SYM-AM-26-153

Excerpt from the Proceedings
of the
Twenty-Third Annual
Acquisition Research Symposium and
Innovation Summit

Volume III
"Accelerating Warfighting Capabilities"

Cognitive-Technical Scheduling:
A Framework for Addressing DoD Weapon System Acquisition Failures

Published: April 30, 2026

The research presented in this report was supported by the Acquisition Research Program, Graduate School of Defense Management at the Naval Postgraduate School.

To request defense acquisition research, please contact:

Acquisition Research Program
Department of Defense Management
Naval Postgraduate School
E: arp@nps.edu
www.acquisitionresearch.net

Cognitive-Technical Scheduling: A Framework for Addressing DoD Weapon System Acquisition Failures

Charles Pickar—has retired from the faculty of the Naval Postgraduate School (NPS), 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. He has also served as Chair of the IEEE Systems Council's Systems Education Technical Committee. 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 as a 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

Defense acquisition programs continue to face schedule overruns due to cognitive biases, overly optimistic planning assumptions, and limited visibility into evolving program behavior. This paper proposes a cognitive-technical scheduling system that combines three key capabilities: Defense Acquisition Visibility Environment's (DAVE's) and other DoD databases' empirical performance data, AI-enabled forecasting, and system dynamics modeling. DAVE offers the historical data needed to address base-rate neglect and anchor estimates in proven outcomes. AI systems trained on this data could produce probabilistic forecasts, identify emerging deviations, and explore alternative execution paths, thereby offsetting human limitations in perception and memory. System dynamics models illustrate feedback loops, delays, and accumulations that influence schedule behavior, aligning mental models across the PMO and IPTs and enabling the development of schedule-level digital twins. Collectively, these components create a continuously updated simulation environment that forecasts results, visualizes secondary and tertiary effects, and supports proactive interventions. The paper presents a conceptual model where Cognitive Technical Integration enhances schedule performance both directly and through improved situational awareness, bias reduction, and adaptive learning. This framework provides a way to achieve more realistic scheduling, stronger governance, and greater strategic flexibility across the defense acquisition system.

Introduction

Complex weapons systems development continues to face persistent failures, especially in defense technology innovation and acquisition (GAO, 2023, 2025; Oakley, 2025). While technical risk and requirements instability are traditional explanations, research suggests these outcomes reflect deeper cognitive and behavioral constraints in complex sociotechnical environments (Mortlock, 2021; Stingl & Geraldi, 2021; Stingl et al., 2025). Decision-makers operating in high-complexity, high-uncertainty environments frequently exceed their capacity to process information and anticipate outcomes, producing systematic biases that undermine the accuracy of project estimates (Flyvbjerg, 2021; Lovallo & Kahneman, 2003; Tversky & Kahneman, 1974). Foundational cognitive science literature on complex problem-solving shows that humans struggle to manage nonlinear, dynamic systems with delayed feedback, uncertainty, and interdependent variables (Bui, 1996; Dörner, 1980, 1996; Dörner & Guss, 2022). This paper argues that recurring performance challenges in DoD acquisition stem from fundamental cognitive limitations rather than solely from procedural or technical shortcomings (Arena et al., 2006; Buehler et al., 1994; Flyvbjerg & Gardner, 2023).

Literature Review

The integration of cognitive science, system dynamics, artificial intelligence (AI), and digital twin technologies offers a novel avenue for addressing these limitations by embedding human problem-solving insights into adaptive, data-driven project scheduling systems. The following literature review examines key theoretical and empirical works that inform this integrated framework. The synthesis draws on sources ranging from foundational cognitive science to contemporary behavioral project management research, where inferences extend beyond what individual sources explicitly claim; this is noted.

Cognitive Foundations of Project Scheduling

Baptiste & Grunder's (1999) cognitive activity network model represents an early approach to developing a project scheduling formulation that accounts for both the cognitive and operational characteristics of project planning activities. It was developed primarily from an Operations Research/Constraint Programming perspective, but it provides a basis for understanding how scheduling models can address real-world operational complexities rather than purely idealized mathematical representations. This provides a methodological link to the current framework. Models that cannot represent real-world operational complexity will be unable to support the situational decision-making used by human project managers and their staff.

Stingl et al. (2025) extend this perspective through a conceptual essay that frames project cognition, understood as the way the mind acquires, processes, and enacts information in project contexts, as the underlying basis for all project behavior. Their essay calls for a stronger engagement with cognitive science in project research, arguing that it is not differences in thinking itself but differences in the context in which thinking is applied that make projects a distinct and relevant arena for cognitive study. The authors emphasize that cognitive framing, including how project managers perceive uncertainty, interdependencies, and time pressure, is foundational to understanding project execution behavior, and they call for integrating cognitive models into project management research and tools to improve situational awareness and decision quality. This aligns with challenges in DoD system development, where cognitive overload from complex program interactions frequently contributes to schedule slippage.

