluminous data and rigorous test methods. Criticisms of the system inevitably arose, but their claims could be easily refuted. While the testing was very broad based and robust, special pains were taken to conduct even more extensive testing on specific systems attributes that were known to be problematic.

3.4.3 Targeted Testing and Validation of Fixes

The Army established three primary areas of technical concern during the DT/OT II testing of the M-1, including the turbine engine's ability to operate in dusty environments, hydraulic system leakage, and track retention (the track would fall off in deep sand), and work began immediately to correct them.[70] Congress, the press, and other critics were especially worried about the turbine engine's maturity and dust ingestion (Congress even funded a diesel engine program in case the turbine did not work out, against the Army's wishes), but the Army was able to fix the turbine's problems by reengineering several parts.[71] The fixes were validated extensively—test planners realized that the turbine engine could become a huge barrier in the tank's development, and adapted the test plan accordingly. In 1976, despite heavy criticisms from Congress, the Army requested an additional $30 million in order to conduct 30,000 hours of testing specifically on the turbine, far above the 4,400 recommended by the contractors.[72] The "engine maturity program" introduced fixes such as new air filters, seals, and fuel injection, and had completed 5,400 hours of testing in a lab and 20,000 miles driven by May 1979, with a goal

---

[69] House of Representatives Committee on Armed Services. *Status of the Army Tank Program*, November 5, 1981. 1-20. See also House of Representatives Committee on Appropriations. *Department of Defense Appropriations for Fiscal Year 1980*. May 1, 1979. 4-20.

[70] House of Representatives Committee on Appropriations. *Department of Defense Appropriations for Fiscal Year 1980*. May 1, 1979. 4-6.

[71] House of Reps. Committee on Armed Services. *Status of the Army Tank Program*, November 5, 1981. Page 9.

[72] House Committee on Armed Services, *Oversight Hearings on the Status of the Army XM-1 Tank Program*. March 25, 28 1977. 70.

of 8,530 hours and 36,000 miles by November 1979.[73] By mid-1981, the dust ingestion problem was fully solved, and the vehicle was tested both in desert conditions at 20 times zero visibility at Yuma Proving Grounds and driven 30,000 miles at Ft. Knox with no malfunctions. In FY1983 hearings, the commander of Army Training and Doctrine Command, General Otis, confidently stated that there had been no dust ingestion failures since 1979, and that they problem would not crop up again.[74]

Track retention and hydraulic fixes were also implemented quickly and then tested. The track retention problem was discovered because of the deep, moist sand environment at Ft. Bliss, and the solution was to modify the design of the sprocket and hub assembly and to tighten the track tension. The hydraulic leakage caused the loss of pumping fluid, resulting in the overheating and failure of the main and auxiliary hydraulic pumps, and the fix was to redesign fittings with better tolerances, and to shorten and reroute hydraulic lines. In addition to the original test plan, five pilot vehicles were given fixes, and each was tested a total of 6,000 miles to ensure that the problems had been solved (the total amount of miles driven to test for these specific problems were 100 times more than the DIVAD received in 1984 operational tests). The problems did not return again with future vehicles, as the fixes were implemented in time for the wave of production that followed pilot vehicle production.[75]

The Army also decided early in development that survivability was the first priority for the M-1 tank. In the context of programmatic consequences, this makes a lot of sense—while track throwing and hydraulic leakages could presumably cause programmatic delay or, if not fixed, program cancellation, there is little that could turn public opinion more rapidly than perceived deficiencies in armor plating that could put U.S. troops at risk. To test the extensive crew protection measures on the tank, the Army engaged in a very elaborate set of tests against both individual system components like armor plates and against a fully loaded tank. These tests included small arms, large caliber tank rounds, antitank missiles, and antitank mines.[76] When all

---

[73] House of Representatives Committee on Appropriations. *Department of Defense Appropriations for Fiscal Year 1980*. May 1, 1979. 4-8.
[74] House of Representatives Committee on Armed Services. *Status of the Army Tank Program*, November 5, 1981. 31. See also Senate Committee on Armed Services. *Department of Defense Authorization for Appropriations for Fiscal Year 1983, Part 4: Tactical Warfare*. February-March 1982. 2338.
[75] House of Representatives Committee on Appropriations. *Department of Defense Appropriations for Fiscal Year 1980*. May 1, 1979. 5-7. See also Senate Committee on Armed Services. *Department of Defense Authorization for Appropriations for Fiscal Year 1983, Part 4: Tactical Warfare*. February-March 1982. 2337-2339.
[76] House of Reps. Committee on Armed Services. *Status of the Army Tank Program*, November 5, 1981. Page 6.

was said and done, 3,500 rounds had been fired at the tank, and General Otis could confidently state that the M-1 was "the most survivable tank in the world," with 2.5 to 4 times the survivability of the M-60 (the previous Army main battle tank).[77]

As with the DIVAD, there was no lack of media of negative media attention for the M-1. They lambasted the turbine, which they thought to be an unreliable and technologically unsound, and broadly criticized the track failures and hydraulic system. However, because of the Army efforts to mitigate criticism through extensive testing, the M-1 was able to avoid the worst of the programmatic consequences.

3.5 Conclusions

The primary takeaway from the examples of the DIVAD and the M-1 tank is that testing is a shield with which services can defend themselves from programmatic consequences like program termination. Without rigorous testing that makes sure each attribute is within specification to high confidence, a project can easily fall prey to the media, congressional, and GAO criticism that follows any system, regardless of its merits. Military systems are simply too large and complex to be built problem-free—the key is to fix problems rapidly and to validate those fixes before media criticism can take control of program choices. In short, the DIVAD's haphazard test plan and execution did not provide enough ammunition to the system's proponents to allow it to survive in a hostile media environment, and premature operational testing fueled the system's critics; the M-1's test plan, with its voluminous amount of test data and targeted testing strategy for rapidly validating fixes, gave the system's critics very little to work with. The programmatic results reflected the test results: the DIVAD ended in cancellation, leaving the Army with no short range air defense gun to this day, while the M-1 became the premiere battle tank in the world, with production still ongoing.

---

[77] Senate Committee on Armed Services. *Department of Defense Authorization for Appropriations for Fiscal Year 1983, Part 4: Tactical Warfare.* 2339-2342.

4

Operational Consequences Case Studies

Operational consequences are consequences that result from the failure of a system attribute during a military operation that has disproportionately large negative effects. Even if the attribute's probability of failure is small, large consequences mean that the risk associated with the attribute will still be high. As discussed in Section 2.3, it is essential that the confidence intervals associated with the probability of failure of high consequence attributes are very tight, because the value of higher confidence is much greater for large consequences. It is thus very important to test specific system attributes above and beyond standard confidence intervals to ensure that they will not fail—if testers do not fully evaluate the potential operational consequences of each attribute's failure, the test plan they create might miss attributes that need extensive testing for a specific failure mode.

The following seeks to bring operational consequences to life by applying the operational consequences theory to actual cases. Each of the two case studies looks at a system (or, in the first case, a system of systems) and assesses it in three parts, first by giving brief background on the case itself, then by examining the operational consequences that could result from the system's failure, and finally by prescribing a rudimentary test strategy that could be used to mitigate that failure. The basic premise of testing for operational consequences is that if the tester is not absolutely sure that the probability of failure is low, he needs to conduct more testing.

Please note that the following is not meant to provide a full list of high stakes attributes for the studied systems. Such an endeavor would take a fair amount of analysis by looking at the system's intended and potential uses, and evaluating the impact should the system fail in any one of several failure modes. My sole intention in this section is to demonstrate a methodology for how failure mode analysis could be conducted, so that testers know which system attributes need special testing considerations. Both of my cases involve unmanned systems to give focus to the analysis.

It is important to note that analyses of attributes and test plans below are by no means comprehensive—I am not an expert on any of these systems—but are merely meant to present an example of the analysis methodology.

