Unmanned System Safety Precepts

NDIA 2018 Ground Robotics Capabilities Conference and Exhibition

Presenters:
Robert Alex, Booz Allen Hamilton
Presenting for: Mr. Michael H. Demmick,
DoD, UxS Safety IPT Chair

UxS Safety IPT Objectives

- ✓ **Updated 2007 Guide and Developed New Precepts**
  - ✓ **Filled critical gaps in AI, Autonomy, V&V**
  - – Subsequent to the 2007 UMS Safety Guide, the DoD perspective on autonomy evolved
  - – 2016 study by the Defense Science Board titled, “The Role of Autonomy in DoD Systems,” highlights need for a dynamic approach to evolving DoD policy regarding autonomous systems
- ✓ **Interfacing with Services**
  - – DOA – integrate Networked Munitions Requirements
  - – DON – interface with DASN UxS & RDT&E
  - – DAF – interface with USAF Safety Directorate
- ✓ **Collaborating with stakeholders**
  - – Collaborating with DOS [*the UN CCW LAWS talks*] and Defense Science Board
  - – Ensure unique interests, capabilities, and concerns are shared, leveraged, and addressed
  - – Integrate other Federal Agencies with similar interests
- – **Institutionalize UxS Safety Guidance**
- – **Align System Safety Engineering Criteria & Requirements with:**
  - – Programmatic, Design, and Operational Requirements

Unmanned System Safety Guide

- **The purpose of this guide is to aid the PM's team, the operational commander, and the systems engineer in recognizing and mitigating system hazards unique to partially or fully autonomous design capabilities.**
- **It augments the tasks within MIL-STD-882 with additional details to address UxSs and the incorporation of greater levels of autonomy and machine learning.**
- **Autonomous capabilities create unique safety challenges beyond those addressed in other safety guidance.**
- **This guide lists safety precepts that must be followed in order to address safety with respect to programmatic, operational, and design considerations**

Guide sets threshold of rules of behavior that manage programmatic, design, & operational characteristics & aligns requirements

Programmatic Safety Precept (PSP) = Program management principles that help insure safety is adequately addressed throughout the lifecycle process.

Operational Safety Precept (OSP) = Directed at system operation setting operational rules to be adhered to. OSPs may generate the need for DSPs.

Design Safety Precept (DSP) = Provides Design guidance & facilitates safety of the system and minimizes hazards. Safety design precepts are intended to influence, but not dictate, solutions.

UxS Safety Challenges

Critical Gaps

[*no substantive safety guidance or policy in place*]

1. Diverging & Missing Definitions
2. Authorized Entity Controls
3. Flexible Autonomy
4. Fail Safe Autonomy
5. Autonomous Function V&V
6. Artificial Intelligence (AI)

- **Highly Complex & Evolving Technologies**
  - Understanding technological complexities associated with Gap areas and their relationship to safety
- **AI technology advancing faster than expected and with less safety assurance**
- **Unmanned Systems (UxS's) cross many boundaries & environmental domains**
  - Cross Service and Cross Agencies - all Department of Defense (DoD) services and operational domains
  - Research & Development and S&T organizations
  - Various Federal Agencies & Industry e.g., DOT, NGA, DOE, DHS, USCG, etc.
- **UxS Lexicon**
  - Taxonomy gap bigger / more central than expected
  - To ensure guidance is effective terminology, lexicon, and definitions must align
    - New and unique terms evolve as a result of on-going scientific research and engineering
- **AI risk mitigation methodologies and techniques are at best immature**
  - E.g., V&V; Probabilistic software analytics; code level analysis techniques; etc.
  - Difficulties exacerbated in a Rapid Acquisition environment

