Ethical Control of Unmanned Systems

Mission Design and Semantic Web Exemplars for Human Supervision of Lethal/Lifesaving Autonomy

Don Brutzman

Information Sciences, Undersea Warfare Academic Group, Modeling Virtual Environments Simulation (MOVES)

CRUSER, Naval Postgraduate School (NPS)

12 July 2021

Synopsis: Ethical Control of Unmanned Systems

- **Project Motivation**: ethically constrained control of unmanned systems and robot missions by human supervisors and warfighters.
- **Precept:** well-structured mission orders can be syntactically and semantically validated to give human commanders confidence that offboard systems
  - *will **do** what they are told to **do***, and further
  - *will **not do** what they are **forbidden to do***.

Paraphrase: can qualified robots correctly follow human orders?

- **Project Goal:** apply Semantic Web ontology to scenario goals and constraints for logical validation that human-approved mission orders for robots are semantically coherent, precise, unambiguous, and without internal contradictions.
- **Long-term Objective:** demonstrate that no technological limitations exist that prevent applying the same kind of ethical constraints on robots and unmanned vehicles that already apply to human beings.

https://savage.nps.edu/EthicalControl

Autonomous Vehicle Command Language (AVCL)

- AVCL is a command and control language for humans supervising autonomous unmanned vehicles.
  - Clarity arises from close correspondence to human naval terminology.

Mission Tasking

- Structured vocabulary defining terms and relationships for mission planning, execution, conduct, recording and replay across diverse robot types.
- Common-ground XML representations for
  - Mission agenda plans, mission scripts, and post-mission recorded telemetry results.
  - ***Future work:*** defining unit tests and expected results for verification and validation.
- Operators have single archivable, validatable format for robot tasking, results
  - directly convertible to and from a wide variety of different robot command languages.

https://savage.nps.edu/Savage/AuvWorkbench/AVCL/AVCL.html

Example AVCL mission agenda, as pseudo-code XML

```
<?xml version="1.0" encoding="UTF-8"?>
<UUVMission>
  <GoalSet>
    <Goal area=”A” id=”goal1”>
      <Search nextOnSuccess=”goal2” nextOnFailure=”goal3”/>
    </Goal>
    <Goal area=”A” id=”goal2”>
      <SampleEnvironment nextOnSuccess=”goal3”
          nextOnFailure=”recover”/>
    </Goal>
    <Goal area=”B” id=”goal3”>
      <Search nextOnSuccess=”goal4” nextOnFailure=”goal4”/>
    </Goal>
    <Goal area=”C” id=”goal4”>
      <Rendezvous nextOnSuccess=”recover” nextOnFailure=”recover”/>
    </Goal>
    <Goal area=”recoveryPosition” id=”recover”>
      <Transit nextOnSuccess=”missionComplete”
          nextOnFailure=”missionAbort”/>
    </Goal>
  </GoalSet>
</UUVMission>
```

AVCL is readable by human or robot, captures logic of mission tasking

XML ensures syntactically correct, well-defined, numerically valid

Needed: semantic representation to check ethical, logical consistency

Mission clarity for humans – and robots

- Simplicity of success, failure, and (rare) exception outcomes encourages well-defined tasks and unambiguous, measurable criteria for continuation.

Confirmable beforehand: can a tactical officer (or commanding officer) review such a mission and then confidently say

  - "yes I understand and approve this human-robot mission" or, equivalently,
  - "yes I understand this mission and my team can carry it out themselves."

Converse:

- if an officer can't fully review/understand/approve such a mission, then likely it is **ill-defined** and needs further clarification anyway.

Added benefit: missions that are clearly readable/runnable by humans and robots can be further composed and checked by C2 planning tools to test for group operational-space management, avoiding mutual interference, etc.

Wrong question, right question

Wrong question to ask first when planning a tactical operation:

- "What are my *robots* doing out there?"

