SYM-AM-24-049

Excerpt from the Proceedings of the Twenty-First Annual Acquisition Research Symposium

Acquisition Research: Creating Synergy for Informed Change

May 8–9, 2024

Published: April 30, 2024

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

The research presented in this report was supported by the Acquisition Research Program at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School

Navigating the Labyrinth: Unraveling Schedule Complexity

Charles Pickar—retired from the Naval Postgraduate School (NPS) faculty, where he taught 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 also 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. HiHeas also served as 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@gmail.com]

Raymond Franck—retired from the NPS faculty in 2012. He retired from the Air Force in 2000 in the grade of Brigadier General. His active-duty career included operational tours, staff positions, leadership assignments and, finally, head of the Department of Economics and Geography, United States Air Force Academy. His published work includes journal articles, book chapters, and research reports on military innovation and defense acquisition management. [cfranck215@aol.com]

Abstract

This paper continues our efforts to help advance the state of the art for estimating defense acquisition program schedules. Accurate schedule estimates provide valuable benchmarks for program managers and reliable dates for the availability of new systems for warfighters. But credible (much less accurate) schedule estimates are difficult, at best. This is due to many factors, including programs' inherent complexities and the likelihood of unfavorable developments outside the control of program management. Recognizing this difficulty, our inquiries center on improving the art of estimating schedules, and also making program schedules more resilient with respect to those difficulties. Accordingly, we also focus on the role of Systems Dynamics in program execution (illustrated with case studies) and on improving management information through prediction markets. Despite the difficulties, we remain convinced that improvements are possible.

Introduction and Overview

We have identified three significant areas of interest in our continuing inquiries into the art and science of managing acquisition schedules. First, credible, available prior to program start (*ex ante*); second, during program execution (*in media res*); and finally, lessons learned from completed programs (ex *post*).

However, even the best ex-ante schedule estimates do not fully prepare program management for the difficulties ahead. In this paper, we focus on project dynamics and the difficulty (arguably impossibility) of foreseeing what unforeseen events (some external to the program itself) can affect acquisition schedules.

A central theme of our previous papers (e.g., Franck et al., 2017) is understanding project dynamics in schedule execution in weapon system development projects. Scheduling is a unique and challenging problem requiring mathematical and human perspectives.

A system is complex if it has many elements, interactions, and variables, and if it is difficult to predict its behavior and outcomes. A system is dynamic if it changes over time and if its changes depend on its current state and history. Projects and weapon system development programs are systems. Typically, such programs are dynamic, complex, interrelated, and changing systems. Project dynamics refer to those internal and external factors that can influence the course of a development project throughout its life cycle. These factors can be unpredictable and constantly evolving, posing significant challenges to maintaining a project schedule. According to conventional wisdom, project management can plan and execute projects by identifying and controlling the cause-and-effect relationships among various factors inside the system. However, this approach is inadequate in dealing with the uncertainties and interdependencies within the system and during the project life cycle. Since most people are prone to linear thinking, they have difficulty with systemic and nonlinear thinking, especially when faced with complex and dynamic situations that involve multiple variables, uncertainties, time delays, and trade-offs (in abundance in weapon system development projects; Dörner, 1996). Dörner also notes that people often fail to recognize the interrelated and long-term effects of their decisions, leading to unintended and sometimes disastrous outcomes (Dörner, 1996, p.45). The fact is that 50% of technology projects fail to meet their cost and schedule objectives (Flyvbjerg & Budzier, 2015).

Closely related to the idea of dynamics is the value of using data analysis for schedule planning and execution. Data analysis of past projects can provide insight into the problems experienced in other developments. Data can provide insights into the project's progress, resource allocation, and potential roadblocks, allowing project managers to make informed decisions. Data can also help to give indicators of the project's progress against the schedule, allowing for timely interventions and adjustments. Data can help identify problem areas and opportunities for improvement, leading to more efficient project execution. In past papers, we have used data to better understand trends in schedule execution, as well as provide insight into the causes of schedule delays. To identify the causes of schedule execution dynamics, collecting and analyzing data related to these factors, such as historical, current, or forecast data, is necessary. A significant data source for qualitative and quantitative analysis of previous studies was the Selected Acquisition Report (SAR). We reference some past data-focused studies on the determinants of schedule delays (Pickar, 2018). We had also hoped to make greater use of Selected Acquisition Reports for this study, but the lack of data for the two case studies in this paper prevented their use.[1]

So, while reasonable *ex-ante* schedule estimates are helpful, they are insufficient to successfully complete the program. Many problems in today's defense acquisition world occur because managers fail to recognize that the various tasks in the program are not isolated problems; they interact within their own system through personnel and activities and with multiple stakeholders, environments, and events that can influence and change project outcomes. Therefore, project managers need to adopt a more dynamic and adaptive mindset to anticipate and respond to the dynamics of complex projects. In short, project managers must acknowledge defense projects' dynamic nature.

