SYM-AM-19-062

PROCEEDINGS
OF THE
SIXTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

THURSDAY SESSIONS
VOLUME II

Acquisition Research:
Creating Synergy for Informed Change

May 8–9, 2019

Published: April 30, 2019

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

Making Time From Data: Toward Realistic Acquisition Schedule Estimates

Charles Pickar—is a member of the NPS faculty where he teaches project management, defense acquisition, and systems engineering. Before joining NPS, he led the Applied Systems Engineering Program Area at the Johns Hopkins University Applied Physics Laboratory. He is a retired Army officer with extensive experience in the U.S. defense industry, including director and VP levels at Lockheed Martin, Northrop Grumman, and SAIC. He is the current Chair of the Systems Education Technical Committee of the IEEE Systems Council. His research and published work focus on applying systems engineering and system dynamics analytical approaches to defense acquisition problems. [ckpickar@nps.edu]

Raymond Franck, Brig Gen, USAF (Ret.)—retired from the faculty of the Graduate School of Business & Public Policy (GSBPP), Naval Postgraduate School (NPS) in 2012. He retired from the Air Force in 2000 in the grade of Brigadier General. His active duty career included a number of operational tours staff positions, and head of the Department of Economics and Geography, USAF Academy. His published work includes a number of journal articles and research reports on military innovation and defense acquisition management. [cfranck215@aol.com]

Abstract

This paper continues a research agenda started in 2016 with an aim of more realistic acquisition program scheduling estimates, especially for the development (SSD) phase. We discuss acquisition management as a system, and its execution (especially with respect to schedule) from the perspective of Systems Dynamics (SD). We then present two episodes from F-35 program history. We then essay an integration of the SD method with these episodes using Cooper's (1998) failure modes. Finally, we present a discussion of system performance as a potential metric for schedule estimation and analysis (through schedule estimating relationships.)

Introduction

This paper is the *fourth* in this series of investigations into identifying both alternatives to the way we do schedule estimation (process), and the schedule dynamics that impact weapons system development execution (effects). It builds on the research agenda proposed by Franck et al. in 2016 and furthered in Franck et al. in 2017 and 2018 (Franck, Hildebrandt, & Udis, 2016, 2017, 2018). The goal of this ongoing project is to examine weapons systems development schedules to both identify current state and contributing causes of schedule estimating difficulties and suggest ways to more accurately predict development duration.

Before proceeding, it is worthwhile recap the genesis of the three previous research efforts. The original intent, unchanged, is to pursue a research agenda aimed at producing more accurate schedule estimates with a focus on major defense acquisition programs. The original research questions included the following:

- What is the current state of schedule estimation and control? What's needed?
- Where are the gaps?
- How can operational performance metrics better capture contemporary operations?
- What model(s) best capture the trade-offs among program cost and schedule, as well as operational capability of fielded equipment? Can those models give

insight into “troubled programs,” with difficulties in cost, schedule, and performance?

- Analyze previous case studies (e.g., from Kennedy School of Government) for insights into program schedule drivers.
- What estimating relationships best capture time to field new hardware? What schedule drivers are generally most important?
- Based on available data, formulate and empirically test models with hypothesized schedule drivers.
- Formulate and test prediction markets for cost and schedule problems.

While many of these questions have been considered, we have not yet been able to fully answer them. This paper continues the quest to better understand the schedule estimation process and why, after so much research and practice, we still have not come to terms with accurately estimating and executing development schedules. These are some interim findings from the past three years:

1. Data science, analysis, and empirical models show the type of analysis that can be accomplished using Selected Acquisition Reports (SAR) data.
2. The mining and analysis of acquisition data helps to identify reasons for schedule delays. The reasons (Schedule Delay Factors SDF) inform planners and schedulers on additional activities and sources of delays that must be considered in schedule planning and execution.
3. Systems Dynamics and other network models that include program schedules as an integral part of the modeled acquisition process have value in explaining the nature of schedule delays.
4. Exploration of more sophisticated mathematical models that interpret the causal structure associated with program schedule achievement show promise but need more work.

Why should we care about schedule delays? The primary reason is the impact on the warfighter. Systems scheduled to reach or provide initial operational capability that are delayed by years or even decades impact the DoD’s ability to fulfill its ultimate mission of protecting the country. Contractors care about delays because delays contribute to cash flow problems, and ultimately future contracts. Taxpayers care because delays not only can ultimately increase the cost of the development but may also result in canceled program and money wasted (Stumpf, 2000).

Exploring the Concept of Schedule

Review of the literature and discussion with defense acquisition scholars and practitioners interested in schedules reveals a fundamental distinction in the concept of weapons system development schedules. The first group focuses on the time it takes to develop a weapon system (Drezner & Smith, 1990; Pugh, 1987; Rothman, 1987; Tyson et al., 1989; Van Atta et al., 2015). This is the most prevalent research focus driven by the concern in the length of time necessary to field systems. This emphasis identifies schedule as a problem of technology maturity, cost overruns, cost estimating, budget formulation, and the time it takes to deliver weapons to the field. One of the aims of this aspect of schedule research is identify ways to reduce the time necessary to field systems.

The second interest and the one pursued in this research agenda asks the question, why did it take so long? This approach, focused on the mechanics of the system development, explores the issues of realism in creating and executing weapons system development schedules. For schedule creation, we focus on the schedule development process, task duration estimation, and the fundamentals of the Critical Path and Program Evaluation Review Technique (CPM/PERT). For schedule execution, we examine the reasons the established schedule is overrun. Instead, we concentrate on the challenges of bureaucracy, high-tech, technological complexity, and maturity and ultimately accept that serendipity has a role to play in the development of advanced weapons systems. Thus, we accept the fact that acquisition programs take longer to complete. Instead, we are interested in examining the details and decisions of weapon system development, and how those details and decisions can affect the dynamics reflected in program execution length.[1]

In order to effectively examine the creation and execution of schedule, we use three main approaches. The first is a systems approach emphasizing the dynamics of both schedule creation and execution. This systems approach is based in part on the idea that planning, scheduling, and project execution must be examined as a system—that the project or program does not consist of separate and unrelated variables (Senge, 2006). The second approach uses the case study approach. Because of its interest and size, our current efforts examine the F-35. Our case study approach uses a mixed-methods analysis using data, interviews, and qualitative analysis of program reports. Finally, we have been examining schedule through a quantitative approach through earned value management.

Systems, Complexity, and Schedules

A critical point to be made when discussing weapons system development is that the act of development, that which we call a program is actually a system. A system consists of activities or parts that interact to produce something. A system uses inputs and operating through constraints and mechanisms to produce an output. An effective way to visualize a system is by using an IDEF model. IDEF (**I**ntegrated Computer-Aided [ICAM] **DEF**inition) was developed by the U.S. Air Force in 1973. IDEF was derived from a well-established modeling language, the Structured Analysis and Design Technique (SADT) (Marca & McGowan, 2005). IDEF is useful when exploring the activities of a system by identifying what functions are performed (inputs and outputs), what is needed to perform those functions (controls), and who or what is performing those functions (Mechanisms). Figure 1 shows an elementary model of a weapons system development project as a system.