4.1 Case Study 1: Stopping SCUD Missiles with Sensor Networks

The following case study is adapted from both Chapter 14 of C.E. Dickerson and D.N. Marvis's *Architecture and Principles of Systems Engineering* (2009), and from C.E. Dickerson and Ricardo Valerdi's article "Using Relational Model Transformations to Reduce Complexity in SoS Requirements Traceability." In it, a network of UAVs and ground systems in communication with a central communications node[78] find, fix, and track SCUD launchers that are maneuvering in west-central Sudan, and may soon launch a weapon of mass destruction (WMD) at a sensitive target. It is analytically interesting in the context of operational consequences because while the objective of the system is inherently high stakes (the tracking of very important targets for potential destruction), each system component does not necessarily need the same testing rigor (that is, the probability of failure of some components is not correlated to the system of system's probability of mission failure).

4.1.1 The Sensor Network

As discussed above, the purpose of the sensor network is to track ground vehicles in order to keep tabs on a high-value target (in this case a SCUD missile launcher traveling with an entourage of other vehicles). The metrics associated with the mission, or key performance parameters (KPPs), are that the SCUD launcher is tracked for the maximum amount of time possible (to enable a strike), and that the system identifies the target vehicle amongst its escort vehicles with above 90% accuracy.

To execute this mission, there is a four-stage process in which the detection network:

1. CUEs the sensor system, or actually detects the enemy convoy

[78] In military parlance a Joint Intelligence Center (JIC), which collects information from the sensors, in conjunction with a Supporting Arms Component Commander (SACC) who calls in strike packages (that destroy the target).

2. FINDs the convoy's exact position through use of a Moving Target Indicator radar (MTI)
3. FIXes the target vehicle (the SCUD launcher) by identifying it with video imagery
4. TRACKs the target vehicle so that a lock is maintained on its position while a strike package moves in

For simplicity's sake, we can break the system down into four primary components, comprised of three sensor types and one communications node (see Figure 4-1):

1. The Supporting Arms Component Commander/Joint Intelligence Center (SACC/JIC), which is the command center for the sensor network, that calls back to higher headquarters for a strike package after the target is identified. Once the SACC/JIC picks up a signal from MIUGS (Micro-Internetted Unattended Ground Sensor, the CUE), it immediately sends a Global Hawk to FIND the position and a Predator to FIX the target. The SACC/JIC has communication links to all sensors and also has communications back to higher headquarters.
2. A number of MIUGSs which are small, immobile, ground-based acoustic and seismic sensors with detection ranges of one mile that can detect the approach of ground vehicles. These each have the ability to CUE the system by notifying the SACC/JIC that they are picking up heavy vehicle traffic. Because of their limited range and immobile nature, they must be deployed in an array.
3. The Global Hawk unmanned aerial vehicle, which has a 42-hour endurance and 14,000 mile range,[79] is equipped with a MTI, or Moving Target Indicator, that can FIND moving vehicles precisely and has a 62-mile range.[80] With its long loiter time, the MTI-equipped Global Hawk can TRACK the convoy for a long time. The Global Hawk sends its data back to the SACC/JIC through a data link.
4. The Predator unmanned aerial vehicle, equipped with an advanced video camera. The Predator's job is to FIX (identify) the SCUD launcher so that a strike hits the right vehicle.

[79] Available at airforce-technology.com's page on the Global Hawk, http://www.airforce-technology.com/projects/global/specs.html. Accessed May 13, 2010.
[80] Available at Unmanned-Aerial-System.com's page on the Global Hawk, http://www.unmanned-aerial-system.com/list/global_hawk_rq-4.html. Accessed May 13, 2010.

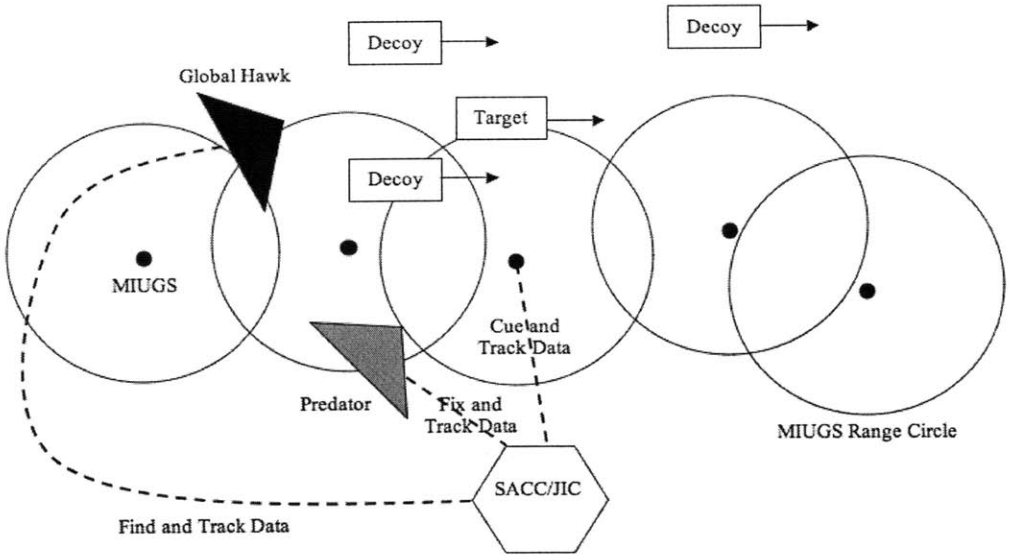

Figure 4-1: The Sensor Network Tracking a Moving Convoy.

The story goes as follows: A convoy approaches the MIUGS network. Once the convoy gets within range of one of the sensors, the sensor is triggered, CUEing the system, and sends back data to the SACC/JIC. If a Global Hawk happens to be in the area of the convoy, it can also CUE the system, but since there are so few Global Hawks the system is most likely to be cued by the MIUGS. The SACC/JIC then orders a Global Hawk to go to the area around the sensor, with a Predator following close behind. Once the Global Hawk gets within range, it uses the MTI to locate the convoy precisely (or FIND it), and sends that information back to the JIC, which forwards it to the Predator. The Predator then flies in and gets video of the convoy, FIXing the SCUD launcher amongst all the other convoy vehicles, and sends back the message of positive or negative ID to the JIC. If there is no SCUD launcher, the mission is aborted. If there is a launcher, the Global Hawk/Predator/MIUGS team continues to track the convoy until the strike package kills the target.

4.1.2 The Sensor Network and Operational Consequences

Thus far we have been looking at the sensor network in a fundamentally backwards way. Instead of looking at the sensor network as a system with many potential missions, and finding

which attributes could have operational consequences on those missions, we instead took the operational consequence (WMD launch) as a given and then examined the system. Before we continue, we should take a step back and view things from the perspective of the system developer as he is planning the test.

The sensor network system has two clear likely sets of uses. In a conventional war, the network would be used to detect the advance of an enemy mechanized formation, giving the friendly force the ability to better prepare for the attack and avoid being surprised. The identification capabilities of the system would allow friendly forces to readjust based on enemy force composition. While this is clearly a helpful advantage, the network's failure cannot be considered an operational consequence (assuming that the friendly forces do not completely rely on it), as the consequence of a network failure would not prove disastrous. While the friendly forces would lose some of their ability to detect surprise, the result might be thought of as a small decrease in fighting effectiveness—they would still fight on, and once contact is made the battle would proceed as normal. In short, without this system, friendly forces would be just as well off as they were before the system ever arrived, and it is doubtful that the system would have a disproportionate impact on the battlefield.

When unconventional warfare is considered, however, the picture changes, as the sensor network goes from being a nice capability to have to being essential to attacking hostile high-value ground vehicles. These ground vehicles could range from a jeep carrying a key insurgent leader (knowledge of whose movements intelligence provides) to a scenario similar to that in the case study, where a missile launcher or other weapons system is moving to wreak havoc on an allied city. For unconventional warfare, it is clear that mission failure for the system can have operational consequences. The killing of an insurgent leader could mean that an insurgent group is irreparably weakened, causing it to fall apart and greatly reducing the threat to friendly forces; the destruction of a WMD armed missile launcher could save countless lives, ensure that a shaky coalition does not fall apart (which might happen after a missile attack), and can deter future other would-be attackers. In short, a failure in finding and tracking moving ground targets, when those targets have a high value, does have consequences disproportionate to the action of tracking a vehicle, which is done frequently across the military for less valuable targets. The decision to label the moving target as a SCUD launcher in this case is merely an artifact of the available case study.

