The Test and Evaluation of Unmanned and Autonomous Systems

Fil Macias

U.S. Army White Sands Missile Range,
White Sands Missile Range, New Mexico

Current Department of Defense test and evaluation capabilities and methodologies are insufficient to address testing of weapon systems operating in non–deterministic and unscripted modes characteristic of the unmanned and autonomous system. Task complexity and adaptability to the environment are critical for evaluation of unmanned and autonomous performance. This represents a new challenge for the test and evaluation community. Verification of system performance and interactions will require the tester to understand the nuances of multiple technical domains such as physical/Battlespace, information/knowledge, and cognitive/social.

Key words: Autonomy; behavior adaptation; cognitive agents; complex systems; test arena.

Those of us who can recall portions of our misspent youth watching *Lost in Space*, *Star Wars*, and *Star Trek* may find it somewhat ironic that we would be involved in testing robotic/intelligent machines for the U.S. Department of Defense (DoD). Although the technology is still nascent and advancing, we are faced with the innovator's dilemma and the need to improve unmanned autonomous systems (UASs) through advancing test and evaluation (T&E). We certainly do not test B9s or R2D2s at our test ranges and can agree that it is much too early to invoke Asimov's Laws of Robotics. Yet, intelligent machines in the form of UASs are now rapidly finding their way into the hands of the warfighter and are envisioned to provide amazingly new tactical capabilities in the near future in a variety of areas including mission assurance; command and control; and intelligence, surveillance, and reconnaissance. UASs will enable the warfighter to expand capabilities in the large arena of dull, dirty, and dangerous jobs typically performed by the soldier.

Introduction

The evolutionary nature of UAS acquisition must be met with evolutionary test capabilities yet to be discovered and developed. Test capabilities must be deployed at a faster pace than UAS deployment to satisfy the demand for warfighter improvements. The DoD is stimulating this new area of innovation with ongoing plans to invest more than $11 billion during the FY 2007–2013 timeframe out of a projected $24 billion-plus total budget for unmanned systems as referenced by the Unmanned Systems Roadmap 2007–2032 (DoD 2005).

The Pentagon's investment in UAS technology will drive innovation toward technology with increasing levels of intelligence and autonomy. UASs will enhance the tactical capabilities of warfighters by advancing decision making while performing dangerous, difficult, and monotonous tasks. Consider the example of a Predator unmanned autonomous vehicle designed for surveillance yet now also configured for weapons delivery. The specific needs of UAS employment are mission driven. As UASs become more intelligent, the potential for cognitive agents to enable and enhance these areas is expected to accelerate. UAS missions are expected to expand beyond simple remote sensing to include target illumination and weapons delivery. Physics-based testing is fundamental for knowledge rendering, expanded decision making, and collaborative action. The issue for T&E is that it must become more prominent during the acquisition process and frontloaded to create an advantage. What is that advantage? The advantage is to create a more viable test infrastructure that transfers UAS technology to the warfighter in a fast, efficient, and cost effective manner. Hence, testing is an important step in the evolutionary progress of any innovation.

Report Documentation Page

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **2008**	2. REPORT TYPE	3. DATES COVERED **00-00-2008 to 00-00-2008**

4. TITLE AND SUBTITLE **The Test and Evaluation of Unmanned and Autonomous Systems**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Army White Sands Missile Range,White Sands Missile Range,NM,88002**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**8**	

Current T&E techniques are suitable for systems with tightly coupled tethered operations. As we approach infinitesimally close to fully autonomous systems over a 30-year horizon, testing becomes enormously more difficult. Therefore, to address these and other testing limitations, the Test Resource Management Center established the Unmanned and Autonomous System Test (UAST) focus group, whose mission is:

1. Develop the technologies required to test and evaluate our transforming military capabilities. This includes any system that makes our warfighter more survivable and effective in combat:

a. lethal and non-lethal weapons;

b. manned and unmanned ground, sea, air, and space systems;

c. intelligence, surveillance, and reconnaissance systems;

d. information systems.