Figure 1 is almost deceptive in its simplicity until one considers the volatile mix of the variables named. Inputs to the system include warfighter needs effectively translated into valid requirements. Controls or constraints include Congressional oversight and funding, as well as the constant challenge of shifting priorities. Acquisition and engineering personnel provide the mechanism for the process of development to actually occur. The output is the completed weapon system delivered to the warfighter. While easily diagrammed, no one would argue that this process is not a complex undertaking. And, while we are only examining a part of this system, no discussion on creating or executing weapons systems development schedules would be complete without considering the complexity involved.

---

[1] While the field of schedule development also includes operational research approaches to schedule development and estimation (e.g., Van de Vonder, Demeulemeester, & Herroelen, 2007; Vandevoorde & Vanhoucke, 2006), this aspect is not included in our study.

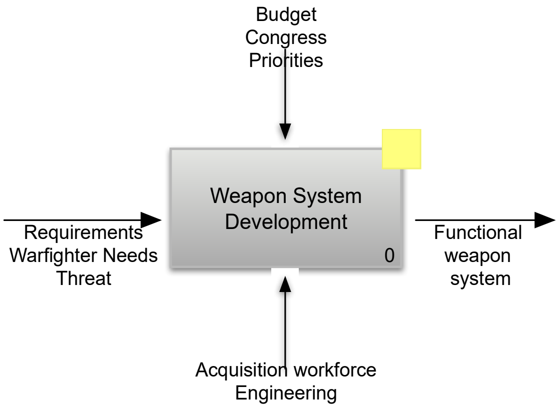

Figure 1. Weapon System Development as System

Because complexity science is, well, complex, we limit this discussion of complexity to three recognized types, structural complexity, detail complexity, and dynamic complexity (Dörner, 1990; Perrow, 1999; Senge, 2006; Williams, 2002). The first type, structural complexity is a construct developed by Williams that effectively captures the later classification of detail and dynamic complexity and includes the idea of uncertainty as a complexity contributor. Figure 2 shows a modified structural complexity construct (Williams, 2002). The revised graphic acknowledges the Williams' structural complexity and uncertainty, but suggests that decision dynamics is a more suitable result of uncertainty.

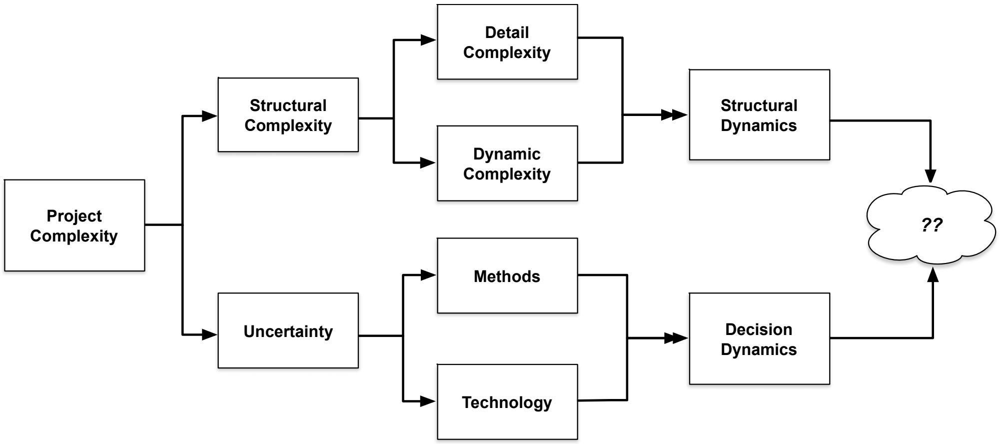

Figure 2. Complexity Model
(Adapted from Williams, 2002)

Detail complexity is about the size, scope, and/or the amounts of "things" in a system. It is concerned with the number and differentiation of the quantities of parts, dollars, pages in a contract, subsystems, or the size of a system, in other words, the number of variables (Baccarini, 1996). Detail complexity can often be overwhelming, but that is caused by the sheer number of elements one has to consider. Detail complexity is also the most familiar and thus addressable of these two forms of complexity because detail complexity can be captured in a spreadsheet.

Dynamic complexity is about interdependence and interrelationships and the feedback loops of various events of the development (Dörner, 1990). It is dynamic complexity that is central to the idea of schedule. We find dynamic complexity in "situations where cause and effect are subtle and where the effects over time of interventions are not obvious" (Senge, 2006, p. 70). Dynamic complexity is one of the greatest challenges that

PMs have to overcome. It is insidious in its effect because the results of dynamic complexity are not immediately apparent. Time is a critical factor in dynamic complexity:

> We rarely have trouble dealing with configurations in space. If we're not entirely sure of what we're looking at, we can take another look and resolve our uncertainty. We can normally look at forms in space again and again and in this way precisely determine their particular configuration. That is not true of configurations in time. A time configuration is available for examination only in retrospect. (Dörner, 1997, p. 100)

Managers in every industry make decisions and expect to see quick results of those decisions. In fact, this almost immediate feedback has become central to the U.S. stock market, for example. Market and industry analysts drive investors to expect to see the results of decisions often within the next quarter. However, dynamic systems and the associated complexity may or may not react in defined time frames. In reality, "*Conventional forecasting, planning and analysis methods* are not equipped to deal with dynamic complexity" [emphasis added] (Senge, 2006, p. 70).

A major manifestation of dynamic complexity is the time frame. The greatest threat to the success of a system development is not a quick, single catastrophic act, but instead the slow, almost imperceptible changes in the system that result from PM decisions (Senge, 2006). In fact, many PMs will not see the effects of their decisions before they move on to another position. This is the end state of decision dynamics.

And the problem continues because we learn best from experience. This is the benefit of experiential learning whether it is part of a curriculum or a result of on-the-job training. However, we rarely directly experience the consequences of many of our most important decisions (Senge, 2006, p. 30). This idea of project dynamics is one the DoD tends to ignore, but one we will continue to explore to better understand and explain how we can build and execute better weapons system development schedules.

Schedule Processes as System

Project planning is a well-defined and generally well understood process detailed in both DoD and the Project Management Institute (PMI) documents. Figure 3 shows a modified version of the generally accepted schedule development process from activity definition to execution. The work breakdown structure (WBS) identifies the tasks necessary for system development. WBS feeds these tasks into the scheduling process by providing activity definitions. The activity definition part of schedule estimation focuses on those activities defined by the WBS. If an activity is not named in the WBS, it is not included. This requires consideration of those tasks/activities that may not have a direct link to engineering tasks but are still essential to system development. The activities include other events, such as those imposed by the customer, in this case the DoD.