What makes this case somewhat unique in the discussion of operational consequences is that the system's mission itself is inherently high stakes. In other cases, like collateral damage from a guided bomb strike, the mission of killing insurgents is in itself not high stakes (it only has tactical consequences), but the failure mode of the guidance system that has a bomb hitting civilian areas is most definitely high stakes. In the following section, we discuss how to evaluate the failure modes of a system whose mission is inherently high stakes.

4.1.3 Attribute Analysis

If the sensor network's mission is inherently high stakes, how can we determine which system components need additional testing to ensure that operational consequences do not occur? While the mission is high stakes, the failure of any component does not necessarily mean mission failure. In fact, the failure of any one sensor may have little impact on the mission at all. In this section, I hope to present a rudimentary methodology for analyzing which system components need additional testing to ensure that operational consequences do not occur. In this method, the first step is to break the system down into its constituent parts. For a networked system of systems, this breakdown takes the shape of that in Figure 4-2, with the SACC/JIC and data links to the SACC/JIC as the network, Component 1 as the MIUGS, Component 2 as the Global Hawk, and Component 3 as the Predator. Below each component are several system attributes, each of which has a probability of failure and must be evaluated for operational consequences. For example, the Predator aircraft's attributes include a certain range, airspeed, and optical sensor resolution. After breaking down the network into its constituent components, we will evaluate each component independently and then evaluate the network itself to see which parts of the system of systems have failure modes that could have operational consequences. It is important to note that the attributes listed below are by no means a comprehensive list but are merely meant to present an example of the analysis methodology.

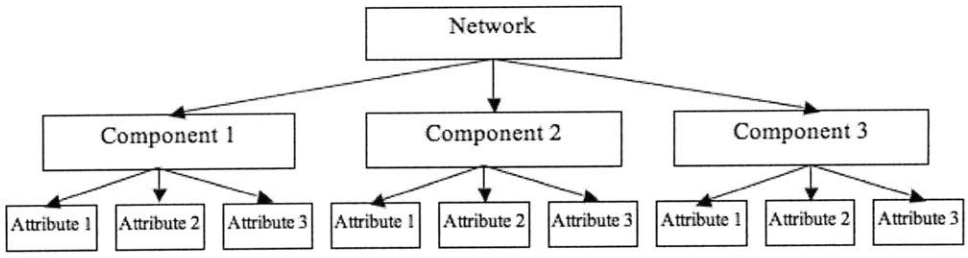

Figure 4-2: The Sensor Network Decomposed.

4.1.3.1 The MIUGS Sensor Array

The first component we evaluate is the MIUGS array. Because the MIUGS only has a range of one mile, using its acoustic and seismic sensors to pick up vehicle activity, we can assume that the sensors are spaced in a grid with overlapping fields of observation (see Figure 4-3 for an example of such a grid of 70 sensors with half-mile spacing between sensors and 22 square mile coverage area). The MIUGS sensor has four primary attributes that can vary based on the age of the MIUGS and manufacturing defects: the range of the sensor, the probability that the sensor registers a false positive (that is, there is no vehicle in range but it still says there is one), the probability that the sensor registers a false negative (there is a vehicle in range but the sensor does not register), and the mean hours between failure of the sensor (how long it takes before the sensor stops working).

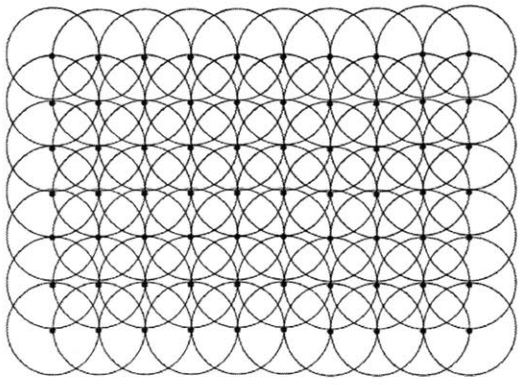

Figure 4-3: 70 Sensor MIUGS Grid with Half-Mile Spacing (22 Square Mile Coverage).

A failure on any of these attributes will not mean operational consequences, as long as the sensors are deployed in a robust grid as in Figure 4-3,[81] because the redundancies provided by the array of sensors mean that no one failure will mean an overall failure to detect the incoming SCUD launcher. Even in an extreme case, where 20% of sensors fail at one time, the network will still retain a very large percentage of its original coverage (see Figure 4-4), especially because the SCUD launcher will be moving through the sensor array, and thus will be hitting many sensors' area of observation before reaching a launching site. The false positive attribute also will not be a problem—even if several sensors are out of specification and have a tendency to register false positives, the SACC/JIC will probably not be induced to commit air assets because several other sensors will still register negative. Our conclusion from the analysis of the MIUGS array is that there are no attributes whose failures will cause operational consequences.

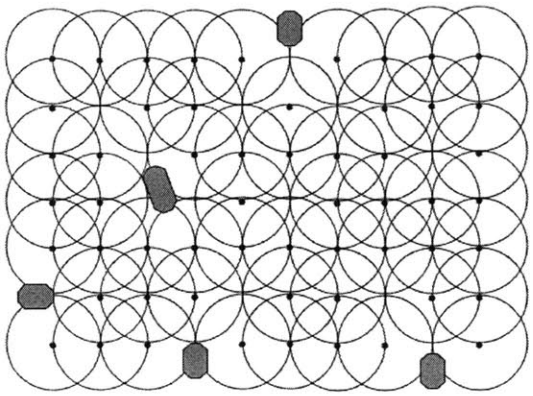

Figure 4-4: 20% Random Attrition in the MIUGS Array (Gray Areas Denote Coverage Blackouts).

4.1.3.2 The Global Hawk and the MTI

We have just established that the MIUGS network, which cues the dispatch of a Global Hawk UAV to the general area of enemy activity, does not have any operational consequences associated with system failure. The Global Hawk similarly has no operationally consequential attributes. For the purposes of this case, we will assume that friendly forces have multiple Global

[81] Even if the spacing between sensors is much larger than that in Figure 4-3, the failure of one sensor will have little impact on the effectiveness of the array, because other sensors along the SCUD launcher's path will still probably pick up the movement.

Hawks that they could launch to locate the target in the event of one Hawk's failure, though they only deploy one to track the SCUD launcher. The Global Hawk and its Moving Target Indicator radar (MTI) each have multiple attributes, but a failure on any one of these attributes is highly unlikely to cause operational consequences, and hence the attributes do not need extensive targeted testing:

- Global Hawk
    - *Airspeed.* While variations in airspeed will mean that the aircraft may close less quickly than desired on the target, the air speed differential will probably still be at a minimum 240 miles per hour, assuming that the target is moving away from the Global Hawk at 60 miles per hour and that this particular aircraft, because of manufacturing defects, has a cruising speed 100 mph below that of its peers. That is, even if the Global Hawk is well below specification, it will still be able to complete its mission.
    - *Mean Miles Between Failure.* The Global Hawk may crash on its way to the target or while tracking the target, causing there to be a blackout in MTI coverage. If the Global Hawk crashes on the way to the target, another backup aircraft can be put in the air or brought from another surveillance area to locate the convoy. Though this is suboptimal, all is not lost, as the MIUGS network still knows where the convoy is within a one mile radius—with the fast closing speed of the aircraft, the convoy will still be acquired relatively quickly. The Global Hawk crashing while tracking the target is not a problem, as a Predator will already be tracking the target vehicle in conjunction with the Global Hawk.
    - *Survivability.* We do not need to worry about how survivable the Global Hawk is for two primary reasons: first, if the Global Hawk gets shot down, another one can come to take its place (while hopefully evading the first one's fate). Second, the foes that friendly forces are facing in an unconventional war probably do not have the resources to obtain high altitude capable surface to air missiles (SAMs), the only things that could possibly threaten the Global Hawk at its altitude of 50,000-plus feet.
- Moving Target Indicator