Knowledge of program dynamics can conceptually equip program managers to cope with and mitigate the ill effects of unpredicted developments. Such knowledge can provide a basis for making weapon systems development programs more resilient to untoward developments. Moreover, the quantity (and quality) of information available to program management can significantly increase resilience in program execution.

In this context, recent developments in data analytics can provide valuable information for getting acquisition programs through their generally turbulent stages. It's also vital that data quality includes the situational contexts from which it is filtered through hierarchies and widely distributed. Therefore, including local knowledge in the program's

---

[1] In past research, we used two main sections of the SAR: the Executive Summary and the Schedule Change Explanations. The F-35 Executive Summary changes very little from reporting period to reporting period. The Schedule changes provide information of what changes, but do not always provide the reasons for the schedule changes.

management information system is essential. One way to do that is through "prediction markets."[2]

In this paper, we consider the art of managing defense acquisition programs as complicated and incompletely understood systems through the lens of system dynamics—which we illustrate through case studies. We also consider improving the quality of information available to decision-makers through "prediction markets" regarding upcoming program events and milestones.

System Dynamics and Rework in Schedules

Scheduling is about allocating resources, tasks, and time to achieve a desired outcome in a project. It is a crucial aspect of project management, affecting the project's quality, cost, and scope. However, scheduling is neither straightforward nor simple, as it involves many uncertainties, complexities, and interdependencies. Moreover, scheduling is not only a technical or mathematical problem but also a human and social one, as it involves the behavior, motivation, and interaction of the project manager, team, and stakeholders.

In a previous paper, we introduced the concept of rework (Franck et al., 2017). Figure 1 shows the rework cycle (Richardson, & Pugh, 1981; Cooper, 1993a). Rework is correcting or modifying work that has already been completed. It is common in many projects and can significantly impact project outcomes in terms of cost, schedule, and performance.

The rework cycle is a feedback loop that describes the negative and positive effects of rework on project dynamics. The rework cycle is a critical aspect of system dynamics studied extensively in project management. The rework cycle is characterized by feedback effects that can be negative, controlling, or positive, re-enforcing, often "vicious circles." The rework cycle can also lead to knock-on or domino effects within or between work phases and between projects. Software engineering and development is prone to rework due to errors, changes, and learning in the software development process (Abdel-Hamid & Madnick, 1982, 1983). Rework can affect the cost, schedule, and performance of software projects and the developers' and users' satisfaction and motivation.

Figure 1. The Rework Cycle (adapted from **(Cooper, 1993a)**

[2] Prediction markets are described below.

Rework is a common and pervasive phenomenon in project management. It occurs when a task or process must be redone or corrected because it was not done correctly or completely the first time. Rework can have negative consequences for a system's cost, schedule, and performance, as well as for the satisfaction and motivation of the stakeholders involved. It can also create delays, waste, errors, and rework cycles, worsening the system's behavior and outcomes.

Rework is influenced by various dynamics, such as the complexity and uncertainty of the task or the process, the quality and availability of the inputs and resources, the skills and experience of the workers, the communication and coordination among the participants, the feedback and control mechanisms, the standards and expectations, and the external environment and disturbances. Rework can also vary in nature, extent, and impact, depending on the type, stage, and scope of the task or the process, as well as the detection and correction time and method (Love et al., 2002).

There is a general agreement that rework involves the need to redo or correct a task or a process that was not done correctly or completely the first time. However, the criteria and standards to determine what is correct or complete can vary depending on the context and the stakeholders' perspective. For example, a task or a process can be considered correct or complete from a technical or functional point of view but not from a customer or user point of view, or vice versa. Similarly, a task or a process can be considered correct or complete at a given time but not later due to changes in the requirements, specifications, or expectations.

Our past research agenda has included examination of the reasons for schedule delays. We have used qualitative and quantitative techniques to identify the causes of schedule delays (extracting the data from the Selected Acquisition Reports). Of course, these studies depend on the PM reported reasons for the delays. These studies yielded the following broad categories of causes.

Table 1. Schedule Delay Factors. (Pickar, 2018, p. 82).

Schedule Delay Factor	# instances
Administrative changes to the schedule, including updates to APB, ADM changes as well as changes resulting from Nunn-McCurdy processes and program restructuring	460
Technical	291
Testing delays	283
Delay in availability of key capabilities/facilities (launch vehicle/testing facilities/IOT&E units)	3
Budget/funding delays	52
Delays attributed to the contractor	50
Delays because of rework	16
External events such as inflation, earthquakes, labor strikes, etc. *(force majeure)*	4
Delays due to contracting/contract negotiation/award delays	29
Actuals (updating previously reported dates to actual occurrence)	172

The number of instances of rework cited is of note. This study examined 20 years of Selected Acquisition Reports (SAR). The delay causes were extracted from the SARs' reported schedule explanations. The low number of reported cases of rework could come from two reasons. The first is that the contractor may not have reported specific instances of rework instead of the result of the rework (e.g., contractor delays or the catch-all category of administrative problems) to the government program office. The second reason could be a lack of understanding of the factors causing rework and the resulting problems caused by rework.