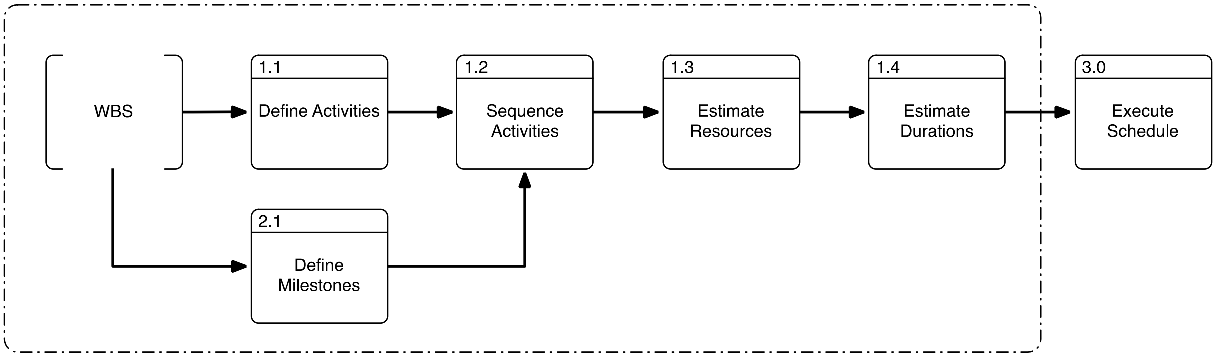

Figure 3. Schedule Planning Process

Activity sequencing is the process of sequencing the tasks identified through the work breakdown structure. The project planning team determines the logical sequence of tasks necessary to develop the system. At the same time, the planning team identifies those activities dependent on other activities (e.g., activities that can't start until another is finished). Correct sequencing drives efficiency in execution. However, scheduling decisions from activity definition to execution depend on the recognition and an appreciation of schedule factors that are often beyond the traditional scheduling considerations. This is reflected in the box, Function 2.1. For example, a WBS will often identify testing as an activity required to be performed many times during a development as initial assemblies are completed through integration of those assemblies into a component or subsystem. The WBS will also identify contractor reviews of testing results. However, the WBS cannot identify management attention manifested as questions to be answered (contractor and government) on the testing and potential retesting (rework), as well as emphasis on reviews that may occur if problems are identified, wherever they occurred.

This is where an appreciation of the project dynamics, the associated dynamic complexity, and ways of addressing dynamic complexity including system dynamics can be useful. While the normal scheduling process focuses on the actual tasks related to the completion of the development, system dynamics allows the addition of other, recognized relationships and their effects to the basic schedule. This allows the program manager to better anticipate potential problems.

Schedule Dynamics

Figure 4 is a graphical representation of the DoD "triple constraint" of cost, schedule, and performance. The goal on each axis is to move to the center, the cloud that depicts system completion. The red marks show the incremental attainment of the various targets of cost, schedule, and performance. The bi-directional arrows indicate the "one step forward, two steps back" progress often seen in system developments. For example, the contract point is a critical, established event, but one that is often revisited in the course of a system development. Cost is re-evaluated, performance is re-assessed, and schedules are redone. The dynamic changes occur in both directions representing the idea that the dynamics of the development consists of both success and failure (as measured through cost, schedule, and performance)—a back and forth.

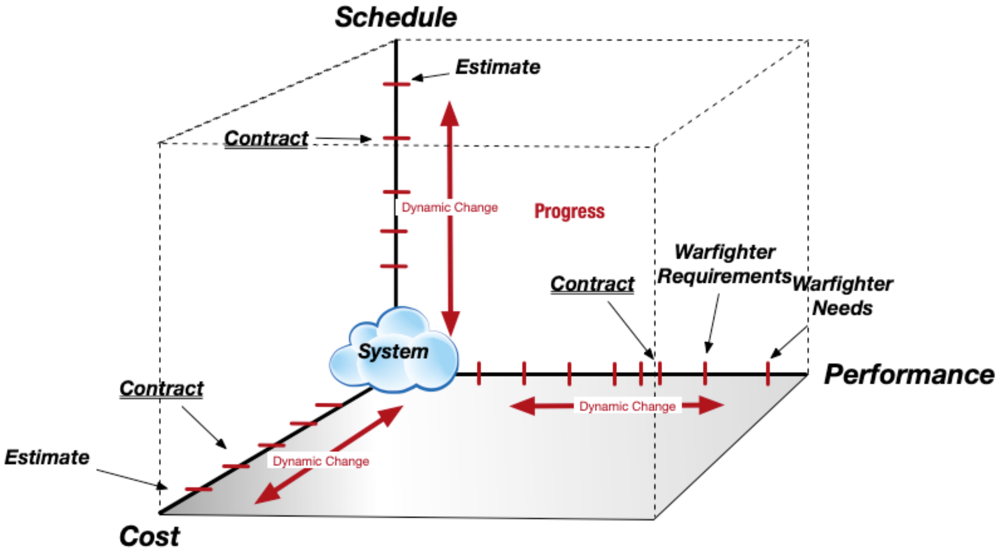

Figure 4. The Dynamic Environment of Cost, Schedule, and Performance

While Figure 4 emphasizes the dynamic nature of the entire project planning and execution process, it is at best a simplistic view of an extremely dynamic process. The current test of the whether the schedule was planned correctly and executed flawlessly is whether the cost, schedule, and performance axes are addressed and kept moving towards system completion. Unfortunately, in the defense world, we focus almost exclusively on cost, and to a lesser degree performance, while ignoring for the most part, the impact on and of the schedule. We have discussed this emphasis on cost and performance in previous papers.

Simply stated, the planning process—focused on cost, schedule, and performance—is itself a dynamic system. The activities on these three axes (and within the system that is the development project) change on their own through the dynamic processes of the development effort. In the execution process, the activities on these three axes are also changing. This movement creates time pressure forcing PMs to act, often with incomplete and/ or imperfect information. They can't wait to act before making a decision as failure to act, also has dynamic consequences.

> We cannot content ourselves with observing and analyzing situations at any single moment but must instead try to determine where the whole system is heading over time for many people, certainly those associated with weapon systems development, this is an extremely difficult task. (Dörner, 1997)

Consideration of the dynamic nature of the project/program management system is a question of the program manager's perspective, both government and industry contractor, and what is being measured. Anyone that has experienced a program review knows the focus is on quantitative metrics. Using an Integrated Master Schedule (IMS) and other quantitative tools including earned value, the review focuses on how we are performing to schedule and cost. We measure schedule and cost efficiency using accurate and extremely precise measures such as the cost and schedule performance index (CPI & SPI). Unfortunately, this accuracy can be misleading in light of the actual dynamics that are likely occurring. Culturally, we tend to accept metrics and computed numbers over real life. In fact, we often distort our view of real life because of computed interpretations.

The quantitatively measured progress of system development is potentially overestimated because of the focus on quantitative metrics at the expense of what is actually happening in a development (Cooper & Mullen, 1993). Specifically, the difference between the actual progress in a development effort and the actual completion rates can and often are very different. Those with project management experience will always recall the development project that slowly progresses until the "last 10%," which then takes an inordinate amount of time. That last 10% is most often due to rework, whether is a software development, hardware development, or an integration activity.

Unfortunately, the accuracy and precision afforded by our quantitative focus become accepted as the "ground truth," which leads to some of the problems we discuss in the F-35 development (Hennessy, 1996). Basically, we have created an illusion of accuracy and understanding that is not real. Further, this illusion can also affect our risk assessment, sometimes leading to false conclusions. That is not accurate. We frequently tend to ask questions focused on uncertainties. And we address the uncertainties through mathematical models based on deterministic statistical probabilities that fail to account for the exponential effects of interdependencies.