2. Provide the required T&E technologies in time to support developmental and operational tests to verify performance before production or deployment.

UASs as complex systems

Recognition that UASs represent a new type of technology with a new engineering genus is key to development of a UAST strategy. In fact, these systems may be characterized as complex systems (Braha et al 2006). Why? Consider the fundamental differences between a traditional system (what we test) and a complex system (what we must test). Systems designed under traditional means are expected to perform predictable tasks in bounded environments. A complex system, on the other hand, functions and operates in open, non-deterministic environments. These systems are composed of interconnected parts having one or more properties (behavior among the possible properties). The complex system is always greater than the sum of its parts. For example, humans are complex systems and so are robots. Complex system testing involves more than optimized testing procedures because many interactions happen among systems and subsystems. Testing is expected to be more fluid coupled with a component of uncertainty, commonly referred to as chaos.

Testing a UAS cannot be limited to the physical domain aspects of the individual system but must consider the systemics of the entire collaborating unit: humans, systems, and mission. Complex systems are studied by many areas of natural science, mathematics, and social science. Fields that specialize in the interdisciplinary study of complex systems include systems theory, complexity theory, systems ecology, and most importantly cybernetics.
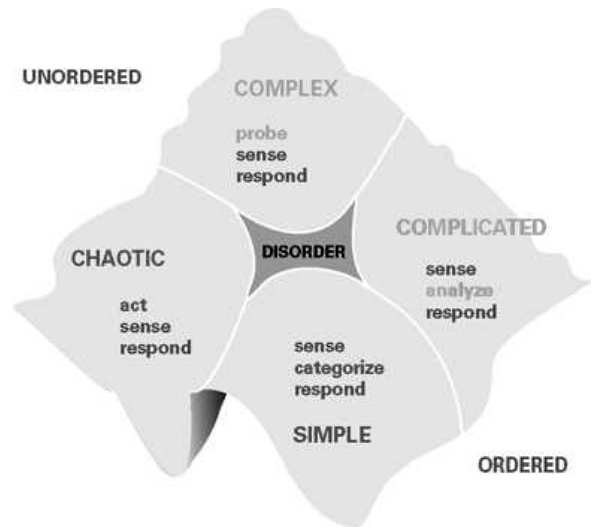

Figure 1. The Cynefin framework (Kurtz and Snowden 2003)

Traditional T&E is limited to single system focus with life cycle development numbered in years. System Interoperability is managed through interface requirements and integration of components. The focus is on reliability, maintainability, and availability within a centralized acquisition and management framework. Needs and requirements are primarily considered fixed at concept phase. The Cynefin framework is a sense-making device developed to help people make sense of complexities, see *Figure 1* (Kurtz and Snowden 2003).

In the ordered world of T&E, we have well developed guidance for simple and complicated systems but not for complex systems. As UASs trend to more complex and chaotic dynamics we must relax our strategies to enable us to establish emergent practices for complex systems and novel practices for chaotic systems deployment. During the next 30 years, both industry and academia are expected to provide UASs that are emergent and unanticipated, and T&E must be prepared to handle both ordered and unordered test requirements as capabilities mature. Testing UASs demands new tools and methods to address complex systems and action-based environments for chaotic system scenarios.

Autonomous, intelligent systems such as a UAS and its operators, will execute outside of predictable, stable behavior within carefully optimized situations. In order to be useful to the warfighter, a UAS must have the capacity for adaptation and change and be able to perform the unexpected. Consider the acquisition, deployment, and current mission/utilization of the MQ-1 Predator, a medium-altitude, long-endurance, remotely piloted air vehicle. The U.S. Air Force considers the Predator not just an aircraft, but a system consisting of four aircraft, a ground control

station, and other pertinent equipment. The Predator failed Operational Test and Evaluation (OT&E) (Coyle 2003). The problem for T&E is that the Predator went on to huge success on the battlefield. Even though the Predator failed to meet measures of performance, nothing is in place to give equal significance to measures of effectiveness. The assembly line process of acquisition and T&E failed to ensure that a viable solution could be passed. In this situation, T&E recommendations were overridden and the system was deployed. Although it would be easy to conclude that we no longer need T&E, this would be a hasty ill-advised conclusion with significant long-term consequences. Test, as it is practiced today, has huge overhead and is highly optimized for yesterday's problems. This is the reason why we need to think about new ways of testing and focus more on mission, capabilities, and effectiveness instead of the measures of performance that failed Predator.