Critical Gaps

#	Critical Gap Name	Rationale for Declaring a Critical Gap, and Gap Description	Impact on UMS Safety Document
1	**Diverging & Missing Definitions**	***Rationale:*** Ensure that safety guidance is interpreted and applied in a manner consistent with the intent of DoD directives and policy, and mindful of international influences. ***The Gap:*** The 2007 UMS Safety Guidance definition of "UMS" diverges from policy. Definitions are missing for: "autonomous system", "semi-autonomous system", "autonomous function", "cognitive autonomy", "LAWS", "LARS", "Human Control", "Human Judgment", and more.	Rewrite Section 1 with best available definitions.
2	**Authorized Entity Controls**	***Rationale:*** Ensure that unmanned systems include Human Control that is appropriate and meaningful, per DoD directive and U.N discussions and in accord with safety precepts. ***The Gap:*** Current guidance allows for any function to be taken over by autonomous systems. There is no guidance ensuring Human in the loop at any level.	Changes throughout document; New PSP, OSP, and possibly DSP.
3	**Flexible Autonomy***	***Rationale:*** a. Enable continued legal use of systems as policies evolve. b. Keep up with evolving technology, adversaries, and CONOPs by enabling safe, rapid insertion and upgrade of autonomous functions. c. Support filling Critical Gaps 2, 4 and 5. ***The Gap:*** No safety precepts to ensure timely system safety upgrades as requirements evolve.	Changes throughout document; New DSP and perhaps OSP.
4	**Fail Safe Autonomy**	***Rationale:*** Mitigate multiple hazards that are new or more critical for autonomous functions** ***The Gap:*** No precepts to mitigate autonomy critical hazards, such as: a. Loss, or suspected loss, of data feed integrity, b. Hack by autonomous system usurping functions that, by law or policy, require human control, c. Hack by enemy, and d. Fail safe on "Terminator Scenario"***	New OSP(s) and DSP(s).
5	**Autonomous Function V&V**	***Rationale:*** S&T efforts indicate that new methods are required for autonomous function V&V, and are developing new methods, such as trust based validation.**** ***The Gap:*** Lack of guidance for safety testing of autonomous functions.	New document section; New DSP, OSP, and edits to SPs added for Critical Gap 2.
6	**Aritifical Intelligence (AI)**	***Rationale:*** Consider new precept[s] that address the use of AI in system decision making; presently UMS precepts focus on Software based logical transitions that are pre-programmed and pre-determined to occur with pre-determined sequencing. AI would potentially impose unpredictability into the equation. ***The Gap:*** Lack of guidance for safety analysis of AI level software or functions	This Gap may have an effect on how Gaps 2 – 5 are addressed, i.e. precepts for 2 – 5 could be written to address AI

* Source of Critical Gap Name: Air Force doc "Autonomous Horizons" (June 2015).
** See MIL-HDBK-516c (Dec 2014) for further discussion regarding such hazards.
*** Term used by RDML Selby at 2nd NSWC Dahlgren Unmanned Systems Integration Workshop and Technical Exchange Meeting.
****Perhaps interact with S&T community to mature new V&V methods for autonomous functionality and with G48 to evolve MIL-STD 882 accordingly.

