Some Musings on Test and Evaluation of Unmanned Ground Vehicles

Ellen M. Purdy

Joint Ground Robotics Enterprise,

Office of the Under Secretary for Defense (Acquisition, Technology, and Logistics), Washington, D.C.

The U.S. Department of Defense has come a long way in the last 5 years when it comes to fielding and employing unmanned ground vehicles. In 2003 with operations in Afghanistan, 3 Packbots were sent to the theater of operations to support cave reconnaissance. The systems worked so well that there were calls for more. Then came the idea of using ground robots to defeat improvised explosive devices. Experiences with these robots in the theatre proved such a success that today there are more than 6,000 unmanned ground vehicles in the inventory, the majority of them serving in Iraq and Afghanistan. No doubt about it, ground robotics are and forever more will be part of the materiel that enables our Service members to conduct their missions. With such a rapid adoption of this technology, most of which was acquired through rapid means, it is worth taking some time to consider whether the Department is fully prepared to test and evaluate the new robotic systems that will be called for in the future.

Key words: Autonomous tactical behavior; unmanned systems; hijacked system; rogue behavior; safety; reliability & mission duration; test metrics; trust; weight; speed.

Recently, Service experts with unmanned systems experience were engaged in an exercise to identify potential missions that could feasibly be conducted by unmanned systems during the next 25 years. For the ground domain, missions ranged from reconnaissance to casualty evacuation, range clearance, runway inspection, dirty bomb interrogation and disposal, tunnel exploration and mapping, and facility protection, to a myriad other missions. These experts also projected that in 25 years unmanned systems would likely advance from the teleoperation of today to fully autonomous tactical behaviors. They would likely be able to fully and autonomously team with manned systems and other unmanned systems across air, land, and sea.

These technological feats are unprecedented in military history, so it begs the question of whether there is work to do when it comes to conducting appropriate and reliable test and evaluation (T&E) of these systems. The purpose of this article is to encourage dialog now, while unmanned systems are still in their infancy, particularly when it comes to full autonomy. Such dialog should take place not only among T&E personnel

Ellen M. Purdy

but also with users. As testers and evaluators wrestle with the difficult issues associated with the unique aspects of unmanned systems, they are in a position to assist users with sorting through appropriate measures of operational effectiveness.

Test considerations: T&E of unmanned systems is really no different for manned systems, or is it?

Unmanned systems, by their nature when they cross the boundary into fully autonomous behaviors, will be doing many of the perception and reasoning tasks that have up to this point always been done by people. Do we have all the right test methodologies to test these capabilities? Maybe, maybe not. A case in point is a study published in January 2008 by the Massachusetts Institute of Technology (MIT) regarding the commonly adopted methodology for assessing the effectiveness of computer vision. According to MIT News:

"For years, scientists have been trying to teach computers how to see like humans, and recent research has seemed to show computers making progress in recognizing visual objects. A new

		Form Approved OMB No. 0704-0188
Report Documentation Page	

1. REPORT DATE **DEC 2008**	2. REPORT TYPE	3. DATES COVERED **00-00-2008 to 00-00-2008**
4. TITLE AND SUBTITLE **Some Musings on Test and Evaluation of Unmanned Ground Vehicles**		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) **Office of the Under Secretary for Defense,(Acquisition, Technology, and Logistics),Joint Ground Robotics Enterprise,Washington,DC,20301**		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)**	**3**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

MIT study, however, cautions that this apparent success may be misleading because the tests being used are inadvertently stacked in favor of computers.” (Delude 2008)

When it comes to unmanned systems in combat situations, it is imperative we get our testing right so that we accurately characterize what these systems can and cannot do.