In the future that we are facing today, UASs will be deployed on a timeline of months instead of years. Systems that are being developed in industry and academia have utility today to a warfighter who is facing enormous challenges. The question of how to start testing these systems in parallel with development may require us to move beyond the traditional test focus and towards a test strategy that covers the entire acquisition cycle from cradle to grave. The challenges of testing UASs are moving from simple system test toward the world of complex systems engineering.

Terminology is important in understanding the concepts of complex systems. Consider the following definition of a complex system (Kirshbaum 2008):

A complex system is any system that involves a number of elements, arranged in structure(s) that can exist on many scales. These go-through processes of change are not describable by a single rule nor are reducible to only one level of explanation; these levels often include features whose emergence cannot be predicted from their current specifications. Scientists are finding that complexity itself is often character-ized by:

■ *self-organization,*
■ *non-linearity,*
■ *order/chaos,*
■ *emergent properties.*

Classical test approaches emphasize real-time sup-port, communications, networking, and command and control that continue to be optimized to satisfy problems with declining returns. Current methodolo-gies are no longer relevant due to inherent need, expanding system interdependencies, and increasing complexity. A UAS has multiple interactions between many onboard components and an increasing set of external agents. The properties of a complex system cannot be completely explained by understanding only its component parts. In the case of a UAS, the complexities of human interaction, multi-system op-eration, cloud computing knowledge frameworks, sophisticated behavior models, collaboration, and expanding mobility all combine to create emergence that leads to even more complex adaptive behavior. Only by instituting positive feedback and negative feedback test frameworks, can these systems be sustained through intergenerational development. UAST provides the potential for expanding un-manned-based warfighter capabilities in requisite ways for better addressing mission and sustainability.

The latest Defense Acquisition Guidebook (DAG), Section 4.4.11.8, specifically devotes a section to Unmanned Systems (UMS) stating "…UMS and unmanned variants of manned systems are being rapidly developed and fielded to meet critical warfigh-ter capability needs…" (DoD 2007). Testing for UAS is currently being fast tracked but within a traditional T&E infrastructure that is limited in its ability to address the challenges of UAST. These challenges, however, are increasingly a function of human/UAS awareness/interaction (cognition) and autonomous control levels (autonomy) *(Figure 2)*. Complexity at the systems level is also matched by even more complexity when these systems are aggregated with other UAS and manned systems in system-of-systems and complex system scenarios. In addition, the requisite variety of products being fielded in the UAS technology sector is a function of both emergent needs and accelerating technology. Yet, T&E remains very systems-centric while the scope of UAS development has expanded to include human dynamics, cognition, knowledge representation, and autonomy (Braha et al 2006).

Missiles, aircraft, and ships are all complicated with well-defined and well-understood boundaries for test. They are expected to perform predictable tasks under the watchful eye of well-trained operators. Traditional engineering practices are ordered and linear. It is a goal oriented process that seeks to achieve known specific ends through well-defined means (Cook 2008). This process can be described through the following milestones:

- Functional Specification (what the system is expected to do),
- Design (how the system and components may look and function),
- Testing and Validation (procedure that sets procedure intended to establish the quality,
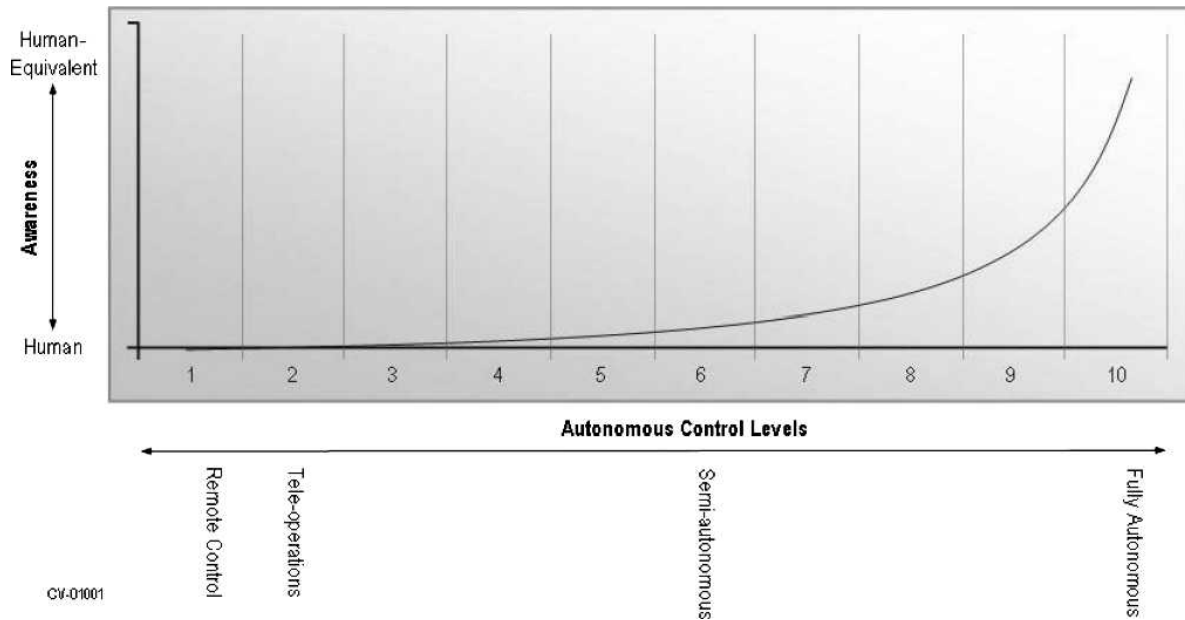
Figure 2. Unmanned systems levels of awareness versus levels of control (Braha et al 2006)

performance, or reliability of something; conditions designed to recreate reality to ensure it performs as needed to discover flaws and correct them),

- Production and Manufacturing (once designed and tested, copies are made).

Current T&E techniques are suitable for systems with manned operations and limited autonomous behavior. The nature of the UAS makes testing considerably more difficult. Whereas the mechanism for creating affordable systems with operational effectiveness is well understood, there are aspects of UAS that create new challenges for UAST. *Figure 3* from the latest Defense Acquisition Guidebook, Section 4.4, (DoD 2007) describes affordable system operational effectiveness with emphasis on measures of performance and measures of effectiveness, including emphasis on the traditional "-ilities" (reliability, maintainability, and supportability). However, for UAST, new "-ilities" have come to the forefront of discussion, namely flexibility, adaptability, and composability, especially as we target system-of-system and complex system scenarios.

UAS are inherently "complex systems" and traditional engineering and subsequently testing practices are not sufficient to address a complex support paradigm characterized by:

- systems-of-systems and complex systems;
- collaboration, extended enterprises, and federated operations;
- net-centric and mobility communications;
- capability driven acquisition and testing;

- sustainability of systems;
- design for flexibility, adaptability, and composability;
- uncertainty in the environment and predictability of human/systems collective action;
- model-based engineering and mission driven development;
- spiral processes, evolutionary acquisition, and open-source dynamics.

A framework for UAST

The UAST Framework consists of four categories: models and architecture, aspects and protocols, testbeds and test environments, and analytics. Each of these categories have been selected to enable a framework of dialog for research and development of a UAST infrastructure and supports an unfolding technology sector with expanding diversity and requisite variety. *Figure 4* illustrates the key areas of the UAST Framework, which will hereafter be referenced as the four technical topic domains.