Safety Issues with UxS

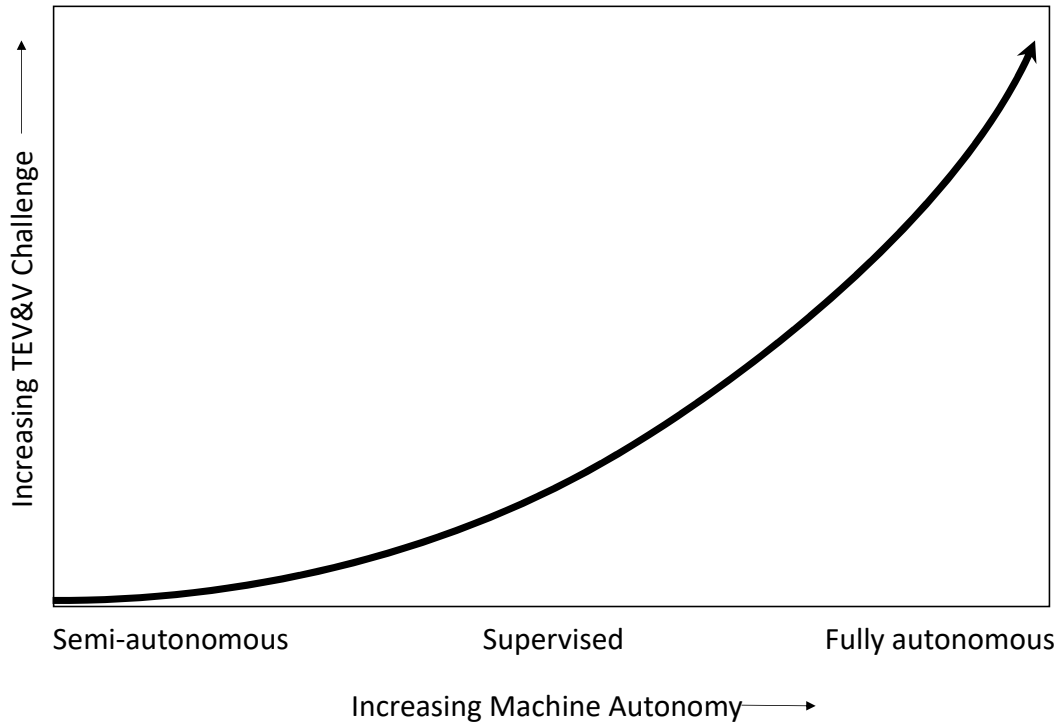
- **Autonomous UxSs inherently introduce potential mishap risk to humans for many different reasons, ranging from unpredictable movements, to loss of absolute control, to potential failures in both hardware and software.**
- **Weaponized UxSs present even more significant and complex dangers to humans.**
- **Typical safety concerns for military UxSs, that apply across semi-autonomous, supervised, and fully autonomous UxSs include:**
  - Loss of control over the UxS
  - Loss of communications with the UxS
  - Loss of UxS ownership (lost out of range or to the enemy)
  - Loss of control of UxS weapons
  - Unsafe UxS returns to base
  - UxS in indeterminate or erroneous state
  - Knowing when an UxS potentially is in an unsafe state
  - Unexpected human interaction with the UxS
  - Inadvertent firing of UxS weapons
  - Erroneous firing of UxS weapons
  - Erroneous target discrimination
  - Enemy jamming or taking control of UxS

Key Autonomy Safety Focus Points

- **Achieving Safety with Autonomy**
  - When tasks are assigned, the assigner bounds the assignment when issuing the task, and checks the bounds when the plan is generated
  - When autonomous functions are operating in a semi-autonomous mode, the human does the bounds checking
- **Bounding Autonomous Functionality**
  - Once the human is out of the loop (fully autonomous), deterministic bounded software becomes a real-time validator of the autonomous function or a notification for a human that an autonomous activity is taking place
  - Without separate deterministic bounding software, hazards may increase and trust may decrease when novel solutions are offered by the autonomous functions
- **Managed Machine Learning & Learning Mode**
  - A side effect of machine learning is the potential to execute unsafe decisions
  - The use of machine learning is expected to increase
  - Managed machine learning, or the concept of “Learning mode”, provides a tool to enable or disable machine learning and a mitigation to associated potential risk
- **Flexible Autonomy**
  - Flexible autonomy allows, without reprogramming, rapid safe reconfiguration of the system based on validation results, field experience with the system, changing mission parameters or rules of engagement, DoD policy and more.
  - It allows people to rapidly grant the system more autonomy as trust is developed. It also allows people to rapidly revoke autonomy where trust has been compromised.