- *Circular Error Probability* (CEP) of the radar, or the accuracy at which the convoy can be tracked.[82] It might be a problem if the target was far from the sensor—however, the MIUGS places the Global Hawk within a mile of the target, meaning that the CEP is inherently small, even if there is great uncertainty in how accurate the MTI antenna is. Let us say, for example, that the specification for the antenna error angle is one degree (typical for high end systems), but this particular MTI is five times spec, and has an error of five degrees. At the maximum range of one mile out, this changes the CEP from 92 feet to 460 feet, using a basic geometric calculation.[83] When the Predator comes overhead at an altitude in the tens of thousands of feet with an optical surveillance system, a ground difference of a few hundred feet in the location of the target simply will not matter (especially if the Global Hawk has closed on the target after getting an initial reading).
- *Radar Range.* The range for the radar is measured in tens of miles. The radar that is hunting for the SCUD launcher can be severely underpowered, as the MIUGS network will bring it within a mile of the target.
- False returns. The MTI could pick up the wrong target. It is doubtful that this will lead to mission failure, as the Predator will identify the target and the ground crew will be able to make a determination about whether it is the SCUD launcher or not.
- *MTI malfunction or shut down.* See Mean Miles Between Failure, as the consequences of the plane crashing and the MTI failing are the same (once the MTI fails the Global Hawk loses mission effectiveness).

4.1.3.3 The Predator

The Predator, unlike the Global Hawk and the MIUGS, does have failure modes that can trigger operational consequences. While the airframe itself has no attributes that could result in

[82] CEP is not the exact term for this, as the "CEP" will actually more closely resemble an ellipse due to the operational characteristics of the radar system.
[83] Lecture on Intelligence, Surveillance, and Reconnaissance Systems by Dr. Owen Cote in Understanding Military Operations, MIT Security Studies Program.

operational consequences (see section above, as the Predator's airframe has the same general attributes as the Global Hawk's), the sensor package does. Where the Electro-optical Video Camera sensor package is similar to the MTI in that its range (or the width of the lens angle) and potential failure (that is, it turning off unexpectedly) are not important in the context of operational consequences, the performance of the system when it gets to the target with its optical sensors is immensely important.

There is one primary performance parameter that dictates whether the Predator will be able to perform its mission on arrival and get a positive identification on the target: the resolution of the camera at high altitude. If the Predator cannot perform to this parameter, no positive identification will be possible, and it will be impossible to strike the SCUD launcher (an operational consequence)—the optical sensor needs to be able to distinguish between the SCUD launcher and other vehicles. If the resolution is not high enough, the Predator will not be able to find the right vehicle to strike, potentially allowing the SCUD launcher to get away, and will also not necessarily be able to identify false positives being tracked by the MIUGS and Global Hawk network. The Predator needs to be able to perform this identification procedure at a high enough altitude so that it can evade enemy small arms fire from the ground (and hopefully avoid detection altogether)—if the Predator must fly 1,000 feet over the enemy to get a positive ID, it may never get a chance to complete the mission. While it is okay for the MIUGS and Global Hawk to have leeway in most of their specifications and performance, the Predator must be very capable in its optical sensing capability.

4.1.3.4 The Network

A good way to evaluate the network that connects the systems is in terms of *modularity* and *coupling*.[84] Modularity is the ability of a system to swap components in and out without having to redesign the system architecture or other components within the system. In our context, modularity means that the failure of one component's attribute does not affect other parts of the system. Within the SCUD tracking sensor network, the modular parts of the system are the sensors out in the field. Assuming that the SACC/JIC and sensors have the proper data link and that there is enough bandwidth, sensors can be added to the network and some sensors can fail

[84] Shah, Nirav Bharat. "Modularity as an Enabler for Evolutionary Acquisition." 2004.

without affecting the operability of other parts of the system. For example, if one MIUGS stops operating, the other MIUGS will continue operating, as will the Predator, Global Hawk, and SACC/JIC. Coupling, on the other hand, means that if one part of the system fails, it will make other parts incapable of performing their mission. Coupled parts of the system are extremely important, as their failure can propagate and cause mission failure for the whole system. In systems testing, the most coupled part of a system is often seen as the "brittle" portion of the system and is designated to receive the most testing.

The element of the sensor network that we have not looked at yet is the communications infrastructure behind the sensors. The SACC/JIC takes in information from sensors via data links, and communicates instructions back to the sensors, thus orchestrating the whole SCUD finding mission. If the SACC/JIC were no longer able to communicate with the sensors, the MIUGS would not be able to cue Global Hawk missions; the Global Hawk would not be able to transmit MTI data, and thus could not call in the Predator; the Predator itself would not be able to transmit a video feed back to the SACC/JIC, making a positive identification and subsequent air strike impossible. The network becomes especially important in cases like this, where the system of systems is predominantly made up of unmanned vehicles with no means of communication besides the data links and no capability to make decisions on their own.

For operational consequences, the implications are obvious—if the SACC/JIC goes down, the sensor network will be unable to locate the SCUD launcher, causing mission failure. Our idealized version of the SACC/JIC has several attributes, most of which have operational consequences (here we are treating the "SACC/JIC" as the both a communications node and as a data link all rolled into one, for simplicity's sake):

- *Bandwidth.* A system failure in bandwidth would not have operational consequences. While dramatically lower than expected bandwidth would increase the time it takes for information to get passed back to the SACC/JIC, the slow speed of the convoy and the time it takes to set up a SCUD launcher means that a seconds long time delay in decision-making will not have a catastrophic effect on the probability of mission success.
- *Error Rate in Communications.* Like bandwidth, the error rate in communications is unlikely to cause operational consequences. Even if a communication were wrongly transmitted (the MIUGS, for example, sends out a detection notice but the notice never

reaches the SACC/JIC), a simple implementation of proper procedures where signals are sent multiple times would mitigate the problem. The time it takes to catch the problem would slow the system response time, but overall this would have little effect on the sensor network's operations.

- *Range of Communications*. If the communications range is less than expected, and no testing was done previously to identify the problem, then serious operational consequences could result. The SACC/JIC would not be able to support a geographically wide enough range of sensors, meaning that not all parts of the area of operations would be covered, potentially leaving room for the SCUD to maneuver outside of sensor range unbeknownst to friendly forces. In this case, the total range of the system is the communications range of the SACC/JIC plus ½ miles, assuming that the farthest out sensors are placed at the edge of communications range.
- *Mean Hours Between Communications Node Failure*. Even though there are undoubtedly redundant systems in the SACC/JIC (extra generators and additional communications antennas, for example), if the whole node goes down for any reason and stays down for a long time, there would be significant risk of mission failure. Once the SACC/JIC goes down the MIUGS-Global Hawk-Predator network becomes incapable of transmitting information, and thus becomes mission incapable. It is essential that the SACC/JIC be given endurance tests (especially in adverse weather conditions) to ensure that it does not fail. Failure might also be induced by enemy direct and indirect fires, which might mean the destruction of the SACC/JIC if its constituent equipment is not hardened sufficiently.
- *Vulnerability to Electronic Countermeasures (ECM)*. If the opponent is somewhat sophisticated, he may attempt to block the signals using jamming to prevent the transmission of data between the SACC/JIC and the component sensors. A successful jamming operation would mean that the sensor network would lose its ability to track and defeat the SCUD threat, leading to operational consequences. Testing the system's ability to resist jamming is thus essential.
- *Ability to Operate in Austere Terrain*. If significant ground obstacles like mountains are able to disrupt communications, and the system is deployed in such a region, then the SACC/JIC may prove unable to reach significant portions of the sensor network.

From the above analysis, we have derived a list of four system attributes that can have operational consequences. In the next section, we aggregate the results of our analysis of the whole system of systems and present a very basic concept for a testing plan.