Stingl et al. (2025) also warn that viewing bias correction as merely fixing biased thinking to predict what people will think about their projects is dangerous. Referencing Stingl and Geraldi (2021), Stingl et al. (2025) argue that heuristics can be suitable for their specific context and that Kahneman and Tversky's (1974) distinction between good and bad decisions may not directly apply to project environments.

The tension between the Kahneman–Tversky tradition and the Gigerenzer (2024) informed position represented by Stingl et al. (2025), referencing Stingl and Geraldi (2021), reflects an unresolved debate in behavioral project management research. Kahneman and Tversky (1974) treat heuristics primarily as sources of systematic error, while Stingl and Geraldi argue that heuristics can be well adapted to specific project environments and that applying a universal standard for good and bad decisions may not be appropriate in project contexts. The proposed Cognitive-Technical Scheduling System (CTSS) does not resolve this debate. Rather than functioning as a bias-correction system that presupposes heuristic reasoning is deficient, the CTSS is designed as a decision-support system that expands the information environment in which judgment occurs. It surfaces feedback-loop dynamics, historical reference-class data, and probabilistic risk distributions that heuristic reasoning was not designed to process efficiently. When the system detects patterns consistent with known systematic biases, such as optimism bias or the planning fallacy[1] (Kahneman & Tversky, 1982), it will flag those patterns for human review. The project manager (PM) retains decision authority. The system does not substitute a corrected estimate but presents evidence against which existing judgments can be evaluated. This approach is compatible with both traditions as it does not assume that heuristic reasoning is categorically flawed. Instead, it provides the project management staff with a signal when the evidence suggests systematic error is likely.

Human Limitations and Complexity in Project Settings

Dörner's seminal works ("On the Difficulties People Have in Dealing with Complexity" (1980) and *The Logic of Failure* (1996)) are essential for understanding why human decision-makers struggle with complex project systems. The 1980 article, published in *Simulation and Games*, presents experimental evidence that individuals systematically mismanage feedback-rich environments due to limited working memory, misinterpretation of delays, and overreliance on linear causal reasoning. In *The Logic of Failure*, Dörner extends this framework, explaining how even experienced professionals tend to oversimplify dynamic systems, leading to compounding errors over time.

Dörner and Güss (2022) build on this framework by developing a taxonomy of 24 human errors in complex problem-solving and dynamic decision-making, drawing on microworld simulation studies. When applied to project scheduling, these insights imply that human planners tend to underestimate nonlinear interactions among technical subsystems, resulting in systematically optimistic schedules, an observation corroborated by empirical project data from defense programs.

Flyvbjerg (2021) consolidates related findings through a behavioral science perspective, highlighting the ten most important behavioral biases in project management. Among the most impactful are strategic misrepresentation, optimism bias, and the planning fallacy, all of which skew schedule baselines and performance tracking. Importantly, Flyvbjerg emphasizes that equating behavioral bias solely with cognitive bias is a mistake. Political bias, driven by deliberate strategic distortions stemming from power and organizational incentives rather than cognition, is equally important, especially in large defense projects. His analysis underscores the need to address these biases through data-driven feedback loops, adaptive modeling, and digital simulation, while also considering the organizational conditions that foster strategic misrepresentation.

System Dynamics, AI, and Digital Twin Approaches

To overcome the cognitive limitations identified by Dörner (1980, 1996) and the behavioral distortions documented by Flyvbjerg (2021), system dynamics (SD) offers a formal method for representing and simulating complex feedback structures in weapon system development. By capturing interdependencies among cost, schedule, risk, and resource variables, SD models provide planners with a simulation environment to explore what-if scenarios and visualize the downstream consequences of early decisions. When augmented with AI, these models can detect emergent bottlenecks, predict likely schedule deviations, and recommend mitigation strategies based on historical data. Digital twin technologies extend this paradigm by providing real-time synchronization between the virtual model and the physical project environment. In this hybrid environment, the project management office (PMO) serves as the operational-level sense-maker, while AI-driven models manage tactical complexity. The

[1] The planning fallacy refers to the tendency for people to underestimate how long a task will take, how much it will cost, and what risks it will entail, while simultaneously overestimating the likelihood that the project will succeed.

system dynamics framework provides transparency and interpretability, critical for DoD acquisition accountability, while digital twins enable continuous validation of assumptions against live data from program execution. The organizational and political conditions that enable strategic misrepresentation, as Flyvbjerg emphasizes, must also be addressed through governance structures that embed formal bias checks and empirical benchmarks into milestone decisions; technical tools alone are insufficient.

Summary

The literature reviewed supports the view that cognitive and behavioral factors are significant, systematically under-addressed contributors to schedule failures, operating in conjunction with political and institutional pressures that also warrant attention. The cognitive science literature (Dörner, 1980, 1996; Stingl et al., 2025) reveals the inherent human limitations in managing complexity, while the behavioral research (Flyvbjerg, 2021) shows both the cognitive and political dimensions of systematic bias in project environments. Technical methodologies, e.g., system dynamics, AI, and digital twins, provide complementary tools for externalizing, simulating, and partially addressing these cognitive shortcomings. The organizational and political conditions that enable strategic misrepresentation, however, warrant equal attention alongside technical solutions, as no scheduling tool can fully compensate for institutional incentive structures that reward optimistic projections.