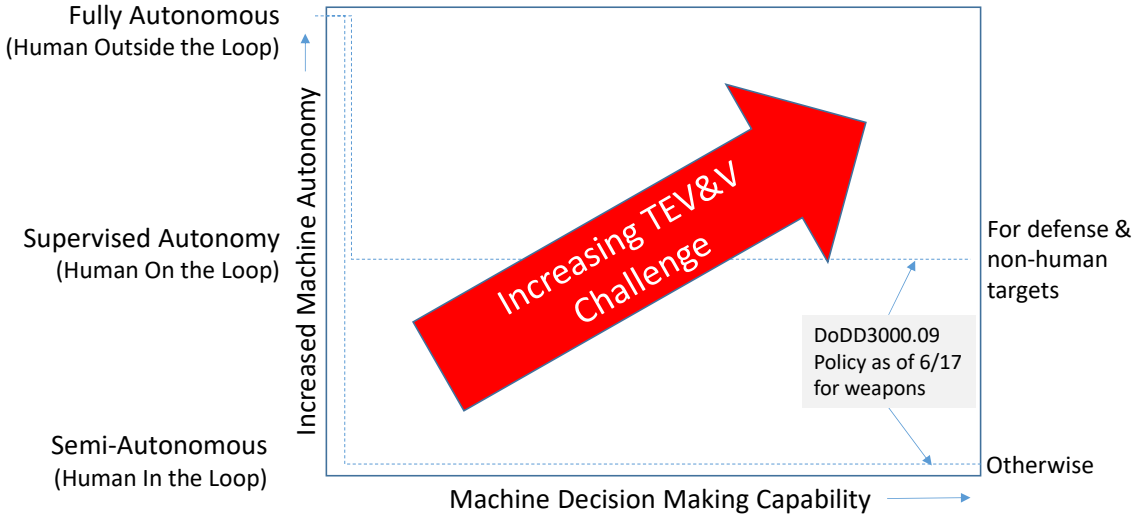
TEV&V Challenges

- **The relative magnitude of the challenge as a function the extent of autonomy in the system has been estimated as being exponential due to state-space explosion and increasing lines of software**



TEV&V Challenges

- **The challenge to make the system capable and safe while meeting policy and passing the TEV&V portion of the acquisition process increases both as the machines decision making capabilities increase and as the degree of autonomy that it is provided increase.**



	Automatic	Automated	Autonomy	
			Behavioral	Cognitive
V&V	Verifiable	Technically Verifiable	Non-Verifiable	TBD
Sample Technology	Data Processing	Expert System	Machine Learning	TBD
Complexity	Simple	Complex	Highly Complex	TBD
Outputs	Deterministic		Probabilistic	TBD

Programmatic Safety Precepts

- **PSP-1**
  - Establish and maintain a Systems Safety Program (SSP) in accordance with MIL-STD-882 (current version) for all life cycle phases.
- **PSP-2**
  - Establish consistent and comprehensive safety precepts across all UxS programs under their cognizance to ensure:
    - Mishap risk is identified, assessed, mitigated, and accepted
    - Each system can be safely used in a combined and joint environment
    - That all safety regulations, laws, and requirements are assessed and addressed
- **PSP-3**
  - Ensure that off-the-shelf items (e.g., COTS, GOTS, NDI), re-use items, original use items, design changes, technology refresh, and technology upgrades (hardware and software) are assessed for safety, within the system.
- **PSP-4**
  - Ensure compliance to and deviation from the UxS safety precepts are addressed during program reviews such as System Safety Working Groups (SSWG), System Readiness Reviews (SRR), Preliminary Design Reviews (PDR), & Critical Design Reviews (CDR) and Internal Program Office Reviews (IPR).