A systems approach to project management provides project management organizations insight into the systemic causes and consequences of rework, enabling them to make more informed decisions and implement effective strategies for its prevention or mitigation. This concept highlights the importance of a holistic and dynamic perspective when analyzing and managing rework, recognizing its role in shaping the behavior and performance of weapon system development. The case studies provide examples of the impacts of rework.

Case Study: F-35 Weight and Other Rework Problems

The F-35 program is the most extensive and expensive military procurement project in history, with an estimated total cost of over $1.5 trillion. This weapon system development is a classic example of a complex dynamic system. The F-35 program has faced numerous programmatic and engineering challenges. One of the most critical and persistent problems was the aircraft's weight, which threatened to compromise its performance, stealth, and safety. In a previous study, we introduced the weight problem from a different perspective (Pickar et al., 2019). This case study examines macro-level F-35 schedule issues from the system dynamics perspective.

Schedule Problems

The F-35 program has experienced significant schedule problems since its inception. The original schedule for the system development and demonstration phase was to complete the initial operational test and evaluation by 2012 and to achieve the initial operational capability by 2013. However, the schedule has been repeatedly revised and extended due to various technical, operational, and political factors. Factors affecting the schedule include the technical complexity of the aircraft and uncertainty associated with the operational and political environment. From a technology perspective, the F-35 is a highly complex and sophisticated system incorporating advanced technologies such as stealth, sensor fusion, and network-centric warfare. It must also meet the diverse and demanding requirements of the different variants and customers. The F-35's technical complexity has resulted in numerous design changes, software issues, testing challenges, and reliability problems, which have caused delays and rework in the development and production processes.

Operational uncertainty reflects the dynamic complexity and uncertain tactical environment the aircraft operates in, as well as the constantly evolving threats, missions, and scenarios project management must address daily. The F-35 also must adapt to the changing needs and expectations of domestic and international customers, who may have changing operational priorities, preferences, and constraints. The operational uncertainty of the F-35 has resulted in frequent requirement changes, specification revisions, and performance trade-offs, which have increased the scope and complexity of the development and production processes. More specific schedule delay issues are associated with TR-3 upgrades, simulator development, and Block 4 modernization. All these issues are potential rework problems.

Rework is a cycle of corrective actions or changes within a project involving feedback effects, knock-on effects, and implications for project management (Rahmandad & Hu, 2010). Rework issues in the F-35 development are omnipresent. These include seemingly simple problems, such as the prime contractor installing a valve backward, as recounted in this script excerpt from the CBS *60 Minutes* program, broadcast on February 16, 2014. The scene is a morning staff meeting between the F-35 Program Executive Officer (PEO), LtGen Chris Bogdan, and his staff (Walsh & Martin, 2014).

CBS 60 Minutes, 02/16/2014

"Lt. Gen. Chris Bogdan is the man in charge of the F-35 and every morning starts with problems that have to be dealt with ASAP. This morning it's a valve that's been installed backwards and has to be replaced.

Chris Bogdan: How long does it take?

Answer: It's about a seven-day operation.

Chris Bogdan: OK. And now you know what I'm going to say next.

Answer: Yes sir.

Chris Bogdan: What am I going to say next?

Answer: You're going to say, "We're not going to pay for it."

Chris Bogdan: That's right. We're not going to pay for it.

Chris Bogdan: Long gone is the time where we will continue to pay for mistake after mistake after mistake." (Walsh & Martin, 2014)

No one would fault the PEO for focusing on the costs associated with the rework necessary to fix the valve issue. However, while cost is a consideration, the schedule implications of this rework issue were equally important. We learned that the rework associated with this effort added approximately six months to that development phase. The impact of the ripples caused by the rework, the other activities that needed to stop while this fix was installed, and the downstream delays throughout the development are unknown. In system dynamics, this phenomenon is called firefighting[3] (Novak & Levine, 2010).

From a schedule perspective, the central issue is whether the project management office (PMO) considered schedule impacts when this issue surfaced. The PMO was focused on cost, which is unsurprising as cost is the most critical consideration for DoD PMs. Here is where the action-reaction-counteraction heuristic can help to identify downstream problems.

[3] Firefighting is a term used in system dynamics to describe a situation where a problem is temporarily solved by applying a quick fix, but the underlying cause is not addressed, and the problem reappears or worsens over time. Firefighting often leads to a vicious cycle of increasing complexity, unintended consequences, and escalating costs. An example of firefighting is using overtime or hiring temporary workers to meet a surge in demand but neglecting to improve the production process or capacity. Firefighting can be avoided or reduced by applying a systems thinking mindset, which is essential for avoiding firefighting and achieving long-term goals.