Applying Cognitive-Technical Insights to DoD Weapon System Acquisition

Major DoD weapon system acquisition programs operate at the intersection of technical complexity, organizational interdependence, and strategic urgency. Programs encompassing advanced aircraft, naval vessels, hypersonic systems, and integrated command-and-control architectures involve thousands of requirements, tightly coupled subsystems, evolving operational demands, and interdependencies among the PMO, contractors, laboratories, and test ranges. Under these conditions, project planners are cognitively overloaded by the scale and dynamism of what they must manage (Dörner, 1980, 1996; Dörner & Güss, 2022). Acquisition personnel tend to rely on simplified mental models that underrepresent the dynamic interactions among engineering risk, staffing, rework cycles, and integration delays. Traditional project management methods compound this problem by assuming linearity, predictability, and static baselines. In complex adaptive systems such as modern weapons programs, those assumptions quickly collapse under feedback delays, uncertain technologies, and shifting requirements.

These cognitive limitations are compounded by institutional incentives. Early program baselines are typically developed under tight time and budgetary constraints to establish political feasibility, creating conditions favorable to optimism bias, anchoring, and strategic misrepresentation. Overconfidence and uniqueness bias are reinforced by narratives that frame each program as unprecedented, despite strong empirical continuity with prior programs. Flyvbjerg (2021) identifies this pattern as a recurrent driver of cost and schedule failure across major infrastructure and defense programs.

Translating Cognitive Science into Program Practice

The work of Baptiste and Grunder (1999) and Stingl et al. (2025) demonstrates that project managers' cognitive framing of time, uncertainty, and causality directly determines schedule outcomes. In the DoD context, these point toward three integrated changes. First, cognitive modeling should be integrated into program management training and tools, enabling planners to reason about feedback loops and dynamic dependencies rather than static task lists. Second, decision support systems should externalize human cognition by visualizing cascading effects, risk propagation, and dynamic schedule resilience. Third, digital rehearsal environments, instantiated as digital twins, should allow teams to simulate alternative schedules under uncertainty, developing anticipatory rather than reactive management.

These changes would directly counter the planning fallacy and illusion of control identified by Flyvbjerg (2021). They also support a shift from human intuition alone toward human–AI cognitive partnership, in which humans provide strategic intent, context, and ethical oversight while AI manages tactical complexity through probabilistic forecasting, emergent risk identification, and real-time anomaly detection. Together, they constitute a cognitive-technical system capable of outperforming either human or machine alone.

System Dynamics, Digital Twins, and Adaptive Scheduling

The CTSS integrates these components across four interdependent functions. System dynamics models simulate the long-term consequences of resource reallocation, technical risk, and design changes before implementation, making feedback effects visible during planning and decision reviews. Digital twins mirror real-time program execution, enabling live calibration of system dynamics models against actual program state. AI algorithms continuously learn from both, applying reference class[2] forecasting at scale, predicting slippages, surfacing leading indicators, and suggesting preemptive corrective actions. Cognitive elicitation techniques surface and align the mental models of program stakeholders, while governance processes embed formal bias checks and empirical benchmarks into milestone decisions.

This closed-loop architecture transforms static Integrated Master Schedules into adaptive planning instruments, responsive to both internal and external change. The practical results include faster detection of emerging risks, real-time schedule adaptation without full re-baselining, and improved synchronization across contractors and subsystems, offering a structured pathway toward more reliable schedule outcomes in complex weapon system acquisition.

Strategic Advantage through Intelligent Speed

The competitive imperative to field capable systems faster demands not reckless acceleration but intelligent speed. The literature is consistent that speed without cognitive support increases the likelihood of systemic failure (Dörner, 1996; Flyvbjerg, 2021). A cognitive-technical approach generates sustainable acceleration through foresight rather than hindsight, learning rather than guessing, and adaptability rather than rigidity. By coupling cognitive science with system dynamics and digital twin architectures, DoD programs can move from rigid waterfall-style development toward adaptive concurrency, achieving both speed and control in weapon system acquisition.

Enhancing DoD Scheduling with DAVE, AI, and System Dynamics

The Department of Defense now possesses, for the first time, the technical infrastructure needed to directly counter the cognitive and behavioral weaknesses identified by Dörner, Flyvbjerg, Baptiste, Stingl, and their colleagues. Three capabilities stand out as especially powerful when combined: the Defense Acquisition Visibility Environment (DAVE)[3] family of acquisition databases (plus separate DoD databases administered by defense agencies), artificial intelligence and machine learning tools, and system dynamics modeling. Individually, each contributes meaningful value. Together, they form a cognitive-technical scheduling

---

[2] A reference class is a statistically meaningful set of comparable events, projects, or entities used to estimate outcomes such as cost, duration, risk, or performance.