> Many projects fail to deliver against their targets because conventional project management techniques are failing to cope with the project's dynamic environment, complex interactions and the multitude of "soft"/people issues. (Mawby, 1999, p. 1)

Factors Affecting Schedule

A 1998 essay titled "Four Failures in Project Management" discusses what at the time were seen as some of the reasons for project management failure (Cooper, 1998). The essay describes the impact of a lack of systems thinking and a failure to appreciate the dynamics of a human-centered management process. Little has changed since 1998, and it is worthwhile not only to discuss the major points of that chapter, but to propose them as a framework to examine aspects of the F-35 development in the context of schedules.

The four failures are as follows:

- Failure to Know What to Expect
- Failure to Know What to Watch
- Failure to Know What to Do (and To Do It)
- Failure to Know What's What

Failure to know what to expect is about setting project targets including schedule:

> Setting and achieving an aggressive schedule is perhaps the most sacred of all sacred cows in the field of project management. It is also the source of the most destructive behavior and phenomena in projects. (Cooper 1998, p. 10)

The results of knowing what to expect are overlapped work stages, schedule pressure, resource inefficiencies, and worked morale (Cooper, 1998, p. 11).

Overlapped work stages occur when, in an effort to show progress, work is started that is scheduled later in the development in order to be able to show project progress, ultimately causing rework because of the out of sequence effects. Schedule pressure is just as it sounds: in an effort to demonstrate progress, the PM and management apply pressure on the workforce. The result of this pressure is a multiplication of the of out of sequence work and the resulting rework. Resource inefficiency occurs when the PM and management apply pressure, forcing overtime and other stress on the workforce.

Failure to know what to watch focuses on the idea of rework and the ultimate measure of quality. The “what to watch” aspect is about using perhaps the wrong tools to actually create schedules, and then not understanding what to do when rework happens. The basic challenge with the CPM/PERT scheduling method is that it does not account for what every PM knows occurs, which is rework. CPM/ PERT is a key problem because of its basic assumptions. The following are the basics of CPM/PERT:

- mean of activity duration = $(a + 4m + b)/6$
- standard deviation of activity duration = $(b - a)/6$

where a, m and b are the minimum, modal, and maximum of the activity duration

PERT uses four basic assumptions (Williams, 2002):

- there is a minimum, a maximum, and a median time provided by the estimator
- standard deviation is (^{1} / ^{6}) of the range $(b - a)$
- the distribution is $Beta$
- the activity durations are independent

The challenge with PERT is these assumptions. First, what is the max (a) and minimum (b)? What is the basis of these numbers? Second, given that a and b are estimates, how valid is the standard deviation? Third, why use a $beta$ distribution? Finally, although the network diagramming side of PERT is meant to disclose interdependencies and relationships, activity durations are rarely independent (Williams, 2002). The reality in today’s complex projects is that the traditional methods of creating schedules are not robust enough or even complete enough for what will inevitably occur. Regardless of the causes of rework, the fact is it always occurs. We see it in the case of the F-35 weight problem discussed below, as well as any human endeavor.

The third failure is failing to know what to do. This failure points directly at the decisions a PM makes and is a result of the dynamics of the system. The fact is that a PM can influence but is hard-pressed to actually control the execution of a complex project. On the industry side, the PM is captured by his or her organization and the organizational process, as well the matrix-driven organizational structure of most defense companies. Knowing what to do is about the decisions PMs make to influence the project. Because Cooper is focused on rework, the focus of this “failure” is concerned with the decisions about how to apply resources when the project gets in trouble. A perfect example of this failure is captured by Brook’s Law, to wit, “adding human resources to a late software project makes it later” (Brooks, 1995). The fact is, adding human resources to any project in progress has the unintended effect of slowing the overall project because the need to get the new people up to speed slows already slow progress, more workers end up getting in each other’s way, and communication among the team members becomes challenging with the increase in numbers (keeping everyone aware of status and changes). The final failure, “What’s What” relates to being able to learn. Otherwise known as lessons learned, this failure looks at an organization and its PMs’ ability to actually learn from previous problems.

Complexity plus the failures provide an initial framework for analyzing existing development programs in general, and the F-35 in particular. Combining the ideas of complexity expressed as structural dynamics and decision dynamics emphasizes the issues of weapons system development complexity and the dynamics these forces create. The failures provide a means to look at development programs from a different perspective, and may also serve to help explain some of the challenges demonstrated in these programs.

Case Study: Two Episodes From the F-35 Program

Overview

The Joint Strike Fighter was originally intended to meet modest expectations: basically, a timely and affordable replacement for the F-16, F-18, and AV-8. Nonetheless it evolved and ended up a very tough task at the beginning of system development (SDD), as noted in a 2001 DoD independent cost estimate, which rated the F-35 program as high risk for both schedule and technical reasons (not an open source, but discussed in Blickstein et al., 2011, p. 37).

In particular, the original list of requirements turned out to be a highly effective way to reduce engineering “trade space.” The F-35 requirements included being stealthy, supersonic, VSTOL capable (B model), and carrier capable (C model) (Blickstein et al., 2011, Table 4.6, p. 49).

The Narrow Path to Success

To accomplish a tough set of tasks in a timely manner, F-35 program management started with a number of highly optimistic fundamental premises (or “framing assumptions”). These, in turn, led to a program strategy that was success-oriented with little margin for error or surprises. Major assumptions included the following:[2]

JSF is readily available. Program management assumed the X-35 (a concept demonstrator) was a Y-35 (prototype for production, Blickstein, et al., 2011). This suggested that a development program (SDD) could proceed on an ambitious schedule and then transition quickly to full-rate production (~200 per year).

This time it’s different. The program was structured (perhaps implicitly) on the promise of improved manufacturing methods and reformed acquisition practices, even though their value in practice had yet to be demonstrated. For example, an abbreviated test schedule was planned, enabled by improved computer simulation capability (unnamed source, 2018).[3] Also, new manufacturing methods, such as unitized wing, would save both time in development and money in procurement.

However, as the program progressed, system testing was generally in a catch-up mode as data from experimental airframes and computer simulations proved less useful than expected. And for example, the unitized wing was abandoned to save aircraft weight (discussed more below) but with a doubling of assembly time (Warwick, 2018).

This time it’s the same. Cost estimates relied on experience gleaned from “legacy” aircraft, such as fourth-generation fighters, not accounting for, for example, increased complexity of the fifth generation. Also, the program started with a 6% weight growth allowance, in keeping with previous practice (Blickstein et al., 2011, p. 47).

Initial weight estimates used methods derived from experience with previous generations. But as one Lockheed Martin (LM) executive noted, “Legacy estimating techniques just don’t work with this family of airplanes,” which are highly complex, with densely-packed components in the airframe (Pappalardo, 2006).

---

[2] Franck et al. (2012) includes one discussion (esp. pp. 80–83).

[3] For which Chatham House Rules apply.

The real problem was perhaps less in the assumptions themselves, and more in the number. Even if each assumption was reasonable, it was also reasonable to expect that not all would work out. And if the road to success depends on all these bets coming in, the plan resembles a house of cards (cascading effects from small perturbations). In the event, the framing assumptions didn't all pan out, and the Joint Strike Fighter program got into trouble rather quickly.