4.1.4 The Sensor Network Testing Plan

Though the analysis above, we have established a list of five system attributes that could have operational consequences and thus require more extensive testing than other system attributes. This section very briefly describes how a prescriptive system of targeted testing might be created to ensure that the specific operational consequence of failing to properly track the SCUD does not occur. These attributes are: the Predator's ability to capture high enough resolution images at high enough altitudes to accurately identify the SCUD launcher (1); the SACC/JIC's communications range (2); the SACC/JIC's mean hours to failure (3); the SACC/JICs vulnerability to ECM (4); and the SACC/JIC's ability to operate in austere terrain (5). The following concepts for test plans are meant to illustrate the type of targeted testing that must be conducted to ensure that operational consequences do not happen.

1. Fly a Predator over a mock target site at different altitudes, alternating the targets randomly (sometimes the convoy might include a SCUD launcher, sometimes not, with randomly created combinations of convoy vehicles and random vehicle placements within the convoy), and see if the operators can properly identify the targets. Operator training plays a large role in this—even the highest resolution camera with untrained operators may have a low probability of proper identification.
2. Test the SACC/JIC's communications range by deploying all three types of sensors (MIUGS, Global Hawk, Predator) at sequentially longer ranges and sending test messages both from the SACC/JIC and to the SACC/JIC. Repeat this process for multiple SACC/JIC's, to ensure that the confidence interval associated with the range attribute is tight.
3. Operate a small sensor network (with all three sensor types) continuously with doctrinal maintenance procedures until the SACC/JIC malfunctions in a way that would disrupt operations for a lengthy period of time, or until the SACC/JIC's operation time is longer

than the projected maximum duration of the anti-SCUD mission. Conduct this experiment concurrently with several SACC/JICs.

4. Operate a sensor network at the SACC/JIC's maximum communication range, and apply jamming that corresponds with the highest suspected enemy capability. Push the bounds high enough until the communications are successfully jammed, to establish a known threshold at which the system can no longer operate. Repeat this experiment multiple times.
5. Test the network with the sensor components and the SACC/JIC separated by a major terrain feature like a mountain range, running a convoy with a SCUD launcher by the network. Repeat this for multiple terrain features. The end result of this test will be a set of terrain features in which the system will and will not operate properly, allowing the end user to implant the system in terrain in which it is very likely to work.

It is worth noting that the above testing procedures (some of which could be rolled into normal operational testing, like numbers 3 and 4) would undoubtedly be very expensive and time consuming. This, however, is the point of testing for operational consequences: high premiums are worth paying to ensure that operational consequences, which have very high costs (in this case, vast damage from a WMD-armed SCUD missile), do not occur. The testing, as discussed before, reduces uncertainty in system attributes so that the user can be very sure that the system will operate correctly when called upon. Operational consequences are too costly to leave to chance.

4.1.5 Conclusions from the Sensor Network Case Study

In this case study, we went through several steps to find and develop test plans for system attributes in the sensor network that might have operational consequences. We first established that the sensor network does indeed have operational consequences that need to be mitigated with a specialized test plan (4.1.2). We then used a framework to break the sensor network into its constituent parts (4.1.3), and evaluated each component based on its attributes to see if the failure in any attribute would have operational consequences in the context of the specified mission (4.1.3.1 through 4.1.3.3). We also evaluated the network that governs the whole sensor

system, and in particular the SACC/JIC, to see which attributes of the communications nexus could have failure modes with operational consequences (4.1.3.4). After these analyses, we distilled a list of five system attributes that could have operational consequences, from which we could derive a rudimentary list of targeted tests that should be performed to ensure that those consequences do not happen (4.1.4).

The primary purpose of the case study is to explore operational consequences and how they can be mitigated. Where programmatic consequences require a broad based testing strategy, operational consequences require that systems be evaluated for potential operational consequences and that they be tested very extensively on a handful of attributes. The idea is to develop a series of targeted tests in addition to regular testing to gain added insurance that operational consequences do not occur.

4.2 Case Study 2: The Predator Drone

The previous case study, as discussed in 4.1.2, looked at a particular mission of a system of systems and developed a test plan to prevent operational consequences for that mission. This brief case study takes a step back and looks at one system, the same Predator drone discussed in the previous case study, to develop a test plan for mitigating operational consequences that could occur when the Predator operates as a lone system. The part of the test plan that dealt with the Predator drone in the first case study is part of a broader range of tests that should be conducted on the Predator to ensure that the drone does not have operational consequences on the wide range of missions it will encounter in its service life.

This case study takes the Predator drone, and looks at a wide range of missions that it may be asked to perform, and looks at each system attribute in the context of those missions to see which attributes might have operational consequences. While the analysis presented cannot be considered comprehensive, as I am no expert on the Predator weapons system and the analysis is largely drawn from my own conjecture, the intent is to provide a methodology by which such analyses can be conducted in the future. The analysis proceeds in four parts:

1. Define the attributes of the system.
2. Define the mission set of the system.

3. Evaluate the attributes in the context of each mission to define which attribute failures may have operational consequences.
4. Develop a test plan for each high stakes attribute.

After this exercise is completed, we will have a good idea of specific tests that must be run on the Predator to ensure that high stakes consequences do not occur. Again, the following is not meant as a comprehensive analysis but instead meant as an example for how such studies can be conducted in the future.

4.2.1 Defining the Predator's System Attributes

The Predator is a 1,130-pound unmanned aerial vehicle (UAV) that can operate for over 24 hours without refueling at altitudes up to 25,000 feet. It is 27 feet long with a wingspan of 49 feet, and cruises at around 80 knots.[85] We can loosely divide the system into four sets of \ components: the airframe, the sensor packages, the armament, and the command, control, communications, and computer (C4) systems that allow it to process signals and interact with ground controllers. In the following, we will expand and, where necessary, describe each of those groups of components in terms of systems attributes using information from GlobalSecurity.org's data on the Predator Aircraft.

- Airframe
    - *Airspeed*. The aircraft's maximum speed.
    - *Maneuverability*. Maximum g-forces that the aircraft can sustain while maneuvering.
    - *Survivability*. Ability to survive enemy fire.
    - *Altitude*. The aircraft's maximum altitude.
    - *Reliability*. The aircraft's mean miles between failure (will it crash while flying).
    - *Flight Duration*. The aircraft's fuel economy on a mission that includes movement to target and time over target.

[85] See GlobalSecurity.org's page on the Predator at Globalsecurity.org/intell/systems/predator.htm. Accessed May 13, 2010.

- *Payload.* The aircraft's maximum payload.

- Sensor Packages
  - *Electro-optic (EO) Camera.* A combination of two cameras: a spotter video camera and a color video camera.
    - *Resolution at Cruising Altitude.*
    - *Zoom.*
  - *Forward Looking Infrared (FLIR) Sensor.* Meant for identification of targets at night and in adverse weather—relies on temperature difference between target and surroundings.
    - *Resolution at Cruising Altitude.*
    - *Zoom.*
  - *Laser Range Finder.* Used to track the range to targets.
    - *Range.*
    - *Accuracy.*
    - *Adverse Weather Capability.*
  - *EO/FLIR and Range Finder Mount.*
    - *Mean Hours Between Failure.*
    - *Precision of Control.* The resolution of the traverse of the sensor pod.
    - *Speed of Traverse.*
    - *Survivability.* Ability to continue operating after enemy fire.
  - *Synthetic Aperture Radar (SAR).* Used to track ground targets in adverse weather—only takes still pictures.
    - *Resolution at Cruising Altitude.*
    - *Susceptibility to Jamming.*

- Armament
  - *Targeting System.*
    - *Accuracy.*
    - *Adverse Weather Capability.*
  - *Missile/bomb Package.*
    - *Explosive Power.*
    - *Accuracy.*

- *C4 Systems* (lumped for brevity's sake—the communications gear includes a satellite and a line of sight data link).
    - *Time Lag*. The time difference between the sending of information and the ground controller's receipt of that information.
    - *Bandwidth*. The rate at which data can be transferred from the sensors to the ground controller.
    - *Susceptibility to Jamming*.
    - *Susceptibility to Hacking or Hijack*. The susceptibility of the aircraft's communication network to unwanted access from enemy information operations.
    - *Reliability in Adverse Conditions*.

This set of attributes, when examined in the context of the Predator's mission set, will allow us to find which attributes need additional testing to ensure that operational consequences have a very low probability of occurring.