[3] For purposes of this paper, we use DAVE to include other DoD acquisition databases beyond the central nature of DAVE and controlled by other DoD entities.

environment that improves both planning and execution across complex weapon system programs.

The DAVE databases aggregate decades of historical acquisition performance across cost, schedule, testing, program structure, contract types, and risk outcomes. This foundation supports decision-making in several important ways.

From Empirical Grounding to Cognitive-Technical Integration

One of the persistent pathologies Flyvbjerg identifies in large public-sector programs is base-rate neglect, the tendency for teams to discount the performance of comparable efforts. Within defense acquisition, enterprise datasets such as DAVE could directly counter this failure by enabling programs to construct rigorous reference classes for key drivers of schedule and technical risk: aircraft software-integration timelines, propulsion-subsystem maturation curves, flight-test durations by complexity class, integration-and-test manpower burn patterns, and subsystem rework cycles. When programs anchor their estimates in the empirical distributions derived from this historical data, forecast ranges become disciplined by demonstrated performance rather than optimistic planning assumptions, making it substantially harder for programs to assert that they are exceptions to established acquisition patterns.

Beyond anchoring forecasts, DAVE also enables analysts to examine how specific drivers such as technical maturity at Milestone B, concurrency levels, supplier fragility, and staffing instability have contributed to schedule slippage in prior programs. Quantifying these relationships allows the government to detect emerging risk patterns early in the acquisition life cycle and intervene before they manifest as schedule failures. This strengthens program-governance mechanisms, supports more defensible milestone decisions, and enhances the rigor of independent technical, cost, and schedule assessments.

The same historical data that enable reference-class construction and causal analysis also provide the foundation for AI-enabled forecasting and early-warning systems. Because DAVE aggregates decades of acquisition program performance, it offers a high-quality dataset or training corpus[4] for machine-learning models that can predict schedule deviations, NRE overruns, integration delays, and the onset of rework spirals. In effect, DAVE supplies the institutional memory that no individual program office can possess, allowing AI systems to learn patterns of slippage and recovery across hundreds of prior efforts.

Instead of providing a single point estimate that may lead to excessive confidence in an estimated completion date, AI models based upon this data can create probabilistic schedules (e.g., P50, P70, and P90 completion dates, task-specific error bounds, and Monte Carlo simulation-derived variance estimates) rather than point estimates. Additionally, they can identify early warning signs when a program improves situational cognition, which enhances awareness of cognitive bias; bias reduction, in turn, enables clearer perception of dynamic program structure. Both processes strengthen the adaptive learning enabled by AI and digital twin systems, increasing organizational schedule maturity over time and shifting DoD scheduling from static prediction to continuous adaptation of the remaining project schedule. These are examples of “weak signals” in a complex environment, a primary area of study found in Dörner’s work. In addition to predicting program outcomes, AI-based tools enable project managers to rapidly simulate scenarios (i.e., alternative execution paths), including, but not limited to, varying the number of staff assigned, adjusting requirements, disrupting supply chains, and changing how components integrate. Therefore, these tools allow project managers to make decisions quickly and confidently without relying solely on their experience and

[4] Machine learning algorithms depend on a dataset from which they can learn underlying patterns

intuition. Additionally, AI tools can automatically find relevant reference cases within DAVE and generate predictive curves for subsystem maturation, variability in performance among contractors, and learning rates related to throughput during testing, thereby embedding base-rate reasoning into the overall decision-making process of a program.

A systems approach is required to understand the structural nature of delay propagation in a program. While statistical and machine learning methods provide insight into schedule behavior, they do not capture the structural aspects of a program. System Dynamics (SD) provides a structural view by capturing the non-linear interactions, feedback loops, lags, and accumulation that commonly drive program outcomes. The structural relationships that exist within a program, for example, the acceleration of re-work due to compressing a schedule, the degradation of productivity, and the increase in defect injection rates resulting from volatile staffing levels, or the impact of concurrency decisions on propagating test failure and redesign, are precisely the same structural relationships that Dörner has shown humans have consistently underestimated. SD models allow program teams to see the causal structure behind their schedule performance—rather than rely on intuition.

SD also provides the behavioral engine for schedule-level digital twins, enabling exploration of alternative acquisition strategies, integration sequences, early staffing plans, and technology maturity trajectories. This gives program managers foresight into second- and third-order effects that are otherwise invisible, complementing the predictive capabilities of AI systems trained on DAVE's historical data. Because flawed mental models are a major driver of project mismanagement, as demonstrated by Stingl and colleagues, SD offers a corrective by providing a shared causal map of the forces acting on the program. This reduces conflict among IPTs (Integrated Product Teams), surfaces hidden assumptions, and creates a common analytic foundation for decision-making. During execution, SD models can be updated with new data to refine forecasts, simulate mitigation strategies, evaluate the consequences of test delays, and anticipate rework cascades before they materialize—directly addressing the monitoring and adaptation failures identified by Dörner and Güss (2022).