The Weight Reduction and Redesign Episode of 2004

Because development of an operational platform was expected to be relatively quick and easy, initial design efforts could focus on cost ("affordability"), which included standard rather than custom parts. These measures added some weight. As one LM engineer put it, "The focus was very much on affordability at the time. People realized there was a penalty to be paid, and that was included in the weight estimates. It was higher than we thought" (Pappalardo, 2006). One likely reason for that situation is that LM's weight estimates were based on previous experience, as noted above.

The weight problems became obvious in 2003. The emerging F-35 design would be significantly over estimated weight, which would jeopardize meeting the program's KPPs (key performance parameters). Accordingly, weight was treated as an existential threat to the program, especially the STOVL model.

Weight Reduction Program Through Redesign (The Mother of All Rework Events)[4]

The weight problem brought the program to a "screeching halt" on April 7, 2004—with a "stand down" day. LM people were told that all work would stop until the weight problem was solved. This effort included substantial redesign work. LM's main focus shifted from affordability to "what's the lightest way to make it," according to another LM engineer.

The work was organized through a special project group called SWAT (Structural Weight Attack Team). SWAT was given very broad powers to waive LM's standard design change guidance and to offer incentives to employees who had weight reduction ideas. Supply chain firms were also involved and were credited with 586 pounds at the end.

Performance tradeoffs were likewise not off-limits. F-35B air-to-ground weapons carriage was reduced from two 2,000-lb bombs to 1,000 each. But a proposal to save structural weight through a reduction in maximum g-loads was disapproved by the DoD Joint Program Office (JPO).

In late 2004, LM declared victory. The exercise implemented more than 500 weight-loss recommendations. F-35B structural weight was reduced by 2,700 pounds; the A and C models 1,300 pounds each. Given the ingenuity of the engineering, some feelings of satisfaction were certainly warranted; according to one observer, "with SWAT, the program has a chance to come to fruition."

However, there were problems looming. One was cost. For example, "quick mate joints," which added 1,000 pounds to structural weight, were abandoned. To protect

---

[4] The main source for this section is Pappalardo's (2006) excellent article, "Weight Watchers."

commonality, the A and C models also lost their quick-mate joints. The result was an increase in manufacturing costs, due to “traditional, time-consuming” methods used instead.

Impacts known at the time were an increase in cost due to re-planning and an 18-month slip in the schedule, estimated at $6.2 billion and 18 months, respectively.

The Program Executive Officer at the time (PEO, Rear Admiral Steven Enewold) noted concerns going forward:

- increases in manufacturing costs (probably manageable);
- increased sustainability costs (unknown); and
- possible loss of durability-enhancing features (“good weight”), which was a matter of concern throughout the test program.

Continuing Concerns: That “Good Weight”

In some sense, the weight reduction exercise exchanged one set of problems for another. Among those problems was durability (operational life), especially for the B model.

Based on recent test data, the A and C models should last at least the planned operational life of 8,000 flight hours. However, estimates for expected B-model life vary considerably, from estimates of 2,100 (Trevithick, 2019) to 3,000 (DoD official), to well over 8,000 flight hours (LM, quoted in Trevithick, 2019). Part of this difference is due to characteristics of earlier vs. later production models (a result of program concurrency).

However, the F-35B encountered problems in durability testing that were significantly greater than the other models (e.g., DOT&E, 2010). At least some of this is due to the weight reduction exercise. For example, the 2010 DOT&E report on F-35 testing noted (p. 16), “The difference in bulkhead material is due to actions taken several years ago to reduce the weight of the STOVL aircraft. However, LM has recently stated that these problems are now solved: “The F-35B has completed full scale durability testing to 16,000 hours. Planned modifications and fleet management of the early contract F-35B aircraft will ensure that they meet the 8,000-hour service life requirement, and aircraft delivering today incorporate these design changes in the build process to ensure they’ll meet 8,000 hours or more” (Trevithick, 2019).

However, DoD’s Director of Operational Testing & Evaluation (DOT&E, 2019) had a less optimistic assessment for the B model. Early production units have expected operational lives significantly less than 8,000 hours, perhaps as little as 2,100 hours. This could mean B-model retirements as soon as 2026 or expensive retrofits. Moreover, the B-model was unable to complete its three-lifetime test profile, terminated due to numerous repairs on the test aircraft (p. 25).

Other issues have emerged. For example, a safety valve removal in 2008 (40+ pounds weight reduction) raised issues of aircraft vulnerability to combat damage (Copaccio, 2013).

An Engine Episode[5]

An interesting, and related, episode concerned the evolution of the F-119 engine (from the F-22) to the F-135.

In the early 1990s' programs, development efforts for a new strike fighter included with Advanced STOVL (ASTOVL) and Common Affordable Lightweight Fighter (CALF) programs. At this time, the strike fighter was viewed as being lightweight; one F-119 engine was deemed sufficient.[6] The problem emerged when specifications grew with JAST, and affordability was pursued, accepting increases in weight. The weight problem was not discovered quickly because of the parametric weight estimating models discussed above (Pappalardo, 2006; Warwick, 2018).

With increased requirements came an effort to increase F-119 thrust; at some point, the upgraded F-119 became the F-135. With the upgrade came a change in the JSF morphology, which necessitated a redesign, with a number of cascading effects, as reported in the RAND Root Cause Analysis (Blickstein et al., 2011). This RAND analysis reported and cascading major effects from this upgrade:

> Changes in the engine contributed to the weight growth of the JSF. Original plans called for the JSF to use the same engine as the F-22—the F-119 engine. However, the F-119 proved to be underpowered for the performance desired of the F-35, so the F-119 engine was altered to generate more thrust and became the F-135 engine. By enlarging the F-119 engine into the F-135 engine, engineering issues such as shaft length and efficiency had to be dealt with. However, the increase in thrust also lead to an increase in the engine size by a reported 1.5 inches in diameter.[7] *This small change in the engine generated a need to redesign the airframe, which in turn changed everything from aerodynamics to stealth signature, all of which needed to be re-baselined.* This engine issue also indicates lack of integration across the major contractors, which was Lockheed's responsibility as the prime contractor. (Blickstein et al., 2011; emphasis added)

However, the record also indicates that a need for a redesigned F-119 engine with increased thrust was recognized early in the program. That was a significant part of a 1997 contract with Pratt & Whitney in 1997 (Keijsper, 2007, p. 192). PW received a 10-year contract to develop the F-135 ("evolved" F-119) shortly after the F-35 source selection (over

---

[5] This is not THE engine episode. Another—in 2014—involved an engine fire traced to engine fan blades rubbing against their grooves.

[6] For example, the ASTOVL program was bound to an empty weight of 24,000 pounds. (Global Security, CALF). The F-119 was capable of supporting STOVL operations at that weight. However, the empty weight of the F-35A is about 29,000 lbs, an increase of 20+%. The F-135 max thrust is about 43,000 lbs and increase of 20+% above the F-119. So, using this back-of-the-envelope comparison, development of the F-135 makes good sense.

[7] There is some ambiguity in the open literature. Standard sources state that the F-119 and F-135 have the same diameter. However, the F-135 is longer: 220 inches vs. 203.