4.2.2 Defining the Predator's Mission Set

The Predator's mission set can loosely be divided into two groups: kinetic (that is, a mission that uses munitions) and non-kinetic. This mission set is by no means comprehensive, but provides us with something we can use to evaluate each system attribute in the context of operational consequences. In both kinetic and non-kinetic missions, there are several variations that can take place on each mission based on a host of variables, like terrain, weather, enemy force composition, and friendly forces composition (for example, a spotting team on the ground that can assist in targeting).

- Kinetic
    - *Close Air Support*. In this mission, the Predator is used in place of manned aircraft to support ground troops by eliminating enemy bases of fire with bombs or missiles.

- *High-value Target Destruction*. In this mission, the Predator is used to destroy a suspected militant commander or a vehicle carrying a dangerous payload (like a SCUD launcher as in Case Study 1).

- Non-Kinetic
  - *Surveillance*. In this mission, a Predator loiters above an area for several hours looking for enemy activity with its sensor packages. A good example of this mission came during the 1999 Kosovo bombing campaign, when Predator drones gleaned valuable images documenting Serbian war crimes.
  - *Tracking*. In this mission, the Predator drone is sent to follow a vehicle or person as they move, possibly to target a strike or a ground mission.
  - *Target Identification*. In this mission (like the scenario in the first case study), the Predator is sent in by another unit (a Global Hawk or a ground team, for example) to get an aerial view of a target.
  - *Target Designation*. Using a laser designator or other targeting device, the Predator targets an enemy emplacement for a bombing run or missile strike by a different system.

4.2.3 System Attributes, Mission Sets, and Operational Consequences

We have just established a set of missions that the Predator might perform in the field and also a set of system attributes that might be relevant to those missions. Going through each of the identified 29 attributes for each of the six missions (yielding 174 points of analysis) would be tedious and is not necessary to explain my methodology for operational consequence analysis (though is recommended for an actual analysis in the future). In the following, I will examine two missions and bring out a sampling of attributes that both have and lack potential for operational consequences to give a flavor for the analysis methodology.

4.2.3.1 High-value Target Destruction Mission

The high-value target destruction mission involves a Predator drone, after an intelligence report, finding, identifying, and attacking a group of important enemy personnel who are

probably on the move in a vehicle or on foot (like a high-ranking insurgent commander). For discussion's sake, we will choose a more difficult version of this operation from that in Case Study 1, where the Predator is operating in terrain with a high population density (a small town, for example, rules of engagement notwithstanding), under somewhat adverse weather conditions (a light rain is coming down, limiting visibility somewhat). There are two primary operational consequences that we must worry about in this sort of mission: collateral damage and the destruction of a high-value target. As discussed previously in this paper (Section 2.3.2), collateral damage could turn tens of villagers into insurgents, make the village less receptive to coalition efforts, and giving the insurgents a base of support that will make them difficult to defeat, resulting in many more friendly deaths and potentially a much longer troop presence in the area. The failure to destroy the high-value target also is an operational consequence, as it allows the insurgent organization to continue unabated and allows their leadership to continue to exercise command and control. Note that this mission is at its core the same as that in the first case study, except for the difference in assumed adverse weather conditions, the urban terrain, and the Predator destroying rather than just identifying the target.

From the set of airframe attributes, we find that there are none with operational consequences. As the Predator is flying a mission against a small group of insurgents armed with small arms, and there is probably another Predator available should the first somehow get shot down, survivability is unlikely to have operational consequences. Likewise, variables like airspeed and maneuverability also have little to do with the Predator's ability to conduct a ground attack. Similarly, reliability will not have operational consequences, because in the unlikely event that the aircraft goes down, another aircraft should be ready to take the first's place.

The EO sensor package attributes, unlike the airframe attributes, could have operational consequences. The EO Camera must be able to properly identify the target at cruising altitude, especially in dense urban terrain and lowered visibility conditions. If the EO Camera does not have adequate resolution to identify the target, and the wrong target is selected, there could be collateral damage that might turn the whole village against coalition forces. Misidentification of targets is a real problem when using air power—one need look no further than the killing of two journalists in 2007, whose cameras were misidentified for weapons by an Apache helicopter

crew, to see how misidentification can lead to operational consequences.[86] The other sensors are less important, as the FLIR, SAR, and Laser Range finder, though helpful in locating and tracking the target, would not be used to identify the target and thus will not have operational consequences.

The armaments are another key area where operational consequences can take hold. The attributes of the weapons system—the accuracy of the missile itself, the accuracy of the targeting system, the explosive power of the missile, and the targeting system's adverse weather capability—are all essential in both ensuring that the target is killed, and, more importantly, ensuring that there is little collateral damage. If the missile or targeting system are inaccurate, the missile may hit civilian homes or shops; if the missile is more explosive than originally planned, even a direct hit could cause civilian casualties; if the system is not as capable in adverse weather as planned, the missile may not be able to launch in the first place, leading to mission failure.

Finally, some of the C4 attributes might have operational consequences, while others will not. A large time lag in the video feed could cause the operator to open fire just as a civilian vehicle is moving into the way, causing collateral damage; a lack of bandwidth could mean that the EO Camera is not being used to its full potential, causing a reduction in resolution that could lead to target misidentification. Since it is likely that the target is traveling in a small group armed only with small arms, it is unlikely that the UAV will face any sort of electronic warfare capabilities, meaning that jamming and hijack will not be a problem.

From this analysis, we find that the EO Camera's resolution and zoom are high stakes attributes, as are all attributes relating to armament, the bandwidth of the C4 system, and the time lag in the C4 system.

4.2.3.2 Surveillance Mission

In this mission, a Predator is sent to a target area to loiter and wait for insurgent activity. If activity is detected, the ground controller can call in an air strike from other air assets or request that ground forces be moved to the area. Unlike the previous mission, where mission failure was an operational consequence because it meant a failure to destroy a high-value target, mission failure (or the inability of the aircraft to complete the mission) in this case is not an

[86] CNN Staff. "Video Shows Deaths of Two Reuters Journalists in Iraq in 2007." 2010.

operational consequence because it simply means that a small band of low-level insurgents might not be identified and killed or a roadside bomb might be planted (this is the presumed role of the Predator's surveillance). While those insurgents may go on to cause some damage to friendly forces, they are unlikely to have disproportionately large effects on friendly operations. Thus, as in the previous mission, the airframe attributes have no potential for operational consequences, because if the Predator goes down for any reason, the only result is mission failure (note that because the Predator is operating far from base, it is unlikely to have a replacement craft immediately sent out in the event of a crash).

In the sensor suite, again, the EO Camera is still the only component with potential for operational consequences, since the other sensors will probably not be used for target identification. If the EO Camera misidentifies a group of insurgents as hostile, collateral damage might result from the ensuing air strike. Since this is a non-kinetic mission (if an air strike is called in, it would be from an outside source), the armament packages will not be on the aircraft, and thus cannot have any operational consequences whatsoever. The C4 systems also will probably not have operational consequences, as the worst-case scenario involving C4 failure is mission failure, which is doubtful to have operational consequences. The lone exception to this is low bandwidth, which might result in greater compression of EO Camera data, and the ensuing lower resolution may result in target misidentification.

4.2.4 Testing Strategy

The final step in developing a plan to test for operational consequences is to develop the actual tests themselves. From the analyses above, we have found several attributes that could have operational consequences. In the bulleted list below, we explore very rudimentary test plans for each of those attributes by mission. The intent is to provide examples for the sorts of tests that might be prescribed for the Predator drone to mitigate operational consequences for the two missions explored.

- High-value Target Destruction Mission
    - *EO Camera Resolution.* Fly the predator with an EO package over a mock urban area such as the one at the Joint Readiness Training Center in Louisiana during a

training operation. During the flight, attempt to identify key targets using intelligence information collected by the ground troops actually taking part in the exercise. The test administrator can grade the Predator's performance in acheiving a successful ID. Key differentiating aspects for targets might be the color of clothing, presence of an armed entourage, or a specific type of vehicle.