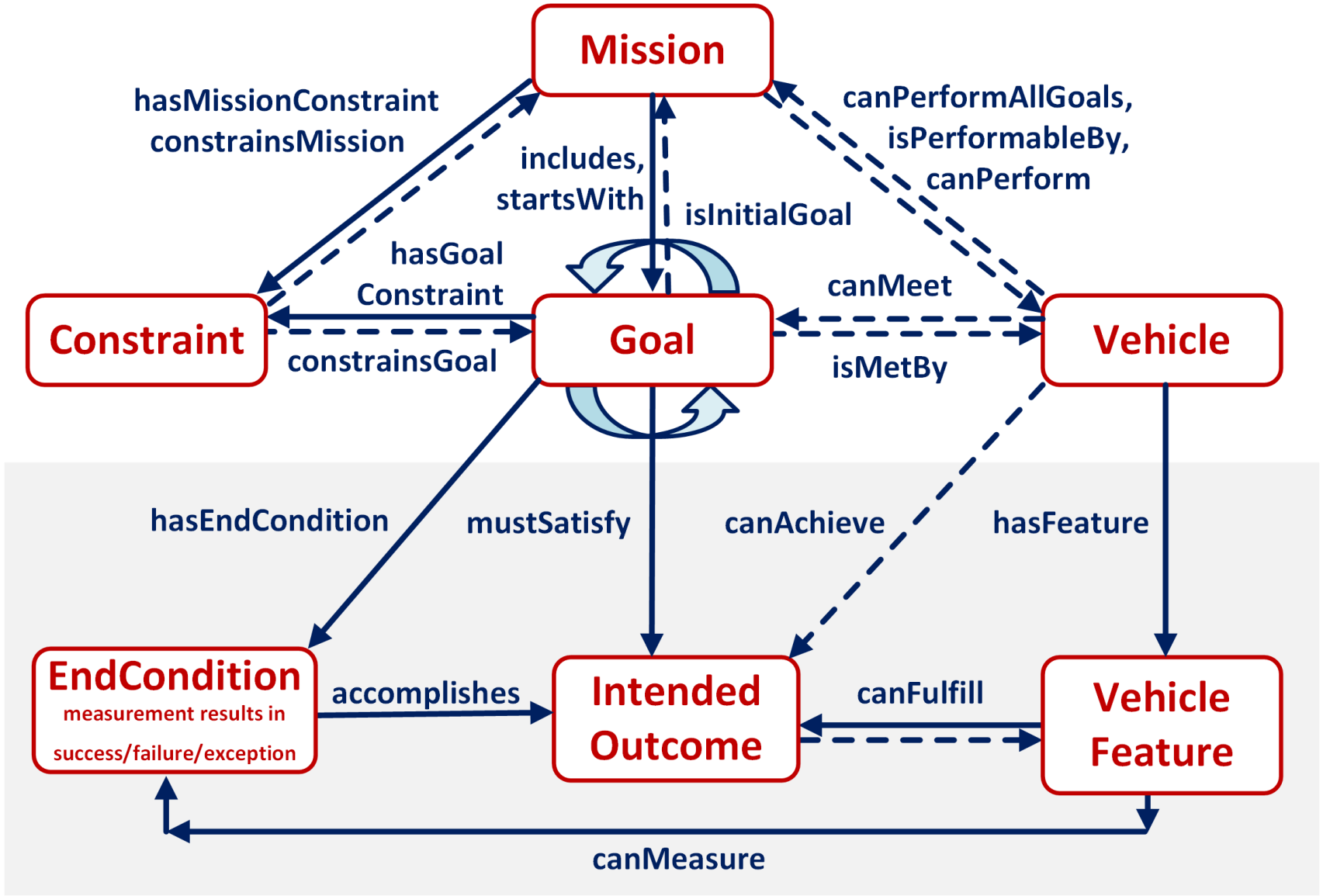
Right question to ask first when planning a tactical operation:

- "What is my *human-robot team* doing out there?"

Human-robot team mission has to be understood first!

- Robots complement humans, who must remain in charge throughout.
- If you don't have an OODA loop, you don't have a competent plan.

Mission Execution Ontology (MEO) 3.0



Asserted = explicitly declared in AVCL mission

Inferred = implicit by semantic relationships

Sailor Overboard Mission: Description

Purpose

- Life saving: single unmanned air/surface vehicle actions to complement human responses when performing *“SAILOR OVERBOARD”* operations.
- Carried out in direct concert with formal *shipboard emergency procedures*.
- Multiple UAVs/USVs might be employed in parallel with ships and aircraft, avoid mutual interference by each following deconflicted mission orders.