Boeing's F-32) in October 2001 (Global Security, Pratt & Whitney F-135 Engine). The first F-135 production unit was delivered in 2009 (Pratt & Whitney F-135, 2019).[8]

Engine Development: From F-119 To F-135

Although this paper focuses on the JSF program after Milestone B, events that preceded selection of the F-35 provide useful context. In May 1994, the Joint Advanced Strike Technology (JAST) program began. Early on the program focused on a single-engine, one-crewmember approach with affordability being a significant part of the rationale.

In July, the Advanced STOVL (ASTOVL) program chose GE, PW (with Allison) to conduct derivative engine studies, leading to demonstrations in FY97. Major issues at the time included single-engine reliability (Navy concern) and thrust.

The JAST and ASTOVL programs merged in October 1994 as JAST. In November, contracts were let for preliminary design of F-119 derivative. GE F-120 received less funding as an alternate engine.

In December 1994, Boeing, Lockheed Martin (LM) and McDonnell Douglas (with BAE) received 15-month conceptual design contracts. In the spring of 1995, all three JAST contractor teams choose the PW F-119 as the preferred engine for their development aircraft (JAST, n.d.).

In May 1996, the JAST program was renamed Joint Strike Fighter (JSF). In January 1997, PW received a contract to develop F-119 derivatives for the Boeing and LM test aircraft (Keijsper, 2007, p. 193). The DoD chose a Government Furnished Equipment (GFE) engine approach. That is, PW would supply engines to the government, which would then be delivered to Boeing and LM as GFE. There were various STOVL-variant problems. But most ground test objectives were met by the end of 2000.

After Milestone B

On October 26, 2001 (shortly after source selection), Pratt and Whitney received a 10-year contract for the design, development, fabrication, and test of the F-135 propulsion system and supporting equipment. It included system test and evaluation. PW was also to provide engines suitable for the F-35 flight testing program ("Pratt & Whitney F-135 Engine," n.d.).

PW assembled its first CTOL/CV test engine in September 2003 and conducted a successful test in October. The first F-135 STOVL propulsion system tests began on April 14, 2004.

In retrospect, however, the maturing engine and airframe designs were not proceeding as a coherent whole. What apparently happened was the F-135 was in development, with implications of the evolving new engine not yet fully known to the LM airframers (Blickstein et al., 2011). In retrospect, this was likely one factor in LM's overreliance on parametric weight estimations (Pappalardo, 2006). If so, it also means that LM not only had to rework the fuselage to save weight, but also to change the fuselage itself to deal with the F-135 engine.

---

[8] The Wikipedia article references a 2009 PW press release. That link is now broken.

Given the RAND findings above, it appears the F-135 (relative to F-119) was not jointly understood by PW and LM. The RAND Root Cause Analysis offers the hypothesis that LM failed to carry out this part of its prime contractor responsibilities (Blickstein et al., 2011).

Another interesting hypothesis is that the DoD decided to deal directly with Pratt & Whitney for the various engine variants associated with the Boeing (X-32) and Lockheed Martin (X-35) development efforts. The DoD would then deliver the engines to the airframers. In effect, DoD was the middleman in these transactions, which is unlikely to have improved information flow from PW to Boeing and LM.[9] That the F-135 (née F-119 variant) was in development at the same time as the F-35 airframe (and a DoD responsibility to boot) might well have been factor contributing to this outcome.

Another factor is that the F-35 airframe and the F-135 engine designs were progressing concurrently. PW assembled its first test CTOL engine in September 2003, and its first test STOVL engine in April 2004. In that regard, it's interesting that LM formed a special team (BRAT) over the 2002–2003 time period to address weight issues and brought F-35 development to a sudden halt, and commenced a redesign effort in April 2004, with a special team called SWAT (Pappalardo, 2006).[10] Engine-airframe program concurrency was a possible factor leading to the weight reduction and redesign episode of 2004.

Conclusion

We have argued that the act of weapon system development is, in and of itself, a system. Because it is a system, it has internal interrelationships and interdependencies that can fundamentally change the internal processes and outputs of that system. The F-35 activities described above are witness to that fact.

Further, we believe the F-35 discussions above serve as examples of the "F-35 Program System" and are thus susceptible to the complexity factors, as well as the four failures Cooper described. The complexity issues create an environment for the failures to occur.

Using the discussion on systems, complexity, and the Four Failures, Table 1 is a summary of the impact the dynamics of complexity and rework can have on a weapon systems development.

---

[9] Given there were two proposals in plan (Boeing and LM), the GFE approach for engines was likely reasonable at the time. However, it did have disadvantages that appeared later.

[10] However, it doesn't appear that the F-135 core engine weight was a problem. The F-119 "dry" weight is 3,900 lbs, while the F-135 weight is 3,750 lbs. Also, the F-119 and F-135 are described as having the same diameter (46 inches). However, the F-135 overall length (including tailpipe) is 17 inches longer (220 vs. 203).

Table 1. F-35 Dynamic Challenges

Events	Failure Modes			
	What to Expect *Excessive optimism in planning/ estimating the schedule*	**What to Watch** *Understanding and accepting the impact of rework*	**What to Do** *Understanding the impacts of complexity and feedback loops*	**What's What** *Lessons learned*
X-35 Prototype Assumption	D			D
Def Acquisition Reform Benefits	S			
Success-Based Development Strategy	S		S	
Cost-Reduction Exercise		D	S,D	D
Estimation Methods		S		S
Weight Reduction Exercise		S,D	S,D	
F-135 to F-119 Evolution		D	S,D	

Note. D = Decision Dynamics; S = Structural Dynamics

Measuring Performance in A Network-Centric-Combat Environment[11]

As indicated in previous reports (e.g., Franck et al., 2017, 2018), there are good reasons to consider the issue of performance measures in developing tools to analyze acquisition schedules. However, performance has become less a matter of platform attributes and more about what the new system adds to capabilities in an information-rich, networked, system-of-systems operational environment. “You look at an effect which you want to create with the overall force and you look at your mix of platforms and determine

---

[11] This section is abridged to conform with proceedings page limits. A more detailed discussion will appear in our final project paper.

which can lead the design change to achieve that effect” (John Blackburn, quoted in Laird, 2018, p. 4).

Also, program managers are (or should be) mindful of trade-offs being made. among the goals of cost, performance, and schedules (CJCS, 2015, p. A-9). With a better understanding of system performance in contemporary operational environments, such decisions could be improved.

Finally, useful measures of system performance can be a useful in estimating schedules—in schedule estimating relationships among other things.

There have been serious efforts in the past to formulate scalar performance measurements. However, previous efforts (e.g., Regan & Voigt, 1988) focused almost completely on platform characteristics and not on force characteristics. Operational capability is no longer a matter by adding up platform characteristics across the force, but by how a mix of different platform types operate together in the combat environment of the near future. As one observer put it, “the focus is less on what organically can be delivered by a new proposed new fighter than on its ability to interact with other platforms to deliver the desired combat effect” (Laird, 2019).