Given that T&E will have to mature right along with unmanned systems, or better yet at a faster pace so testing does not slow down technology development and system acquisition, with what exactly should testers and evaluators be concerned? The first thing that comes to mind is metrics and how they impact test infrastructure. The U.S. Department of Defense (DoD) test community is very good and very experienced at testing mobility of platforms on the ground. Consider this though: so far all the vehicles that have been tested regardless of speed and types of terrain have always been “governed” by the constraints associated with passengers. There are certain speeds and terrains these vehicles do not traverse because human bodies are not built for that kind of punishment. Our vehicle technology is capable of higher speeds and able to traverse harsher terrains as long as there is no human being to worry about getting dizzy, sick, or bruised. In our current test tracks, do we have the kinds of terrain roughness that unmanned ground vehicles (UGVs) will be able to handle across the next 25 years—at UGV speeds?

What about trust? What are our metrics and how do we measure a person’s trust of unmanned systems? Twenty-five years from now there will be an entire generation that does not know what life is like without unmanned systems. Today, however, the size of the population on the planet that has worked with robots is so small it is buried in the noise. Let’s face it, there are just not that many people who have had up close and personal encounters with robots. Should we not have a test for measuring a person’s willingness to let a vehicle autonomously drive them from point A to point B? This would be good information to know before fielding a system so the training support package can take this into account. Experience to date indicates that willingness to trust a robotic vehicle differs widely between people. Some personnel become very comfortable quickly with being a passenger in an autonomous vehicle whereas others with weeks of experience in an unmanned autonomous vehicle still hover with their hands inches from the steering wheel, just waiting to take over should the perceived need arise.

Lastly, what about the basic test infrastructure for unmanned ground systems? The DoD has, over the years, developed extensive test facilities, but unmanned systems bring a new set of considerations. For smaller systems, for example, a “road course” may require obstacles such as curbs, winding stairwells, tunnels, puddles, etc. Instrumentation may pose additional challenges. How will test data be collected when the instrumentation suite normally mounted on a vehicle is now larger and heavier than the system being tested (think micro-robots)? With the current emphasis on tunnels, how will the data collected be relayed to the data collection location? Will global positioning systems provide location data on a system in a tunnel, as is currently assumed?

Evaluation considerations

What about metrics when it comes to perception? Much of our testing with regard to perception relies on human feedback. So how do you get an unmanned system to tell you why it did or did not perceive a negative obstacle (examples of a negative obstacle include a hole in the ground or puddle of water)? How good does perception have to be in order to be operationally effective? As good as a human, better than a human, or can we afford to accept less perception for the trade of human standoff and less exposure to risk?

Now that we are talking evaluation, that is a significant challenge for the T&E community. When it comes to a platform without a human operator on board, just what exactly constitutes operationally effective, suitable, and survivable? In many cases the same measures of effectiveness will suffice, but some will not, and some new ones will likely have to be invented. This question really comes down to redefining the trade space—how much performance are we willing to give up for the sake of greater safety and less risk of loss of limb and life?

Let’s talk survivability—how does one define survivability when it comes to an unmanned system? The easy part is we do not have a human operator to worry about, but where is the knee of the curve when it comes to how much survivability to design into the system? Where is that limit that says do not make the system so valuable that humans are required to protect it? One thing is certain, the requirements community does not really know, so maybe testers and evaluators can collaborate with users to sort through this new frontier of metrics. Sounds like a perfect collaboration, and the sooner the better.

When evaluating suitability, inevitably the notion of reliability comes into play. This has always been an important aspect but when it comes to unmanned systems, suddenly the implications of reliability take on a new meaning. After all, there will not be an operator around to get down from the vehicle to change the flat tire. Clearly the reliability of a system must be no less than that of its expected mission duration. When it comes to unmanned systems, however, the old limits

do not apply. Robots do not have to stop to rest and eat, so mission durations of days, weeks, months, years are being envisioned during the next 25 years. This is not as far fetched as you might think. The Defense Advanced Research Projects Agency is developing a high altitude, long endurance unmanned aerial system expected to provide persistent surveillance during a period of 5 years. When it comes to the air domain we pretty much know how to test for this; after all, our satellites have been doing just fine, thank you.