Models and architecture

UAST models and architecture will become increasingly important based on trends toward model based systems engineering, the guidance of the DoD Architecture Framework, and the increasing relevance of Enterprise Architecture Framework (or Architecture Framework, for short). An architecture framework defines how to organize the structure and views associated within enterprise architecture. In this
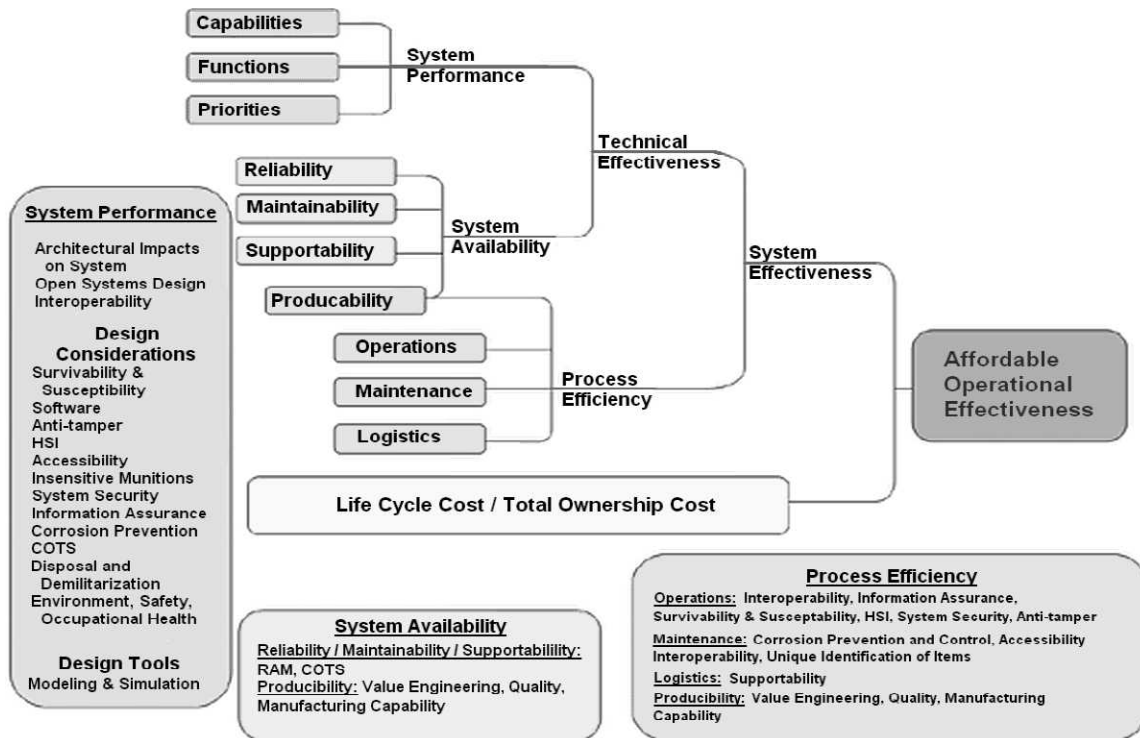
Figure 3. The Defense Acquisition Guidebook operation effectiveness diagram (DoD 2007)

specific case, the enterprise is the mission of establishing a UAS test framework that complements and augments existing initiatives in UAS testing while concurrently working toward a flexible and efficient approach for UAST. Because the discipline of