The F-35 Weight Problem

The F-35's weight problem can be traced back to its inception, when the program aimed to develop a common platform for three variants: the F-35A for conventional takeoff and landing (CTOL), the F-35B for short takeoff and vertical landing (STOVL), and the F-35C for carrier-based operations. The common platform was intended to reduce development and production costs and increase interoperability among the partner countries. However, the common platform also imposed significant design constraints and trade-offs, especially for the F-35B, which required a complex and heavy lift system to enable vertical landing.

According to a report by the Government Accountability Office (GAO), the F-35 program faced a "mismatch between aircraft weight and available engine thrust" since its early stages, and the F-35B was the most affected variant (GAO et al., 2005). The report stated that the F-35B had a low weight margin of 2%, meaning that any increase in weight would hurt its performance and requirements. The report also noted that weight reduction efforts necessitated an 18-month delay (GAO et al., 2005).

The F-35 weight problem emerged in the early 2000s when the aircraft was still in the concept demonstration phase (Tirpak, 2006). The initial weight target for the F-35 was 30,000 pounds, but by 2003, the aircraft had grown to 35,000 pounds, and by 2004, to 36,000 pounds. The weight problem severely affected the F-35's performance, stealth, and safety. A heavier aircraft requires more fuel and thrust, reducing its range, maneuverability, and payload. It would also generate more heat and radar signatures, compromising its stealth and survivability. Moreover, a heavier aircraft would pose more significant risks of structural failure, especially for the STOVL and CV variants, which had to endure harsher landing conditions.

The Weight Solution

The Pentagon and Lockheed Martin recognized the F-35 weight problem as a significant threat to the program's success and viability. In 2004, the Pentagon initiated a comprehensive review of the F-35 program, known as the Joint Estimating Team (JET; Pappalardo, 2006). The JET concluded that the F-35 needed to shed at least 2,000 pounds to meet its performance and stealth goals. The JET recommended management and engineering changes to improve the program's oversight, accountability, and efficiency.

In response to the JET's findings, Lockheed Martin launched a weight reduction campaign known as the Weight Attack Team (WAT), which involved hundreds of engineers and managers from across the company and its subcontractors. The WAT adopted a rigorous and systematic approach to identify and eliminate unnecessary or excessive weight from every part and system of the F-35. The WAT used various tools and techniques, such as computer modeling, simulation, testing, and prototyping, to evaluate each design decision's weight, cost, and performance trade-offs. The WAT also implemented a culture of weight consciousness and discipline, encouraging innovation, collaboration, and communication among the F-35 team members (Pappalardo, 2006).

The WAT's efforts seem to have paid off, as the F-35 shed about 2,700 pounds by 2006, exceeding the JET's target. The weight reduction was achieved by making numerous changes to the F-35's design, such as using lighter materials, optimizing the shape and size of the components, simplifying the wiring and plumbing, and eliminating or consolidating some of the features and functions.

Notwithstanding the success of the WAT efforts, the F-35 program continues to address the weight problem through various measures, such as weight reduction initiatives, engine upgrades, and operational adjustments. The F-35's weight problem is a complex and persistent issue that has affected its performance, cost, and competitiveness. However, the weight problem remains a significant challenge and a source of controversy for the F-35 program. The schedule impacts of the weight issue are still being felt.

The SAWT-driven changes to the design and assembly of the aircraft are estimated at $4.8 billion. This is part of a $6.2 billion replanning to accommodate the additional design cycle required to make the improvements. The replanning caused an 18-month schedule delay in F-35 deliveries (Pappalardo, 2006). The work to reduce the weight of the F-35B also improved the designs for the other variants. However, much of the already completed engineering and manufacturing work had to be redone, causing years to be added to the schedule.

Case Study: Peace Shield

The Gulf War of 1990–1991 exposed the vulnerability of Saudi Arabia's air defense system, which was largely dependent on the U.S.-led coalition forces. The war highlighted the need for a robust and coordinated air defense system that could cover the vast and diverse terrain of Saudi Arabia. The Saudi government pursued a long-term strategic partnership with the United States to upgrade and modernize its air defense system. The United States agreed to provide Saudi Arabia with the latest technology, expertise, training, and maintenance support. The original Peace Shield Program (Peace Shield) was a large-scale air defense system project to design, develop, and test ground-air defense systems for the Saudi Air Force (Sciolino, 1985). The program aimed to modernize and integrate the existing Saudi air defense network with new radars, command and control centers, communication systems, and missile batteries. The overall program was initiated in 1984 and completed in 1996, costing more than $9 billion (Kausal IV, 1996).