What about UGVs? Just what exactly is the test methodology to statistically prove the reliability of UGVs expected to conduct missions measured in months and years? After all, the old joke about the trouble with UGVs is that they operate on the ground. Alot of unexpected things can happen on the ground that are not likely scenarios in the air. Do we need to test for every contingency? That is alot of combinations and permutations—so just how do we crack this nut without bankrupting the program budget?

Another thing about reliability—there is a different dimension when it comes to unmanned systems, that of autonomy. How many hours of operation will be required to statistically prove that the system will not go "rogue"? Today we do not test the propensity of a vehicle to suddenly drive off course and into a building when a driver is behind the wheel, but we will have to do just that with UGVs. Not convinced this is an issue? It happened at the Defense Advanced Research Projects Agency Urban Challenge. With no warning one of the vehicles that was doing splendidly suddenly left the course and nearly plowed into a building. It missed impacting the building by mere inches because the manned chase vehicle following it activated the e-stop. When it comes to rogue vehicles, what are all the protocols that will need to be invoked to prove the unmanned system is safe from being "hijacked"? In other words, how will we test the systems to ensure the enemy cannot retask the system to conduct a mission against us? Do we know how to characterize this risk?

Since we are talking safety, do we have all the test infrastructure needed to ensure safe testing? So much of our safety on the test range today depends on humans. What will we need in terms of infrastructure to ensure safe testing when it comes to 9 ton or larger UGVs. How will we ensure safety when testing a large vehicle at high speed over rough terrain? Do we have sufficient e-stop performance? After all, when it comes to large mass at high speed, it takes more time and distance to stop.

So now what?

The discussion up to this point has been more in the nature of posing questions than in recommending answers. That is deliberate. The T&E community will come up with the answers in due time. What the DoD needs is for that "due time" to well precede the point at which unmanned systems will be ready for testing. In order to ensure timely delivery of needed systems to the warfighter, program managers should not be confronted with surprises because of uncertainty regarding how to test the system. What is so encouraging is that the DoD T&E community is not resting on its laurels and assuming that present methodology is sufficient. They are leaning forward in the foxhole to tackle these issues, and they are doing it at a particularly challenging time when there just are not that many systems to learn from. With continued engagement among the T&E community, the robotics experts in the laboratories, industry developers, and the user community, the answers to the questions posed above will be answered in a timely and effective manner, and eventually the population of personnel working with robots will pretty much number everyone in uniform. ❏

Mrs. Ellen M. Purdy is the enterprise director, Joint Ground Robotics within the Office of the Under Secretary of Defense for Acquisition, Technology and Logistics. She is responsible for oversight of the $1.7B unmanned ground systems portfolio and strategic planning for ground robotics. She prioritizes and manages $50M annual Budget Activity 4 and 5 funding, and ensures ground robotic technology development and acquisition complies with statutory and Congressional mandates.

Previously Mrs. Purdy served in the Program Management Office, Future Combat Systems Brigade Combat Team (PMO FCS (BCT)). During 5 years with FCS, Mrs. Purdy served in a variety of positions including manager, Test, Analysis, Modeling and Simulation; deputy director, Combined Test Organization; director, Combined Test Organization; and integrated product team lead for Integrated Simulation and Test. Mrs. Purdy was responsible for a 60 person organization and a $2.2B test program that included testing at the system of systems level and each of the 18 individual platforms plus the network that comprise FCS.

Mrs. Purdy has a bachelor of science degree in chemical engineering a master of science degree in engineering management. She is a member of the Board of Directors for the Association of Unmanned Vehicle Systems International and a member of the Advisory Board for the Virginia Center for Autonomous Systems. Her awards include two Superior Civilian Service Award medals, two Commander's Award for Civilian Service medals, and the Achievement Medal for Civilian Service. Mrs. Purdy has published over twenty technical reports and journal articles.

References

Delude, C. M. 2008. "MIT Neuroscientists See Design Flaws in Computer Vision Tests." McGovern Institute, January 24, 2008.