Taken together, DAVE, AI-enabled analytics, and system dynamics modeling form an integrated cognitive-technical scheduling system. DAVE provides the empirical foundation that counters base-rate neglect and anchors forecasts in demonstrated performance. AI interprets patterns, identifies emerging risks, and generates probabilistic forecasts, compensating for human limits in perception and working memory. SD models explain why schedule outcomes unfold as they do, aligning mental models across IPTs and improving decision quality throughout the life cycle. Combined, these elements create a continuously updated schedule digital twin that predicts outcomes, visualizes impacts, recommends mitigation strategies, alerts teams to danger patterns, and grounds decisions in empirical evidence. Figure 1 shows the CTSS Framework.

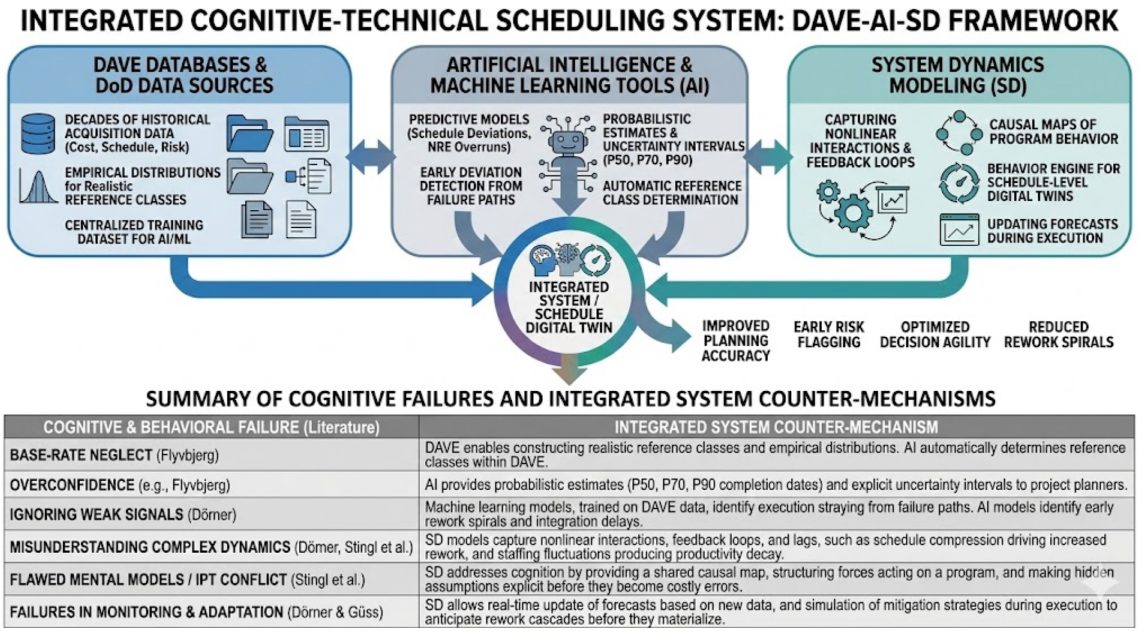

Figure 1 CTSS Framework

Integrating Data, AI, and System Dynamics into a Cognitive-Technical Scheduling System.

The cognitive and behavioral failures identified in the preceding literature review, along with the corresponding DAVE-AI-SD mechanisms, are summarized in Table 1.

Table 1. Cognitive-Behavioral Failure and DAVE-AI-SD Mechanisms

Cognitive / Behavioral Failure	How DAVE / AI / System Dynamics Address It
Oversimplified mental models	System dynamics reveals hidden feedback and delays
Optimism bias	AI generates probabilistic, empirically grounded forecasts
Anchoring	DAVE reference classes counter initial subjective estimates
Overconfidence	Monte Carlo simulations widen confidence intervals
Base-rate neglect	DAVE provides historical performance distributions
Reactive control and poor adaptation	Digital twin supports proactive replanning
Escalation of commitment	AI anomaly detection signals early deviation
Incomplete information search	AI risk intelligence scans vast datasets
Political bias	DAVE and other DoD databases provide a representative sample of similar past projects and use their actual outcomes to predict the current project's timeline. This removes the ability for political stakeholders to "wish away" delays. For example, if the last 10 weapons systems of this size took 6 years, one cannot claim a similar program will take 3 without extraordinary evidence.
Failure to detect side effects	SD models reveal second- and third-order consequences

Conceptual Framework Description

The independent variable is Cognitive-Technical Integration (CTI): the degree to which cognitive models, AI systems, and digital twins are combined to support project scheduling and decision-making.

Three mediating variables link CTI to outcomes. Situational Cognition (SC) refers to the capacity of project teams to perceive interdependencies, feedback, and time delays accurately (Baptiste & Grunder, 1999; Stingl et al., 2025). Bias Mitigation (BM) refers to the extent to which decision-making processes correct or reduce behavioral biases, such as optimism bias, the planning fallacy, and the illusion of control (Flyvbjerg & Gardner, 2023). Adaptive Learning (AL) refers to the extent to which a system continuously improves scheduling accuracy through feedback from digital twins and AI models (Dörner, 1996).