- *Armament Accuracy, Destructive Power, and Adverse Weather Capability.* Testing for these attributes would be best done in a mock urban test range with residential and commercial areas under a range of weather conditions. Targets, mobile and immobile, could be set in various "city streets" and engaged by Predator aircraft, with post-test analyses being conducted for Circular Error Probability (accuracy of the bomb hit) and the bomb blast radius. This would give the tester a good idea for the collateral damage potential of the Predator and associated weapons packages, and could also help develop rules of engagement specific for Predator aircraft.
- *Bandwidth and Time Lag.* Run several missions using the same data links that would actually be used on deployment (i.e. use full satellite networks linked back to operators in U.S. based UAV operations centers) to ensure that time lag and bandwidth constraints will not have an impact on mission effectiveness. These tests can be run in conjunction with the first two sets for efficiency's sake.

- Surveillance Mission
  - *EO Camera Resolution.* This mission is inherently different from the EO Camera resolution in the High-value Target mission in that this mission requires persistent surveillance over a long time period, requiring better crew endurance and the ability to differentiate between civilian and combatant in an environment where there is no confirmed combatant presence. To conduct this test, Predators could be sent to several kinds of terrain (for example urban, mountainous, and desert) and told to loiter over specified areas. Potential enemy personnel could be sent to the target area at random intervals—some groups could be unarmed, some could try to plant an Improvised Explosive Device (IED), and some could be armed. With each group, identification of civilian and enemy could be conducted. It is important to note that identification is not purely a function of camera resolution,

but also relies heavily on crew training, a factor that will have to be accounted for in testing (training is an important part of operational testing, as we saw in the DIVAD case in Section 3).

4.2.5 Conclusions From the Predator Case Study

In this case study, we established a method for establishing plans to test systems for the purpose of preventing operational consequences. First, we provided a list of system attributes, breaking the system into components where necessary. We then established a list of potential missions that the system might realistically be asked to perform during its service life. From the combination of the system attributes and the missions list, we developed a set of attributes that could have operational consequences on each mission, and hence need further testing. Finally, we created concepts for extensive testing to be done on those attributes to ensure that they had low probabilities of failure. The key takeaway from this case study is that, through analysis, prescriptive testing plans can be formulated by which operational consequences can be mitigated.

4.3 Conclusions

In this section, we developed methods for analysis that can be used to prescribe testing plans in order to mitigate operational consequences, and better defined the concept of operational consequences through concrete examples. The first case study illustrated how testing could be done for systems within a defined mission type to ensure mission success. Since this case involved an amalgamation of unmanned sensors configured for a specific role, it made sense to evaluate operational consequences in the context of that role rather than on a system-by-system basis. The final result of the analysis showed us that communications networks in system of systems are crucial, especially when the system of systems is largely unmanned. The second case study provided a method for evaluating an individual system over a broad range of potential missions, with special focus given to the consequences unique to unconventional warfare. In both cases, proper target identification was critical to mission success and prevention of collateral damage.

5

Conclusions

This thesis has tried to expand the theory on the testing of military systems by exploring the high stakes environment in which they operate, and translating the relevance of that environment into implications for testers and program managers. In short, *because of the high stakes consequences associated with the development and use of military systems, testers must adjust their testing strategies to ensure that high stakes consequences do not occur.* One of my primary objectives in the writing of this thesis is to make sure that military system testing theory has a clear connection to realities both in the field and in systems development, an objective that I work towards through the use of realistic cases studies and analysis.

In the first section, we examined the basic statistical theory that drives testing, focusing on how testing helps reduce uncertainty in the risk associated with a system by reducing uncertainty on the probability of system failure. With the very high consequence of failure that military systems often face, it becomes extraordinarily important to reduce uncertainty through testing to the greatest extent possible, because the impact of that uncertainty on risk is very high (recall that risk is the probability of failure times the consequence of failure). We broke down the high consequence space into two types of consequences, programmatic and operational.

Programmatic consequences occur while a system is still under development, and result when insufficient testing is conducted on a system, leading a program manager to have inadequate certainty that the system actually works to specification. When the program comes under the inevitable public and Congressional scrutiny, a lack of testing data can prove a death knell, as without testing data the system will be impossible to defend. To stop programmatic consequences, testers must utilize a broad based and adaptive test plan that ensures adequate testing across all system attributes, as a failure in any attribute might lead to schedule delays, budget cuts, or, at worst, program termination. Especially when a system fails to meet a

specification on a specific attribute, the tester must develop a plan to retest that attribute to rigorously verify that the implemented fix did indeed work.

To connect programmatic consequences to the realities of system development, we examined in comparative perspective the development of the Division Air Defense System (DIVAD) and the M-1 Abrams main battle tank, using testing as an explanation for their dramatically different programmatic outcomes (Section 3). The DIVAD's testing strategy was by no means rigorous, as it involved relatively little testing from which it was impossible to gain certainty about attributes, nor was it adaptive, as it did not adequately test fixes for the system as they were put in place. Because of this, the proponents of the DIVAD were unable to defend the system when it became the target of deep criticism, and the system was cancelled, leaving the Army with no mechanized short-range air defense system to this day. The M-1's testing strategy, on the other hand, was both rigorous and adaptive, and extracted vast amounts of testing data from the handful of prototypes and pilot vehicles that were produced. When certain system elements proved inadequate, the program's engineers quickly developed fixes and program management called for in depth tests to verify that the fixes indeed worked. The end result was testing data easily able to withstand the criticisms thrown at the system by the Congress, the public, and the Government Accountability Office.

Operational consequences, in contrast to programmatic consequences, result from failures of specific attributes in specific failure modes during military operations, after the system has already been fielded. What distinguishes operational consequences from normal systems failure consequences is their disproportionate impacts at operational and strategic levels—operational consequences might result in major shifts in troops, many deaths, or defeat in a particular region. Where programmatic consequences require a broad based and adaptive testing strategy, because it is uncertain which system attributes require additional testing until after they fail, operational consequences require a targeted testing strategy based on analysis of critical attributes. The strategy for this analysis is established in Section 4 through the use of two case studies. The first case examines a sensor network designed to stop SCUD launches in austere areas, while the second conducts an analysis of the potential operational consequences of failures in the Predator drone's system attributes. Where the first case study delves deeply into a single mission for a specialized system of systems, the second looks at many potential missions for a single multipurpose system. Through these two analyses of preexisting cases and systems, we can get a

good grasp of both operational consequences and the testing strategies by which they can be mitigated.

There are certain areas that this thesis left uncovered and need further research. The most important one of these is the need for means by which to measure operational and programmatic consequences in a way that allows testers to make decisions on how to allocate testing resources. While it is helpful to explore and define high stakes consequences, it is important that testers get better metrics than the "rigorous" and "very high" levels of testing discussed in this paper. Creating these measurements will be very tricky, as it is difficult to link the value of operational and programmatic consequences to a specific metric, and tougher still to link that metric back to actual testing. While the case studies here have been very realistic, they do little to answer the "how much testing is enough?" question. A potential way to model high stakes consequences might be to use an exponential rather than linear scale for the consequences of failure, and use that scale to determine a system's testing needs. Through this method, programmatic and operational consequences could be incorporated into test plans. Those attributes with potential operational consequences would be slated to receive extensive testing (based on the exponential measurement of consequences), while other non-operationally consequential attributes would be subjected to the broad based testing plan important to programmatic consequences. When an attribute experiences a failure, and thus could result in programmatic consequences, it too should be slated to receive extensive testing after a fix for verification's sake.

The concepts contained within this thesis could also be expanded both within testing and in other parts of the system development cycle. Within testing, this thesis has exclusively focused on systems, with little attention paid, for example, to tactics, techniques, and procedures (TTP) development. Perhaps more than systems, TTPs can have high stakes consequences as they guide all that soldiers do out in the field—new TTPs often ought to be tested rigorously before they are rolled out to the whole military, if they could have operational consequences. Moreover, the focus on developmental and operational testing alone has missed other critical areas of systems development that can have operational and programmatic consequences. Within the design phase, for example, analyses for attributes whose failures might have operational or programmatic consequences ought to be conducted so that the system is designed with those attributes in mind. Additionally, the focus on developmental and operational testing has led us to ignore bench level testing (which tests individual system components) and integration testing

(which tests system components as they are put together), which, with guidance, could catch many of the problems that might cause operational or programmatic consequences.