enterprise architecture is broad and because the enterprises it describes tend to be large and complex, the models associated with a particular technology sector can also be large and complex. To manage this scale and complexity, an architecture framework defines
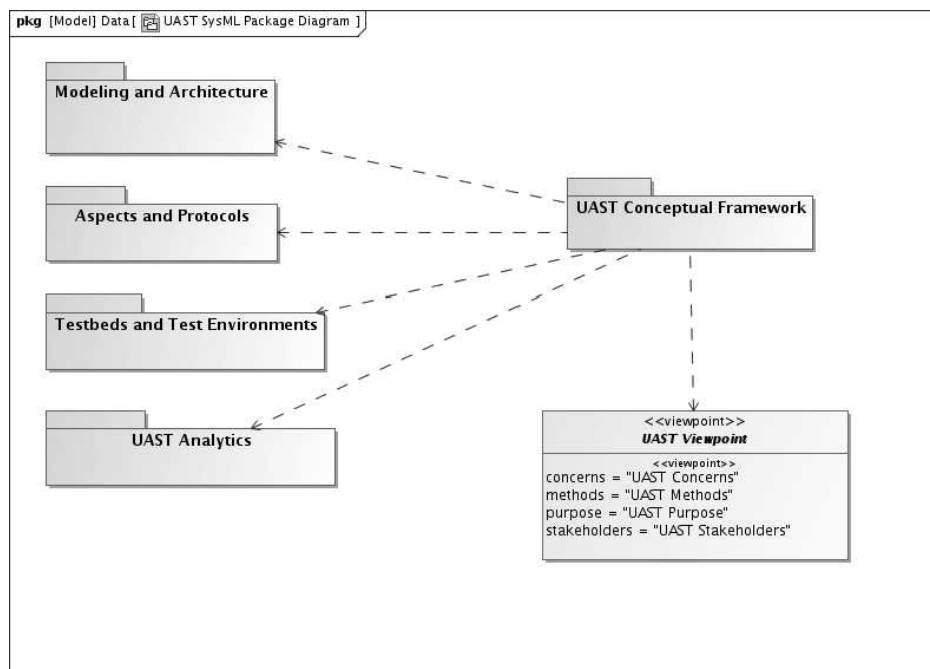

Figure 4. Unmanned and Autonomous System Test framework

complementary projections of the enterprise model, where each projection is meaningful to different system stakeholders. This category of the UAST framework seeks to expand and augment the development of architecture frameworks for UAST in areas such as adaptive architecture frameworks, collaborative tools, human-systems interactions, and decision making.

Analytics

The UAS analytics domain has been established to capture information from systems that inherently are based on awareness and control requirements for varying degrees of onboard intelligence and data gathering capacity. A UAS might be configured to have some level of on-board processing, analysis, and data mining that transforms data to knowledge. A key capability is to identify what information to share (vs data) to enable decision making in support of capability driven development. By understanding what the systems deem to be important information to make decisions, we can support test driven development. This enables identification of what information is important to the system while it is being developed. The information; however, must be assembled with other non-system information to produce knowledge. This knowledge can then be used to facilitate advanced decision making (for development). UAST decision making includes both verification and validation to answer the issues of "system right" and "right system." As we shift from a systems focus to a capabilities focus with the increased emphasis on system-of-systems and complex systems, then the core question becomes what platform can we establish to measure both performance and effectiveness, and how can this information be gathered in data centers to facilitate knowledge creation and system design decision making. The result is that information and knowledge from the testing supports UAS development. This category therefore seeks to identify and mine the information enabling a UAS to focus on mission effectiveness and task completion.

Testbeds and test arenas

Capabilities for UAST are based on sensing, knowledge acquisition and representation, decision making, and autonomous behaviors. In addition, these technologies will be structured to support testing or experimentation in an environment that provides for rigorous, transparent, and repeatable testing of UAS, sensors, simulators, computational tools, and other new technologies. The term for this environment varies across many disciplines; however, it will be referred to as a UAST test arena or testbed. This will support T&E of concurrent real sensor data and simulated entities (Hybertson and Sheard 2008) and it will:

- enable repeatable, controlled, and reproducible measurements and procedures to support federated testing involving multiple command groups and participants;
- provide a robust infrastructure that supports measurement of the UAS and subsystem's performance against *truth-data*;
- provide instrumentation that tracks UAS positions and orientations;
- support advanced sensors to characterize the UAS test environment and provide insight into the UAS functioning and health/status.