The dependent variable is Schedule Performance (SP): the accuracy, stability, and adaptability of schedule development and execution, measured by deviation from planned milestones, rework rate, and response agility to change.

Control variables include program size and complexity, system criticality, acquisition pathway (e.g., Major Capability Acquisition versus Software Acquisition), and contractor maturity level.

Proposed Conceptual Model

In this model, CTI exerts a direct positive effect on Schedule Performance (SP) and also improves SP indirectly through three mediating pathways: Situational Cognition (SC), Bias Mitigation (BM), and Adaptive Learning (AL).

Improved situational cognition enhances awareness of cognitive bias. Bias reduction, in turn, enables a clearer perception of dynamic program structure. Both processes strengthen the adaptive learning enabled by AI and digital twin systems, increasing organizational schedule maturity over time and shifting DoD scheduling from static prediction to continuous adaptation. Figure 2 is the model.

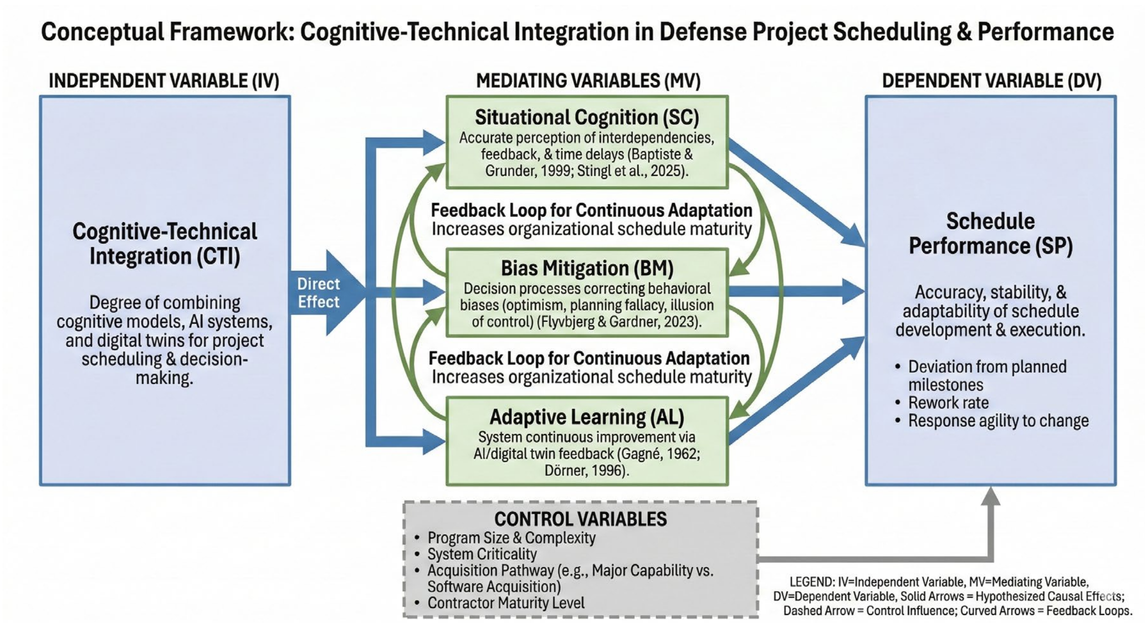

Figure 2. Conceptual Model

Such a design aligns with defense research priorities on human–machine teaming, AI assurance, and digital transformation, and provides actionable insight into where and how CTI interventions yield the greatest schedule gains. Figure 2 shows the proposed model.

This model bridges cognitive science and project management by conceptualizing schedule performance as a cognitive–technical phenomenon rather than a purely mathematical/procedural one. It extends Dörner's (1996) *Logic of Failure* and Flyvbjerg's (2021) behavioral bias theory into a new domain, dynamic, AI-enabled defense acquisition.

For practitioners and policy-makers, the model offers a framework for creating learning-focused scheduling systems that adjust in real time to changing constraints. Using CTI can lead to measurable improvements in schedule realism, program foresight, and strategic agility—key factors in gaining a competitive edge in modern warfare.

Limitations, Human Capital Risks, and the Incompleteness Problem

The case for CTI rests on what the CTSS can do. Intellectual honesty requires equal attention to what it cannot. Three interconnected limitations constrain the performance of a data-driven, AI-enhanced, digital twin system using system dynamics in complex acquisition environments. And each has implications for how the system must be operated, governed, and staffed.

The first limitation is ontological. A digital twin contains known components in known relationships, but program managers should assume from the outset that their model is incomplete. As Box and Draper (1987) observe, all models are wrong; the practical question is whether the inaccuracies are consequential.[5] In well-characterized technical domains, a digital twin can be an excellent likeness of the physical system it represents. In complex weapon system development, where subsystem interactions are imperfectly understood, and requirements continue to evolve, the twin will reliably miss some relationships, and those omissions are more likely to involve interaction effects than individual components. A model is useful to the extent its incompleteness does not matter. In high-complexity programs, that condition cannot be assumed.