Accordingly, this section builds on previous reports (Franck et al., 2017, 2018) with a more general (but still simple) model of air combat in the near future. The essential features of our assumed scenario are as follows:

- two modern, high-technology air combat forces (Blue and Red);
- widely shared (but varying) operational situational awareness;[12]
- decentralized allocation of weapons to identified targets (like a “combat cloud,” Deptula, 2016, esp. p. 3);
- heterogeneous forces,[13] consisting of stealthy scouts (e.g., F-35), and less stealthy weapons carriers (e.g., F-15X);

Winning this engagement (as in all Lanchester-based models) requires inflicting losses on the opposing side. Accordingly, we examine the effects on air battle results of the following variables:

- Relative force sizes (R/B): even with high technology platforms and sophisticated networks, numbers probably still matter a great deal.
- Stealthy aircraft (Scouts) are survivable in a high-threat environment, while Weapons Carriers are not (Harrigian & Marosko, 2016, pp. 2–4, 7).
- Weapons Lethality, measured as a probability of success (kill).
- Battle management capabilities. It’s not “super simple” and “just battle management” (Miller, 2016) after targets have been identified. Moreover, it appears that contemporary combat air arms, such as the U.S. Air Force, do understand these difficulties (USAF, 2016, p. 6).

---

[12] We understand that fully shared situational awareness is still a work in progress (e.g., Laird, 2019).

[13] The U.S. Air Force Air Superiority Flight Plan for 2030 specifically calls for both “stand-in” (stealthy fighters) and “stand-off” (weapons carriers) airborne combat forces (USAF, 2016, esp. p. 7).

The Model

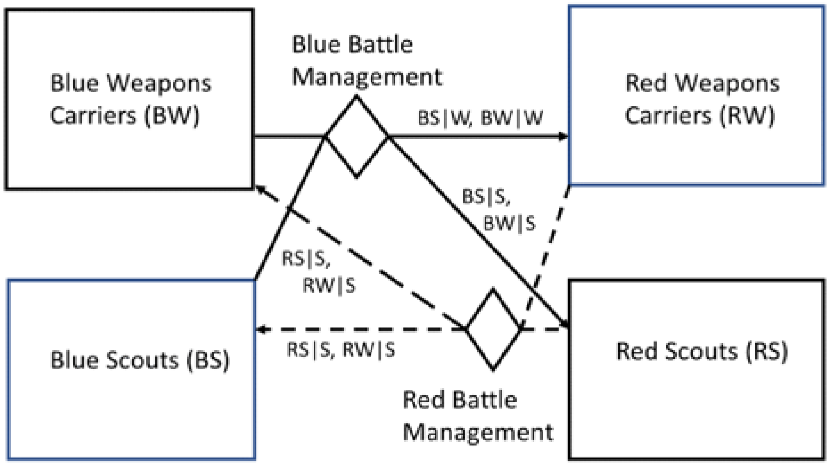

Figure 5. Representation of a Generalized Lanchester Model of Air Combat

Notation. XY|Z is side X (Blue or Red) units of type Y (Scout or Weapons Carrier) targeted against the opposing side's units of Type Z. For example, BS|W is number of Blue Scouts assigned against Red Weapons Carriers.

As noted above, our model involves an engagement of heterogeneous air combat forces: with stealthy scouts (with weapons) and non-stealthy weapons carriers.

Within that framework, we can consider effects of numbers, weapons lethality, stealth, and battle management effectiveness. A battle management decision process assigns Blue (Red) forces to Red (Blue) targets (that are detected and tracked). The air combat assets (both types) then attack their assigned targets.

By varying values for Blue (with Red characteristics held constant), what emerges is both interesting and suggestive. The various capabilities can be substitutes; that is, capability gaps in one characteristic can compensate for shortfalls in another characteristic. For example, Figure 6 depicts battle outcomes primarily as a function of Blue Stealth and Blue Lethality. To the upper right, Blue wins; at the lower left, Red wins. There are two curved corresponding to two levels of battle management denote a "tie."

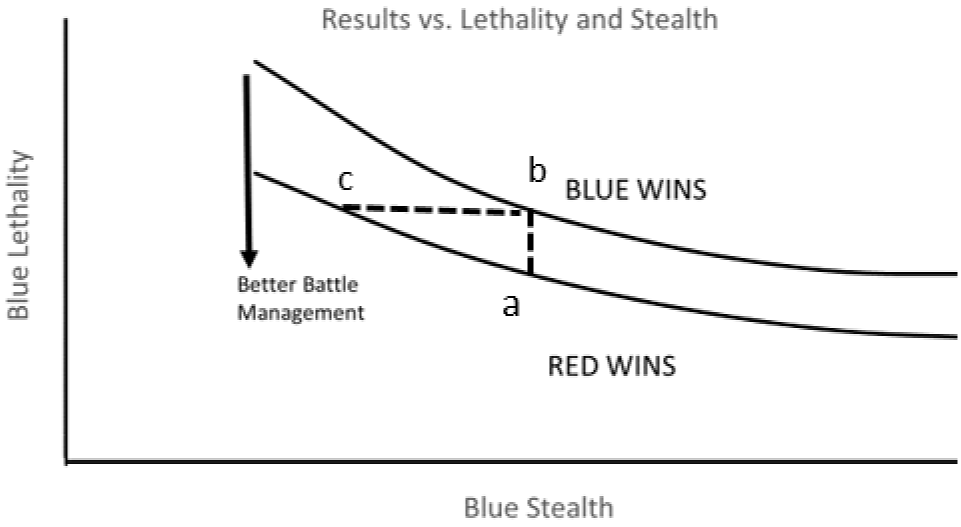

Figure 6. Representation of a Generalized Lanchester Model of Air Combat

Notation. XY|Z is side X (Blue or Red) units of type Y (Scout or Weapons Carrier) targeted against the opposing side's units of Type Z. For example, BS|W is number of Blue Scouts assigned against Red Weapons Carriers.

Also interesting is the relative percentage change in engagement outcome with changes in force ratio,[14] stealth, and battle management (against a Red with specified "baseline" capabilities).

These are given in Table 2.

Table 2: Responsiveness of Outcome to Changes in Force Characteristics

Variable	Force Ratio	Stealth	Lethality	Battle Management
Outcome/Variable Responsiveness	11.3	6.3	1.6	6.0

The magnitude of the numbers themselves should not be taken too seriously. The outcome variable is a measure of the margin of victory over the Red force (or defeat) rather than a raw measure of capability. In any situation of forces with about the same overall capability, any small change ("edge") can have a major effect on the margin of victory. However, the relative values are nonetheless interesting.

---

[14] This is an "elasticity," basically a ratio of percentage changes in outcome and force characteristic.

References

Baccarini, D. (1996). The concept of project complexity—A review. *International Journal of Project Management*, *14*(4), 201–204.

Blickstein, I. et al. (2011). Root cause analyses of Nunn-McCurdy breaches, Volume 1 Joint Strike Fighter. RAND Corporation. Retrieved from https://www.rand.org/pubs/monographs/MG1171z1.html

Brooks, F. P. (1995). *The mythical man-month: Essays on software engineering, anniversary edition, 2/E.* Addison-Wesley.

CALF. (n.d.). DARPA/Navy Common Affordable Lightweight Fighter (CALF) 1993–1994. Global Security. Retrieved February 22, 2019, from https://www.globalsecurity.org/military/systems/aircraft/calf.htm

Cooper, K. G., & Mullen, T. (1993). Swords and plowshares: The rework cycles of defense and commercial software development projects. *American Programmer*, 41–51.

Cooper, K. G. (1998). Four failures in project management. In *The Project Management Institute project management handbook (Jossey-Bass Business & Management Series)*. Jossey-Bass.