A test arena is seen as a set of assets (e.g., instrumentation, targets, etc.), each of which has specifiable relevance to a specific test request, but only if related assets are also available. In addition to test requests, maintenance and calibration requests for each asset must be included. The resulting map of test requests, asset conjoint relevance, and calendar time becomes a complex data structure that must be processed to inform test arena managers in their critical decisions regarding resource allocation and scheduling. We need to compute the test arena commitments for scenarios that will satisfy the greatest throughput of testing per unit of time. To do this, we need (a) a parametric (tailorable) model of a test arena, (b) a language for expressing the test requirements and implied assets, (c) resource allocation and scheduling algorithms, and (d) a dashboard that provides quantitative information about the asset utilization and user satisfaction, all within the context of flexible and efficient testing.

Aspects and protocols

The UAST aspects and protocols category is based on the proposition that UAS can better be understood by using multi-aspect analysis and protocols that offer procedural methodology in the design and implementation of test. In the 30-year timeline over which UASs are expected to fully evolve, a shift from systems assessment to capabilities assessment will be necessary per guidance and by necessity. The T&E of disruptive technology and capability with existing approaches has not kept up with the pace of deployment driven by accelerating need. A shift from an emphasis on system measures of performance toward system measures of effectiveness and mission assessment will continue. A simple component model equipped with protocol framework could be extended to provide aspect analysis. For example, a sensor (component) may have a goal (aspect) to reduce false alarms. Given this goal, it knows to request additional system information (via a protocol). The protocol of an aspect observes the service requests and replies from components and

reacts. A nice feature of this model is that an assembly of aspect and components can be transformed back into an assembly of components. All this is done without breaking the black-box nature of the component. A future model of this approach would be protocols frameworks for evaluating knowledge representation, decision making, and collaboration across the cooperating elements of the scenarios involving both systems and humans.

A test driven approach

The goal of UAST is to refine theory and practice for UASs. This is the scientific method of hypothesis and test toward verified conclusions. Without sophisticated UAST that is faster and prepositioned to meet future test challenges, we will never reach requisite variety in UASs without UAST that can match the pace and tempo of UAST capability deployment. An objective of UAST is to develop the basis of a framework that will enable a new testing strategy that establishes confidence and is constantly evolving and adapting to new challenges. It must offer a value proposition that supports safe test conduct and delivers the information and knowledge to support the decision process that expands opportunity and enables risk reduction. A test driven approach is a front-end process approach to test that considers chaotic and complex test scenarios in a rapid pace environment as real possibilities. This approach will require pre-concept test frameworks that enable a T&E strategy for integrated testing in a test continuum that executes in cycles for periods of months and not years (Braha et al 2006). The software world continues to successfully implement these ideas for the high pace rapid deployment world of internet cloud computing and enterprise software.

If an unmanned autonomous complex system is required to achieve a certain objective, the challenge to the tester is to create a learning environment that can deal with chaotic, complex, complicated, and simple (well known) scenarios. Providing reference problems with measures of performance enables researchers to compare implementations, communicate results, and leverage toward specification. It will always be important to develop test artifacts and measurement methodologies to capture performance data in order to focus research efforts. However, reductionist approaches toward testing often result in problem space explosion that translates into testing strategies that require years when the need is in months. Understanding UAST involves recognizing that traditional approaches to T&E are not sufficient to test the UAS over the 30-year period in which UASs are expected to become more intelligent and collaborative. Collaboration across systems and with

personnel leveraging internet cloud computing environments provides unforeseeable challenges. Traditional test/engineering methodologies continue to lag behind. We need a new strategy that takes into account the theories proposed by a complex systems community researcher like Yaneer Bar-Yam (Braha 2006). He lays out an intuitive framework for several concepts that can be applied to UAST as it continues to evolve:

- focus on creating an environment and process rather than a product,
- continually build on what already exists,
- individual components must be modified in-situ,
- operational systems include multiple versions of functional components,
- utilize multiple parallel development processes,
- evaluate experimentally in-situ,
- increase utilization of more effective parts gradually,
- effective solutions to specific problems cannot be anticipated,
- conventional systems engineering should be used for non-complex components.