The second limitation compounds the first. Complexity, defined not by the number of components but by incomplete knowledge of how those components interact, degrades model accuracy precisely in the environments where schedule support is most needed (Dörner, 1996). AI systems trained on historical data perform well within the scope of that training. Outside it, prediction quality deteriorates and can do so without providing clear warning signals to the operator. This is not a reason to reject AI-enabled digital twins; it is a reason to treat their outputs with calibrated rather than uncritical confidence, and to design governance structures accordingly. The disclaimer most AI systems in use today include that AI can make mistakes and hallucinate; all outputs must be double-checked, which is critically relevant.

The third limitation is the most consequential for acquisition workforce policy. Dependence on AI-powered tools can erode the human judgment required when those tools fail. Agarwal et al. (2018) identify this as a structural problem with automation: as machine-based control becomes routine, human operators accumulate less hands-on experience with the underlying system, reducing their capacity to recognize and correct bad machine outputs. The

[5] Box and Draper's formulation is commonly rendered as "all models are wrong, but some are useful." The fuller context from *Empirical Model-Building and Response Surfaces* (1987) reads: "Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful" (p. 424). The point is that model utility depends on whether the inaccuracies are consequential for the decision at hand, not on whether the model is a perfect representation.

crash of Air France Flight 447 in June 2009 clearly illustrates the mechanism. When automated airspeed indicators failed over the South Atlantic, the flight crew was unable to correctly diagnose the resulting stall, in part because the conditions fell outside the normal operating envelope in which their skills had been exercised (BEA, 2012). The aircraft was lost with all aboard. The parallel for acquisition program management is direct: a PMO that has delegated situational awareness to a digital twin will be poorly positioned to intervene when that twin mis-specifies a risk cascadde or fails to detect an emerging schedule deviation.

Defense acquisition history offers its own cautionary cases. The F-35 Joint Strike Fighter program encountered persistent weight growth across all variants that was not surfaced to senior leadership until it had already foreclosed design options. The eventual weight-reduction effort contributed to airframe life limitations in the B model, and the program's unit cost roughly tripled from the original estimates due to compounding design and concurrency decisions (Pappalardo, 2006). The Constellation-class frigate program provides a more recent example: weight problems identified during construction contributed to the program's termination after only two hulls, following years of schedule slippage and cost growth that accumulated without triggering adequate corrective action (Kington, 2025; O'Rourke, 2025). In both cases, the relevant signals were present in the program data. What was missing was the analytic infrastructure to surface them early and the trained judgment to act on them before options closed. A well-implemented CTSS would have provided the former. Developing the latter is the human capital challenge this section addresses.

The CTSS framework must therefore address human capital development as a first-order design requirement rather than an afterthought. Two phases are most critical. Before program launch, digital twin models can support rigorous cost-schedule-performance trade-off analysis at the system definition stage, the period when Dörner's evidence suggests errors are both most likely and most consequential. Exposing program managers to the structural consequences of early design decisions in a simulation environment develops the cognitive schema needed to interpret model outputs with appropriate skepticism. During execution, the same tools can be used to generate case-study materials, reconstruct management decisions amid adverse program developments, and support after-action learning, thereby building institutional memory in a domain where it has historically been thin (Dörner, 1996).

The goal is not to minimize reliance on AI-enabled tools but to develop the human judgment required to work with them effectively, to recognize the boundaries of their competence, and to exercise independent assessment when the evidence warrants it. A cognitive-technical system that displaces human expertise rather than augmenting it has reproduced the failure mode it was designed to correct.

Conclusion

Defense acquisition programs operate in environments defined by complexity, uncertainty, and persistent cognitive traps. Decades of evidence from Dörner's work on dynamic decision-making and Flyvbjerg's analyses of megaprojects show that human planners systematically underestimate feedback effects, discount historical performance, and rely on mental models that fail under real-world conditions. The Department of Defense now possesses the data and computational tools to break this pattern. By integrating DAVE's empirical data, AI-enabled forecasting, and system dynamics modeling into a unified cognitive-technical scheduling system, the acquisition enterprise can replace intuition-driven planning with evidence-based, continuously adaptive decision-making.

This integration does more than improve schedule accuracy. It strengthens governance by grounding milestone decisions in demonstrated performance, enhances foresight by revealing second and third-order effects before they materialize, and increases organizational learning by embedding feedback loops directly into scheduling processes. The resulting schedule digital twin provides a living representation of program behavior that updates as conditions change, identifies emerging risks, and recommends mitigation strategies with a speed and rigor no human team can match.