Programmatic Safety Precepts

- **PSP-5**
  - Ensure the UxS complies with current safety policy, standards, and design requirements.
- **PSP-6**
  - Ensure that the UxS, by design, does not allow subversion of human command or control of the UxS.
- **PSP-7**
  - Ensure that safety significant functions and components of an UxS are not compromised when utilizing flexible autonomy where capabilities or functions can be added, removed, enabled or disabled.
- **PSP-8**
  - Prioritize personnel safety in unmanned systems intended to team with or operate alongside manned systems.
- **PSP-9**
  - Ensure authorized & secure control (integrity) between platform and controller to minimize potential UxS mishaps and unauthorized Command and Control (C2).
- **PSP-10**
  - Ensure that software systems which exhibit non-deterministic behavior are analyzed to determine safe employment and are in compliance with current policy.

Operational Safety Precepts

- **OSP-1**
  - The control entity of the UxS should have adequate mission information to support safe operations.
- **OSP-2**
  - The UxS shall be considered unsafe until a safe state can be verified.
- **OSP-3**
  - The control entity of the UxS shall verify the state of the UxS to ensure a known and intended state prior to performing any operations or tasks.
- **OSP-4**
  - The UxS weapons should be loaded and/or energized as late as possible in the operational sequence.
- **OSP-5**
  - Only authorized, qualified and trained personnel using approved procedures shall operate or maintain the UxS.
- **OSP-6**
  - Ensure the system provides operator awareness when non-deterministic or autonomous behaviors are utilized in the various phases of the mission.

Operational Safety Precepts

- **OSP-7**
  - The operator should establish alternative recovery points prior to or during mission operations.
- **OSP-8**
  - Weapon should only be fired / released with human consent, or control entity consent and in conjunction with preconfigured criteria established by the operator.
- **OSP-9**
  - When the operator is aware the UxS is exhibiting undesired or unsafe behavior, the operator shall take full control of the UxS. [manual override]
- **OSP-10**
  - The operator must have the ability to abort/terminate/kill the mission of the UxS. [Terminate system]
- **OSP-11**
  - During mission operations the operator shall enable or disable learning mode to avoid hazardous or unsafe conditions. [learning mode]
- **OSP-12**
  - The control entity must maintain positive and active control of the UxS when any transfer of control has been initiated.

Design Safety Precepts

- **DSP-1**
  - The UxS shall be designed to minimize the mishap risk during all life cycle phases.
- **DSP-2**
  - The UxS shall be designed to only fulfill valid commands from the control entity.
- **DSP-3**
  - The UxS shall be designed to provide means for C2 to support safe operations.
- **DSP-4**
  - The UxS shall be designed to prevent unintended fire and/or release of lethal and non-lethal weapon systems, or any other form of hazardous energy.
- **DSP-5**
  - The UxS shall be designed to prevent release and/or firing of weapons into the UxS structure itself or other friendly UxS/weapons.
- **DSP-6**
  - The UxS shall be designed to safely initialize in the intended state, safely and verifiably change modes and states, and prevent hazardous system mode combinations or transitions.
- **DSP-7**
  - The UxS shall be designed to be able to abort operations and should return to a safe state.

Design Safety Precepts

- **DSP-8**
  - Non-deterministic software, as well as safety critical software, shall be physically and functionally partitioned.
- **DSP-9**
  - The UxS shall be designed to minimize single-point, common mode or common cause failures, that result in high and/or serious risks.
- **DSP-10**
  - The UxS shall be designed to mitigate the releasing or firing on a friendly or wrong target group selection.
- **DSP-11**
  - The UxS shall be designed to transition to a pre-configured safe state and mode in the event of safety critical failure.
- **DSP-12**
  - The UxS shall be designed for safe recovery if recovery is intended.
- **DSP-13**
  - Use of the UxS newly learned behavior should not impact the UxS' safety functionality until the newly learned behavior has been validated.

Design Safety Precepts

- **DSP-14**
  - Autonomy shall only select and engage targets that have been pre-defined by the human.
- **DSP-15**
  - Common user controls and display status should be utilized for functions such as: Manual Override (OSP-9), Terminate Mission (OSP-10), and Learning Mode (OSP-11).