Conclusions—a community of interest

UAS will be an unfolding challenge that must be met with an equally adaptive UAST. This is not an area where we can come with a single solution and believe that it will apply to the next 30 years. There is no universal model for a UAS and for that matter UAST. In the recently published biography on John Adams, he is quoted as stating the following: "Our different views of the same subject are the result of a difference in our organization and experience" (Coates 2008). This seems appropriate to describe the current state of UAST. UAST will be unable to succeed in delivering a value proposition to the testing community unless the testing community is involved. We need tester views, opinions, and experience to create a new trajectory for T&E that is UAST. UAST is currently supported by a focus group staffed by exceptional individuals from all over DoD. This group forms the basis of a community of interest to support UAST. This community of interest needs to consider the complexity of testing cognitive agents (UAS complex systems) with increasing autonomy involved in collaboration to achieve mission goals (Hybertson and Sheard 2008):

- Early tester participation,
- Multi-level assessment: Monitoring, assessment, and response occur at multiple levels.
- Plan-based assessment: Monitoring is triggered by an assessment of dependencies and constraints on plan execution.

- Capability-based assessment: Ongoing assessment of vehicle mission-related capabilities is based on subsystem and environment status.
- Predictive assessment: Monitoring and assessment anticipate future events or conditions.
- Team-based assessment: Assessment occurs not just of individual vehicles, but at the team level as well.

This is the challenge of UASs. UASs are needed for increased variety in warfighter capabilities but only with UAST can we test and evaluate for the requisite variety critical to the warfighter. For the T&E community, this should be seen as an opportunity to significantly improve the suitability and sustainability of this emerging technology sector. By enabling the requisite flow of UAS capabilities coming off the assembly line of industry we can better meet the ever-expanding array of problems related to traditional, irregular, catastrophic, and disruptive operations. ❏

FILIBERTO 'FIL' MACIAS is an electronics engineer in the Systems Engineering Directorate at White Sands Missile Range where he serves as supervisor of the Networks and Control Division. Fil has more than 24 years of T&E experience to include missile telemetry, test sensor/instrumentation development, and real-time modeling and simulation. He also serves on numerous advisory panels investigating the application of advanced technology to test and evaluation. In addition, he serves as the executing agent for the Unmanned and Autonomous Systems Project. He has bachelor's in electrical engineering and a masters of business administration from the University of Texas at El Paso. E-mail: Filiberto.a.Macias@us.army.mil

References

Braha, D., Minai, A. and Bar-Yam, Y. 2006. *Complex Engineered Systems*. Cambridge, Massachusetts: Springer-NECSI.

Coates, E. R. Sr. 2008. Life of Thomas Jefferson Reconciliation With John Adams. Available from http://etext.virginia.edu/jefferson/biog/lj37.htm. Accessed August 10, 2008.

Cook, S. 2008. "The Science of Systems Engineering and the Engineering of Complex Systems." *INCOSE Insight Magazine*. 11(1), pp. 8–11.

Coyle, P. E. 2003. Is Missile Defense on Target? *Arms Control Today*. Washington, DC. Arms Control Association. Online journal. Article available at http://www.armscontrol.org/act/2003_10/Coyle_10. Accessed August 10, 2008.

DoD. 2007. "Unmanned Systems Safety Guide For DoD Acquisition." First Edition (Version .96), January 2007. Washington, DC: OUSD (AT&L) Systems and Software Engineering/Developmental Test & Evaluation.

DoD, Office of the Secretary of Defense. 2007. "Unmanned Systems Roadmap 2005–2032." Washington, D.C.: Government Printing Office.

Hybertson, D. and Sheard, S. 2008. "Integrating and Unifying Old and New Systems Engineering Elements." *INCOSE Insight Magazine*. 11(1), pp. 5–7.

Kirshbaum, D. 2008. Complex Adaptive Systems. Available from http://www.calresco.org/lucas/cas.htm. Accessed August 10, 2008.

Kurtz, C. F. and Snowden, D. J. 2003. "The New Dynamics of Strategy: Sense-Making in a Complex and Complicated World." *IBM Systems Journal*. 42(3), PP. Available at http://en.wikipedia.org/wiki/Cynefin. Accessed August 11, 2008. pp. 467–470.