Boeing was awarded the original Command, Control, Communications, and Intelligence (C3I) Peace Shield FMS contract in 1985. The C3I contract was expected to be completed in 1992, with the entire network planned to be fully operational by 1994. However, software integration problems caused significant delivery delays. Among other issues, Boeing significantly underestimated the cost and schedule of the project. In November 1989, the U.S. Air Force issued a cure notice to Boeing with a deadline of December 1989. Initially, the Air Force accepted Boeing's proposed solution. However, delays continued, with the completion estimate pushed into August 1994 (the original contract's completion date was April 1991). Boeing's original delivery schedule of 39 months had extended to 92 months. The Air Force terminated the Boeing contract for default in 1991 (Kausal IV, 1996).

In May of 1991, Pugh-Roberts/PA Consulting (PA) provided proposal support for Hughes to bid on the former Boeing part of the Peace Shield program. PA already had extensive experience in developing program management-based system dynamics models. Key personnel of the PA organization had written about the concept of rework in 1993 (Cooper, 1993a, 1993b, 1993c; Lyneis et al., 2001). Working closely with Hughes's managers, the PA team modified a model they had previously built to support what would turn out to be the winning bid. The model showed that with careful management, staffing, and execution, Hughes could deliver in 54 months. While the Air Force FMS Program Office was skeptical of the Hughes 54-month timeline based on their experience with Boeing, the Air Force awarded Hughes Aircraft a contract of $837 million in July 1991 (Kausal IV, 1996). The contract required Hughes to deliver the processing, displays, and communications equipment throughout Saudi Arabia and provide various equipment and software in 54 months (Forecast International, 2003).

During the execution of the contract, Pugh Roberts/PA Consulting advised Hughes during the program execution using the system dynamics model developed for the proposal. The model was based on the rework cycle and provided planning and execution support during the contract performance period (Lyneis et al., 2001). A central aspect of the model was the software development, which would be “lifted” (reused) from a previous program. The reuse significantly reduced schedule risk.

Given the program’s history to that point, the Air Force offered significant cost and schedule incentives. The contract had a cost and schedule incentive that offered a $50 million bonus for a three-month early delivery and a $50 million penalty for a late delivery (Kausal IV, 1996). Hughes decided to provide 40% of that $50 million contract incentive to incentivize employees (Kausal IV, 1996). The execution was flawless, and Peace Shield was delivered six months and 13 days ahead of schedule and below cost, a clear win for the Kingdom of Saudi Arabia, the U.S. Air Force, Hughes, and their employees.

The program was successful for three essential reasons. First, a Hughes unique “teaming” process provided more oversight and more frequent reviews. This oversight process was simulated in the system dynamics model. Although there was a short-term disruption in the work, the process yielded increases in quality, thus resulting in less rework (Lyneis et al., 2001). The second reason was staffing. Usually, staff doing those tasks would be reassigned as soon as tasks are finished. Instead, the typical “roll-off” of staff, especially software engineering staff, was slowed. This provided a personnel buffer to deal with rework issues that would otherwise add significant time to the schedule. The final reason was the overall approach to the project. From the beginning, Hughes worked with PA to plan and execute Pease Shield in a systems fashion, using system dynamics simulations to test various approaches. This system approach provided visibility from start to finish and allowed managers to anticipate problems before they became overwhelming.

This section examined the criticality of understanding project dynamics in schedule execution in weapon system development projects. The F-35 and Peace Shield case studies provided examples of the importance of realizing that a weapon system development project is a system and that there are feedback effects at work in our programs with rework and its impact on project cost, schedule, and performance of the system being predominant. The F-35 and the Peace Shield case studies also highlight the challenges of recognizing the complex and dynamic forces in our weapon system developments. Finally, a systems approach to project management provides valuable insights into the systemic causes and consequences of rework, enabling organizations to make more informed decisions and implement effective strategies for its prevention or mitigation.

Improving Information Quality: The “Wise Crowds” Proposition[4]

Executives know . . . valuable information is scattered across the organization. They just don’t know how to retrieve it (Thompson, 2012, 1).

Managing complex, imperfectly understood systems (like defense acquisition programs, entails information that encompasses a broad perspective. We believe a good

[4] Our paper proposal promised a “deeper dive” into prediction markets (subject, of course, to editorial constraints). Building upon our previous work in this area (summarized here), we consider issues relevant to using prediction markets as tools for practicing program managers. Hence, our extensive references to Thompson (2012).

understanding of System Dynamics coupled with a solid working knowledge of data science[5] is helpful for understanding that sort of information.

But that does not constitute a panacea. Information distilled into databases has at least two difficulties. First, those data sets are structured to answer (perhaps implicitly) specific questions. It may well prove difficult to use that data to answer new questions. Second, distilled databases are generally assembled at the cost of some helpful information.[6]

Distilled data can take information out of context and distort meaning (Nguyen, 2024). In many cases, there is no substitute for local knowledge. We assert that tools such as prediction markets have significant potential to aggregate information in ways that exploit local knowledge to enhance program management.