A reduction in testing is often seen as an easy shortcut by which to cut programmatic costs and to shorten development cycles. After looking at the high stakes consequences of failure, it is evident that such a strategy risks not only programmatic consequences but, perhaps worse, severe operational consequences out in the field. Testers must ensure that their test strategies adapt to a system's strengths and weaknesses, and must take great pains to make their test plans match the high stakes consequences of system failure during actual operations. By exploring the concept of high stakes consequences, this thesis seeks to bring the complex reality of military systems into the testing literature, and thus help enlighten our understanding of the unique testing measures that military systems require.

References

Babcock, C. "M-1 Tanks Quickly Demonstrate Superiority; Iraqis' Soviet-Built T-72s Prove to be Outmatched So Far in Desert Battlefield Engagements." *Washington Post* 27 Feb. 1991, Final ed.: A28.

Bar-Shalom, Y, Li, X., and Kirubarajan, T.. *Estimation with Applications to Tracking and Navigation.* New York: John Wiley & Sons, 2001.

Clemen, R. *Making Hard Decisions: An Introduction to Decision Analysis.* Belmont, CA: Duxbury Press, 1996.

CNN Staff. "Video Shows Deaths of Two Reuters Journalists in Iraq in 2007." April 6, 2010. Available at <http://www.cnn.com/2010/WORLD/meast/04/05/iraq.photographers.killed/index.html>. Accessed May 13, 2010.

Congressional Budget Office, *Alternatives for the U.S. Tank Industrial Base.* Washington, D.C.: 1993.

Cowart, K. Personal Communications, 2010.

Dickerson, C.E. and Marvis, D.N. *Architecture and Principles of Systems Engineering.* Boca Raton, FL: CRC Press.

Dickerson, C.E. and Valerdi, R. "Using Relational Model Transformations to Reduce Complexity in SoS Requirements Traceability: Preliminary Investigation." IEEE System of Systems Engineering Conference 2010.

Ditton, M. H. "The DIVAD Procurement: A Weapon System Case Study." The Army Lawyer, August 1988. Charlottesville, VA: Judge Advocate General School, Department of the Army.

Howard, R. A. "Bayesian Decision Models for System Engineering." *IEEE Transactions on Systems Science and Cybernetics*, Vol. SSC-1, No. 1. November, 1965.

Keller, B. "Weinberger Faces Decision on Fate of $4.5 Billion Antiaircraft Gun." *New York Times.* 23 August 1985, Final Ed.: A11.

Kenley, B. "Metrics for Testing: Toward a More Strategic Approach to Decision Making for Testing." Working Papers. Lean Advancement Initiative, Massachusetts Institute of Technology 2010.

Krishnan, A. *War as Business: Technological Change and Military Service Contracting.* Aldershot, Hampshire, England: Ashgate, 2008.

Mayer, J. "The Predator War." The New Yorker. October 26, 2009. < http://www.newyorker.com/reporting/2009/10/26/091026fa_fact_mayer>. Accessed May 13, 2010.

Moyer, R. "The Army, Contractors, and Weapons Development: Evidence From Congressional Hearings." Working Papers. Department of Political Science, Massachusetts Institute of Technology, 2009.

National Research Council. *Improved Operational Testing and Evaluation.* Panel on Operational Test Design and Evaluation of the Interim Armored Vehicle, Committee on National Statistics. Washington, DC: The National Academies Press, 2003.

Rooney. "Toyota Announces Gas Pedal Fix." February 1, 2010. Available at <http://money.cnn.com/2010/02/01/autos/toyota_gas_pedal_fix/index.htm>. Accessed May 17, 2010.

Shah, N. B.. "Modularity as an Enabler for Evolutionary Acquisition." Master's Thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology. June 2004.

Sapolsky, H. M., Gholz, E., Talmadge, C. *US Defense Politics: The Origins of Security Policy.* New York: Routledge.

United States Congress. House of Representatives. Investigations Subcommittee of the Committee on Armed Services. *Oversight Hearings on the Status of the Army XM-1 Tank Program.* 95th Cong., 1st Session. S. Doc. 77-H201-25. Washington, D.C.: Government Printing Office, March 25, 28 1977.

United States Congress. House of Representatives. Committee on Armed Services. *Hearings on Military Posture and H.R. 10929 Department of Defense Authorization for Appropriations for Fiscal Year 1979, Part 2: Procurement of Aircraft, Missiles, Tracked Combat Vehicles, Torpedoes, and other Weapons.* 95th Congress, 2nd Session. Washington, D.C.: Government Printing Office, February-April 1978.

United States Congress. House of Representatives Committee on Armed Services. *Hearings on Military Posture and H.R. 1872 and H.R. 2575 and H.R. 3406.* 96th Congress, 1st Session. Washington, D.C.: Government Printing Office, 1979.

United States Congress. House of Representatives, Subcommittee on the Department of Defense, Committee on Appropriations. *Department of Defense Appropriations for Fiscal Year 1980.* 96th Congress, 1st Session. Washington, D.C.: Government Printing Office, May 1, 1979.

United States Congress. House of Representatives. Special Subcommittee on NATO Standardization, Interoperability and Readiness, Committee on Armed Services. *Status of*

Army Air Defense Planning. 96th Congress, 2nd session. Washington, D.C.: Government Printing Office, September 30, 1980.

United States Congress. House of Representatives, Investigations Subcommittee of the Committee on Armed Services. *Status of the Army Tank Program*. 97th Congress, 2nd Session. S. Doc 83-H201-6. Washington, D.C.: Government Printing Office, November 5, 1981.

United States Congress. House of Representatives Committee on Armed Services. *Defense Deapartment Authorization for Appropriations for Fiscal Year 1983*. 97th Congress, 2nd Session. Washington, D.C.: Government Printing Office, 1982.

United States Congress. House of Representatives. Subcommittee on the Department of Defense, Appropriations Committee. *Department of Defense Appropriations for 1984*. 98th Congress, 1st Session. Washington, D.C.: Government Printing Office, 1983.

United States Congress. House of Representatives Committee on Armed Services. *Defense Deapartment Authorization for Appropriations for Fiscal Year 1986*. 99th Congress, 1st Session. S. Doc. 85-H201-23. Washington, D.C.: Government Printing Office, 1985.

United States Congress. Senate. Committee on Armed Services. *U.S. Army XM-1 Tank Program Hearings*. 94th Congress, 2nd Session. S. Doc. 77-S201-10. Washington, D.C.: Government Printing Office, September 14, 1976.

United States Congress. Senate Subcommittee on the Department of Defense, Appropriations Committee. *Department of Defense Appropriations for Fiscal Year 1979*. 95th Congress, 2nd Session. Washington, D.C.: Government Printing Office, 1978.

United States Congress. Senate. Committee on Armed Services. *Department of Defense Authorization for Fiscal Year 1981, Part 5: Research and Development*. 96th Congress, 2nd Session. Washington, D.C.: Government Printing Office, March 5- 26 1980.

United States Congress. Senate. Committee on Armed Services. *Department of Defense Authorization for Appropriations for Fiscal Year 1983, Part 4: Tactical Warfare*. 97th Congress, 2nd session. S. Doc. 82-S201-29. Washington, D.C.: Government Printing Office, February-March 1982.

United States Congress. Senate Committee on Governmental Affairs. *Management of the Department of Defense Hearing: Oversight of the Sgt. York (DIVAD) Air Defense Gun and the DSARC Decisionmaking Process*. 98th Congress, 2nd Session. Washington, D.C.: Government Printing Office, September 28, 1984.

United States Congress. Senate. Committee on Armed Services. *Oversight on the Division Air Defense Gun System (DIVAD)*. 98th Congress, 2nd Session. Washington, D.C.: Government Printing Office, October 2 1984.

Van Voorst, B., and Wilentz, A. "No More Time for Sergeant York." *Time*, 9 September 1985.