As the Department advances digital transformation, human-machine teaming, and AI assurance, CTI offers a practical pathway to institutionalizing analytic discipline across the acquisition life cycle. Programs that adopt this approach will not only achieve more realistic schedules but also gain strategic agility, reduce rework, and improve their ability to deliver capability at the pace of operational need. In an era where on-time delivery is often synonymous with strategic advantage, embedding this approach into acquisition practice is not simply an analytic improvement; it is a competitive imperative.

References

Agarwal, A., Gans, J., & Goldfarb, A. (2018). *Prediction machines: The simple economics of artificial intelligence*. Harvard Business Review Press.

Arena, M. V., Leonard, R. S., Murray, S. E., & Younossi, O. (2006). *Historical cost growth of completed weapon system programs*. RAND Corporation. https://www.rand.org/pubs/technical_reports/TR343.html

Baptiste, P., & Grunder, O. (1999). Cognitive science and project scheduling: More realistic representation. In J. Węglarz (Ed.), *Project scheduling: Recent models, algorithms and applications* (pp. 481–495). Kluwer Academic Publishers/Springer. https://doi.org/10.1007/978-1-4615-5533-9_20

BEA. (2012). *Final report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro – Paris*. Bureau d’Enquêtes et d’Analyses pour la Sécurité de l’Aviation Civile. https://web.archive.org/web/20190530095333/https://bea.aero/fileadmin/documents/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the “planning fallacy”: Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, *67*(3), 366–381. https://doi.org/10.1037/0022-3514.67.3.366

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. Wiley.

Bui, T. (1996). Supporting cognitive feedback using system dynamics. *Decision Support Systems*, *16*(4), 297–316.

Dörner, D. (1980). On the difficulties people have in dealing with complexity. *Simulation & Games*, *11*(1), 87–106. https://doi.org/10.1177/104687818001100108

Dörner, D. (1996). *The logic of failure: Why things go wrong and what we can do to make them right*. Metropolitan Books/Henry Holt.

Dörner, D., & Güss, C. D. (2022). Human error in complex problem solving and dynamic decision making: A taxonomy of 24 errors and a theory. *Computers in Human Behavior Reports*, *7*, Article 100222. https://doi.org/10.1016/j.chbr.2022.100222

Flyvbjerg, B. (2014). What you should know about megaprojects and why: An overview. *Project Management Journal*, *45*(2), 6–19.

Flyvbjerg, B. (2021). Top ten behavioral biases in project management: An overview. *Project Management Journal*, *52*(6), 531–546. https://doi.org/10.1177/87569728211049046

Flyvbjerg, B., & Gardner, D. (2023). *How big things get done: The surprising factors that determine the fate of every project, from home renovations to space exploration and everything in between*. Currency/Penguin Random House.

GAO. (2018). *Weapon systems annual assessment: Knowledge gaps pose risks to sustaining recent positive trends* (GAO-18-360SP).

GAO. (2023). *Weapon systems annual assessment* (GAO-23-106059).

GAO. (2025). *Weapon systems annual assessment: DOD leaders should ensure that newer programs are structured for speed and innovation* (GAO-25-107569).

Gigerenzer, G. (2024). Psychological AI: Designing algorithms informed by human psychology. *Perspectives on Psychological Science*, *19*(5), 839–848. https://doi.org/10.1177/17456916231180597

Kahneman, D., & Tversky, A. (1982). Intuitive prediction: Biases and corrective procedures. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 414–421). Cambridge University Press

Kington, T. (2025). *US Navy nixes Constellation frigate program after two ships half-built*. Defense News. https://www.defensenews.com/naval/2025/11/26/us-navy-nixes-constellation-frigate-program-after-two-ships-half-built/

Lovallo, D., & Kahneman, D. (2003). Delusions of success: How optimism undermines executives' decisions. *Harvard Business Review*, *81*(7), 56–63.

Mortlock, R. (2021). Behavioral biases within defense acquisition. In *Acquisition Research Program 18th Annual Acquisition Research Symposium*. Naval Postgraduate School.

Oakley, S. S. (2025). *Defense acquisition reform: Persistent challenges require new iterative approaches to delivering capabilities with speed* (GAO-25-108528). Testimony before the House Select Committee on Strategic Competition. GAO.

O'Rourke, R. (2025). *Constellation (FFG-62) class frigate program: Background and issues for Congress*. Congressional Research Service.

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

Stingl, V., & Geraldi, J. (2021). A research agenda for studying project decision-behaviour through the lenses of simple heuristics. *Technological Forecasting and Social Change*, *162*, Article 120367. https://doi.org/10.1016/j.techfore.2020.120367

Stingl, V., Gilchrist, A., Lawani, A., Flin, R., & Zwikael, O. (2025). Thinking is for doing: Project cognition as the foundation of project behaviour. *International Journal of Project Management*, *43*(1), Article 102678. https://doi.org/10.1016/j.ijproman.2025.102678

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Acquisition Research Program
Department of Acquisition, Finance, and Manpower
Naval Postgraduate School
555 Dyer Road, Ingersoll Hall
Monterey, CA 93943

www.acquisitionresearch.net