In this section, we continue our previous inquiries regarding prediction markets as a potential method to predict emerging problems in defense acquisition programs. If adverse developments can be expected and lead to actionable information[7], management would be better equipped to mitigate those effects (making program schedules more resilient).

Substantial experience supports the hypothesis that estimates from a group can be more accurate than, say, expert judgments. On the other hand, groups can be spectacularly wrong—e.g., financial bubbles, long-shot winners, and black swans.

A framing assumption for groups being collectively wise (potentially) is that every member has private information. Each person's information includes insights and errors (of various kinds).[8] In a "proper" group setting, the (private) errors across the group tend to cancel out in the collective opinion, while the private sets of information add to the quality of that opinion (Surowiecki, 2004, pp. 10, 41). For example, Hayek (1945/1971, pp. 17, 19–23) and Smith (1776/1937, pp. 13–16) discuss the ability of a crowd of market participants to reach a beneficial economic equilibrium.

Various lines of inquiry have identified **characteristics of "wise crowds."**

- Cognitive diversity is formed in good part by the heterogeneity of private information. Insufficient "cognitive diversity" can lead to "groupthink" and associated pressures to conform (Surowiecki, 2004, pp. 23, 38).
- Independence of members promotes a diversity of errors in the sets of private information, which are more likely to be canceled out.
- Decentralization (in an organizational sense) can foreclose the tendency of hierarchies to filter out information and judgments at lower levels.
- Aggregation produces a collective assessment related to the entire group, which can lead to an evaluation, forecast, or decision. Aggregating a group's collective "wisdom" implies a method other than hierarchical screening, such as prediction markets.

---

[5] As discussed for example, in Pickar and Franck (2021).

[6] E. B. White (1938) provides an interesting (albeit overstated) view of the pitfalls of distilled data.

[7] Joseph and Sconion (2020) offer an interesting perspective regarding this issue.

[8] One can view each set of private information as having two components: useful knowledge and errors, without individuals being aware of how their private information is divided between those components.

Prediction Markets

"The advantage of prediction markets is that they can benefit from the wisdom of crowds. By collecting and weighing the predictions of a large number of traders, they can provide a market-wide forecast that is generally more reliable and balanced than any single expert opinion" (Peters, 2022).

Even with a wise crowd, designing a well-functioning prediction market is easier said than done—even with the substantial body of research and experience available.

Self-Fulfilling or Self-Negating Group Predictions: For example, a group prediction of an untoward acquisition program event can lead to management actions to prevent that event. That is, the group's assessment could be self-negating. This accordingly complicates determining the winning bets.

Positive and Perverse Incentives: Prediction markets look like and can operate like betting markets. As such, there can be incentives to engineer a favorable outcome, which has happened in sports betting markets. This problem can be addressed by limiting the size of the wagers. On the other hand, motivated participation is essential, and this can be addressed by increasing the stakes. However, motivating traders through increased stakes might be substantial incentives to be correct, with the associated potential for perverse incentives (for the enterprise the prediction market is intended to enhance).

Definition of Outcomes: Typically, prediction markets focus on well-defined binary outcomes (like win or lose) that occur at a definite time. But what happens if the outcomes are more complicated?[9] Suppose a wise group identifies an emerging problem in an acquisition program (such as a schedule slip). Suppose that alerted program management also undertakes a remedy that averts the crisis.[10] How do the prediction market rules determine the winner?

One way around this problem is to have more complicated results. For example, group members could choose an outcome in perhaps two parts. "Asking 'If this, then what?' questions usually involve a prediction market with two complementary questions: 'How much will x be if y occurs?' and 'How much will x be if y does not occur?'" (Thompson, 2011, pp. 20–21).

Another obvious issue is that acquisition programs (especially MDAPs) are lengthy and have uncertain termination or milestone dates. Defense acquisition prediction markets operating arena will likely need special care in framing the questions upon which to place bets.

Potential Pitfalls in Structuring a Prediction Market

Organizational Culture: As Thompson (2012, p. 10) put it, "The key to success of the company's prediction process is their corporate culture." Grafting a prediction market into a hierarchy of risk-averse participants is likely to encounter substantial difficulties (Thompson, 2012, p. 180).

Posing Useful Questions: What questions apply to program management, and how should they be identified? This is one major issue related to organizational culture. Should the questions be posed by program leadership (who may be blissfully ignorant of emerging

---

[9] Almenberg, et al. (2009) note this to be a major issue for prediction markets involving questions of scientific research. It's reasonable to suppose that this is also an issue for technical endeavors such as major defense programs.

[10] Not so far-fetched. Miller (2012, pp. 48–49) offers a method which can yield actionable indication of an emerging problem in acquisition programs. It's therefore reasonable to suppose that prediction markets could also provide similar warnings.

problems) or "bubble up" from the workforce? If the latter, how are the best questions identified, and how is a consensus about "good" questions formed? Previous experience indicates "good" questions improve the performance of prediction markets (Almenberg et al., 2009). This seems to be a significant issue related to organizational culture and its receptiveness to bottom-up modes of operation.