Copaccio, T. (2013). Pentagon mulls restoring F-35 safety gear to reduce risk. Retrieved from http://www.f-16.net/forum/viewtopic.php?t=22306&p=252263

Deptula, D. (2016, September). *Evolving technologies and warfare in the 21st century: Introducing the "Combat Cloud"* (Mitchell Institute Policy Papers, Vol. 4). Retrieved from http://docs.wixstatic.com/ugd/a2dd91_73faf7274e9c4e4ca605004dc6628a88.pdf

Dörner, D. (1990). The logic of failure. *Philosophical Transactions of the Royal Society*, *327*(1241), 463–473. http://doi.org/10.1098/rstb.1990.0089

Dörner, D. (1997). *The logic of failure*. Cambridge, MA: Perseus.

DOT&E (Director, Operational Test & Evaluation). (2010). Joint Strike Fighter. Retrieved from http://www.dote.osd.mil/pub/reports/FY2010/pdf/dod/2010f35jsf.pdf

DOT&E (Director of Operational Test & Evaluation). (2019). F-35 Joint Strike Fighter (JSF). Retrieved from http://www.dote.osd.mil/pub/reports/FY2018/pdf/dod/2018f35jsf.pdf

Drezner, J., & Smith, G. K. (1990). *An analysis of weapon system acquisition schedules*.

F-135 Engine. (n.d.). Pratt & Whitney F-135 engine. Global Security. Retrieved February 16, 2019, from https://www.globalsecurity.org/military/systems/aircraft/systems/f135.htm

Franck, R., Hildebrandt, G., & Udis, B. (2016). Toward realistic acquisition schedule estimates. Presented at the 13th Annual Acquisition Research Symposium, Monterey, CA.

Franck, R., Hildebrandt, G., Pickar, C., & Udis, B. (2017). Realistic acquisition schedule estimates: A follow-on inquiry. Presented at the 14th Annual Acquisition Research Symposium, Monterey, CA.

Franck, R., Hildebrandt, G., Pickar, C., & Udis, B. (2018). In the fullness of time: Towards realistic acquisition schedule estimates. Presented at the 15th Annual Acquisition Research Symposium, Monterey, CA.

Franck, R., Lewis, I., Simmonds, H., & Udis, B. (2012). *International defense acquisition management and the Fifth-Generation Fighter quandary*. Monterey, CA: Naval Postgraduate School. Retrieved from http://www.acquisitionresearch.net/files/FY2012/NPS-AM-12-214.pdf

Harrigian, J., & Marosko, M. (2016, July). *Fifth generation air combat: Maintaining the Joint Force advantage* (Michell Forum No. 6). Mitchell Institute. Retrieved from http://docs.wixstatic.com/ugd/a2dd91_bd906e69631146079c4d082d0eda1d68.pdf

Hennessy, G. (1996). From spreadsheets to system dynamic models. *The System Thinker*, 1–3.

JAST. (n.d.). Joint Advanced Strike Technology (JAST). Global Security. Retrieved February 22, 2019, from https://www.globalsecurity.org/military/systems/aircraft/jast.htm

Keijsper, G. (2007). *Joint Strike Fighter*. S. Yorkshire, England: Pen & Sword.

Marca, D. A., & McGowan, C. L. (2005). *IDEF0 and SADT: A modeler's guide*. Auburndale, MA: OpenProcess.

Laird, R. (2018, October 15). F-35 2.0 and changing the defense ecosystem. In *F-35 2.0, Second line of defense information*. Retrieved from https://sldinfo.com/wp-content/uploads/2018/09/F-35-2.0.pdf

Laird, R. (2019, February 6). Recrafting the fighter role. *Defense Information.* Retrieved from https://defense.info/air-power-dynamics/2019/02/recrafting-the-fighter-role/

Mawby, D. (1999). Managing the rework cycle not the schedule: A project management paradigm for the future? *INCOSE International Symposium*, *9*(1), 1175–1182. Retrieved from http://doi.org/10.1002/j.2334-5837.1999.tb00289.x

Pappalardo, J. (2006). Weight watchers: How a team of engineers and a crash diet saved the Joint Strike Fighter. *Air & Space Magazine*. Retrieved from https://www.airspacemag.com/military-aviation/weight-watchers-13117183/#AD2ai6Ui9K7c5tdv.99

Perrow, C. (1999). *Normal accidents: Living with high-risk technologies*. Princeton, NJ: Princeton University Press.

Pratt & Whitney F-135. (2018). In *Wikipedia*. Retrieved September 25, 2018, from https://en.wikipedia.org/wiki/Pratt_%26_Whitney_F135

Pugh, P. (1987). *Where does the time go? A review of development time scales for aircraft and guided weapons*. Presented at the Development Time Scales-Estimation and Control Royal Aeronautical Society.

Rothman, M. B. (1987). *Aerospace weapon system acquisition milestones: A data base*.

Senge, P. M. (2006). *The fifth discipline* (pp. 1–368). New York, NY: The Crown Publishing Group.

Stumpf, G. R. (2000). Schedule delay analysis. *Cost Engineering*, *42*.

Trevithick, J. (2019). USMC's older F-35Bs may only be able to fly around a quarter of their expected service life (Updated). *The Drive*. Retrieved from http://www.thedrive.com/the-war-zone/26269/usmcs-older-f-35bs-may-only-be-able-to-fly-around-a-quarter-of-their-expected-service-life

Regan, J. M., & Voigt, W. J. (1988, July 31). *The TASCFORM methodology: A technique for assessing comparative force modernization* (3rd ed.; Contract No. MDA903-85-C-908). Arlington, VA: The Analytic Sciences Corporation.

Tyson, K. W., Nelson, J. R., & Om, N. I. (1989). *Acquiring major systems: Cost and schedule trends and acquisition initiative effectiveness* (IDA Paper P-2201). Institute for Defense Analysis.

U.S. Air Force Enterprise Capability Collaboration Team. (2016, May). *Air superiority: 2030 flight plan*. Retrieved from https://www.af.mil/Portals/1/documents/airpower/Air%20Superiority%202030%20Flight%20Plan.pdf

Van Atta, R. H., Kneece, R., Patterson, C. M., & Hermes, A. C. (2015). *Assessing weapon system acquisition cycle times: Setting program schedules*. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/1014965.pdf

Van de Vonder, S., Demeulemeester, E., & Herroelen, W. (2007). A classification of predictive-reactive project scheduling procedures. *Journal of Scheduling*, *10*(3), 195–207. Retrieved from http://doi.org/10.1007/s10951-007-0011-2

Vandevoorde, S., & Vanhoucke, M. (2006). A comparison of different project duration forecasting methods using earned value metrics. *International Journal of Project Management*, *24*(4), 289–302. Retrieved from http://doi.org/10.1016/j.ijproman.2005.10.004

Warwick, G. (2018, September 17). How F-35 experience could reduce hurdles to developing fighters. *Aviation Week*, p. 44. Retrieved from http://aviationweek.com/air-dominance/how-f-35-experience-could-reduce-hurdles-developing-fighters

Williams, T. (2002). *Modelling complex projects*. West Sussex, England: John Wiley & Sons.

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

www.acquisitionresearch.net