Manipulating Prediction Market Outcomes: Large, strategically-placed bets can significantly affect the market outcome—somewhat like a short-sales campaign to influence stock market prices. The counterargument is that such actions motivate other prediction market participants to negate those effects (Thompson, 2012, pp. 22–23). As recent experience (Ramkumar, 2024) indicates, manipulating markets (in general) with broader objectives in mind can take on some fascinating forms.

Manipulating Events: This is arguably the most severe potential pitfall. If prediction market results are accepted as credible, and if the market attaches a high probability to an "If x, then y" bet, then an agent (market participant or not) who cared deeply about y could be (perversely) incentivized to take unusual steps to assure x happened.[11]

Legal Complications: These include the following (Thompson, 2012, p. 223–227):

- running afoul of a rather complicated structure of gambling prohibitions at various levels of government;
- potential susceptibility to outcome manipulation through illegal means, and;
- potential to facilitate insider trading (in other markets).

Information Issues and Market Functioning: If little reliable information is available to the market participants, there is not much useful information to aggregate. Accordingly, those participating are more likely to offer guesses than considered assessments (Thompson, 2012, p. 209).

Comments on Prediction Markets

There seems to be a sweet spot for prediction markets in which (a) the participants collectively (but not individually) possess sufficient information and (b) a question not readily solved with specialized (expert) skills.

However, transplanting the prediction market model is easier said than done. The DoD is a top-down organization with members who are generally risk-averse. Potential applications in a DoD context have not been widely explored. The issues and problems raised here are hypotheses warranting further inquiry, including practical experience and lessons learned.

Concluding Comments

As with our multi-year schedule research inquiries, this paper is about improving the state of the art for estimating and executing defense acquisition program schedules and making them more resilient to unforeseen events. This year's focus has been on the role of system dynamics and prediction markets in managing complex projects. The case studies provide examples of the criticality of understanding the system forces at work and that, while essential, cost and technical responses can fail to address the entire problem. There is potential for prediction markets to improve schedule information quality. We believe

---

[11] For example, A prediction market in 2004 gave George W. Bush a 91% chance of reelection if Osama bin Laden were caught, and 67% if not (Thompson, 2012, p. 53). The Bush Administration was, in effect, incentivized to assign an increased (and possibly ill-advised) priority to capturing him. Almenberg et al. (2009) also note the adverse possibilities of strategic behavior.

prediction markets, mechanisms that aggregate the opinions and information of many participants to produce forecasts or decisions, may provide valuable insight into the mechanics of the system that is a weapon program.

In conclusion, this paper furthers our research agenda on schedules and how to make them more resilient to unforeseen events (and what to do when resiliency isn't enough). Systems approaches, case studies, and consideration of prediction markets to improve information quality offer insights into managing complex projects. Understanding the art and science of planning and executing weapon system acquisition schedules is fundamental to delivering systems on time.

References

Abdel-Hamid, T. K., & Madnick, S. E. (1982). *A model of software project management dynamics*. https://dspace.mit.edu/bitstream/handle/1721.1/48555/modelofsoftwarep00abde.pdf?sequence=1

Abdel-Hamid, T. K., & Madnick, S. E. (1983). The dynamics of software project scheduling. *Communications of the ACM*, *26*(5), 340–346. https://doi.org/10.1145/69586.358135

Almenberg, J., Kittlitz , K., & Pfeiffer, T. (2009). An experiment on prediction markets in science. *PLOS One*. https://doi.org/10.1371/journal.pone.0008500

Cooper, K. G. (1993a). The rework cycle: How it really works . . . and reworks. *PM Network*, 25–28. http://scholar.google.com/scholar?q=related:VM52O77UK8sJ:scholar.google.com/&hl=en&num=20&as_sdt=0,5

Cooper, K. G. (1993b). The rework cycle: Why projects are mismanaged. *PM Network*, 5–7.

Cooper, K. G. (1993c). The rework cycle: Benchmarks for the project manager. *Project Management Journal, XXIV*(1), 17–22. http://www.mtas.ru/search/search_results.php? publication_id=449

Dörner, D. (1996). *The Logic Of Failure*. Basic Books. http://books.google.com/books?id=UrjHjgEACAAJ&dq=intitle:The+Logic+Of+Failure&hl=&cd=1&source=gbs_api

Flyvbjerg, B., & Budzier, A. (2015). Why do projects fail? *Project (Association for Project Management), Summer 2015*(283), 22.

Forecast International. (2003). *Peace shield—Archived report.* https://www.forecastinternational.com/archive/disp_old_pdf.cfm?ARC_ID=847

Franck, R., Hildebrandt, G., Pickar, C., & Udis, B. (2017). *Realistic acquisition schedule estimates: A follow-on inquiry.* https://calhoun.nps.edu/bitstream/handle/10945/58970/SYM-AM-17-091-PO_Franck.pdf?sequence=1

GAO, Sullivan, M. J., Hazard, M., Bonner, M., Drerup, M., Lea, M., Schilling, D., Sloan, K., Vodraska, A., & Office, U. S. G. A. (2005). *Tactical aircraft: Opportunity to reduce risks in the joint strike fighter program with different acquisition strategy*. https://apps.dtic.mil/sti/citations/trecms/AD1167885

Hayek, F. A. (1971). American economic review, vol 35. In H. Townsend (Ed.), Price theory: Penguin modern economics readings (pp. 17–31). Penguin. (Original work published in 1945)

Joseph, B. B., & Sconion, D. (2020). Using natural language processing, sentiment analysis, and text mining to determine if text in selected acquisition report executive summaries are highly correlated with Major Defense Acquisition Program (MDAP) unit cost can be used as a variable to predict future MDAP costs. *Proceedings of 17th Annual Acquisition Research Symposium*.

Kausal IV, B. A. T. (1996). Peace shield—A study in motivation. *Program Manager*, 22–24.

Love, P. E. D., Holt, G. D., Shen, L. Y., Li, H., & Irani, Z. (2002). Using systems dynamics to better understand change and rework in construction project management systems. *International Journal of Project Management, 20*(6), 425–436. https://doi.org/10.1016/s0263-7863(01)00039-4.

Lyneis, J. M., Cooper, K. G., & Els, S. A. (2001). Strategic management of complex projects: A case study using system dynamics. *System Dynamics Review, 17*(3), 237–260. https://doi.org/10.1002/sdr.213

Miller, T. P. (2012). *Acquisition program problem detection using text mining methods.* AFIT (Air Force Institute of Technology) Scholar, AFIT/GCA/ENC/12-0.

Nguyen, C. T. (2024). The limits of data. *Issues in Science and Technology, XI*, 94–101. https://issues.org/limits-of-data-nguyen/

Novak, W. E., & Levine, L. (2010). *Success in acquisition: Using archetypes to beat the odds.* Software Engineering Institute CMU. http://repository.cmu.edu/sei/271/

Pappalardo, J. (2006). Weight watchers: With a high-profile military contract like the Joint Strike Fighter, the toughest place to go is back to the drawing board. *Air and Space Smithsonian, 21*(4), 66.

Peters, K. (2022). Prediction market. *Investopedia*. https://www.investopedia.com/terms/p/prediction-market.asp

Pickar, C. K. (2018). *Informing DoD program planning through the examination of the causes of delays in acquisition using acquisition data*. https://calhoun.nps.edu/bitstream/handle/10945/58762/SYM-AM-18-079-019_Pickar.pdf?sequence=1

Pickar, C., Franck, R., Hildebrandt, G., & Udis, B. (2019). *In the fullness of time: Towards realistic acquisition schedule estimates*. https://dair.nps.edu/handle/123456789/2778

Pickar, C., & Franck, R. (2021). *Telling time: Getting relevant data for acquisition schedule estimating relationships, acquisition research program.* Naval Postgraduate School. https://dair.nps.edu/bitstream/123456789/4532/1/NPS-AM-22-013.pdf

Rahmandad, H., & Hu, K. (2010). Modeling the rework cycle: Capturing multiple defects per task. *System Dynamics Review, 26*(4), 291–315. https://doi.org/10.1002/sdr.435

Ramkumar, A. (2024, March 20). Trump is in line for a $3.5 billion windfall from stake in truth social. *Wall Street Journal*. https://www.wsj.com/finance/stocks/trump-is-in-line-for-a-3-5-billion-windfall-from-his-stake-in-truth-social-eae58c9d

Richardson, G. P., & Pugh, A. L. (1981). *Introduction to system dynamics modeling with DYNAMO*. MIT Press.

Sciolino, E. (1985). Saudis to install $4 billion air defense system. *New York Times*, 1.

Smith, A. (1937). *An inquiry into the nature and causes of the wealth of nations* (Ed. E. Canaan). Modern Library. (Original work published 1776)

Surowiecki, James, The Wisdom of Crowds, Doubleday, 2004.

Thompson, D. N. (2012), Oracles: How prediction markets turn employees into visionaries. *Harvard Business Review Press*.

Tirpak, J. A. (2006). Struggling for altitude. *Air Force Magazine*, *89*(9), 38.

Walsh, M., & Martin, D. (Producers). (2014, February). *60 Minutes: Is the F-35 worth it?* [Video]. https://www.cbsnews.com/news/f-35-joint-strike-fighter-60-minutes/

White, E. B. (1938). *IRTNOG: An extremely short work of dystopian fiction*. https://acephalous.typepad.com/acephalous/2007/05/irtnog_by_eb_wh.html

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net