Phases

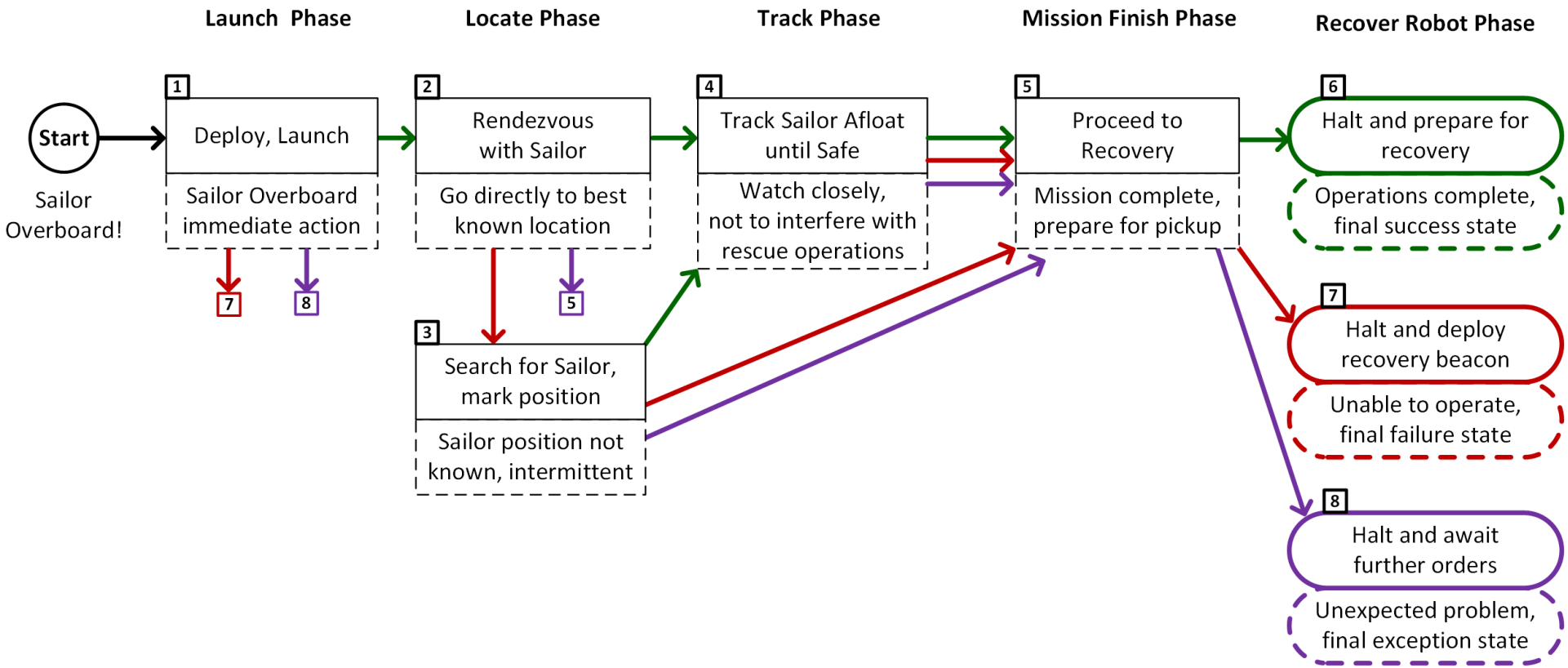
- Deploy/Launch, Rendezvous, Track Sailor until Safe, Return/Recovery.

Human Supervisory Role

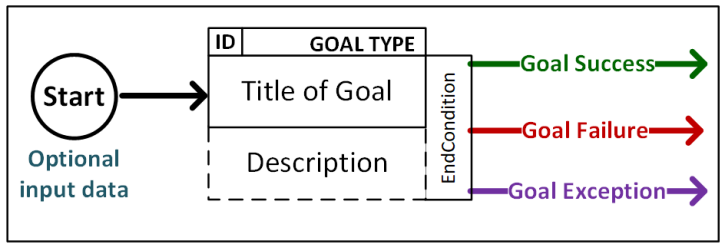
- Order standoff if interfering, manual control is possible due to proximity, can communicate to sailor via loudspeaker or beacon light.

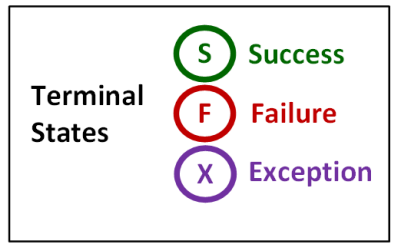
Sailor Overboard, 8 Phases – Mission Execution Automaton (MEA)

Single unmanned air/surface vehicle actions to complement human response when performing "**SAILOR OVERBOARD**" operations, carried out in concert with **shipboard emergency procedures**. Multiple UAVs/USVs can be employed in parallel with ships/aircraft, each following mission orders.



Legend





Don Brutzman and Bob McGhee
Mission upgrade 19 NOV 2019

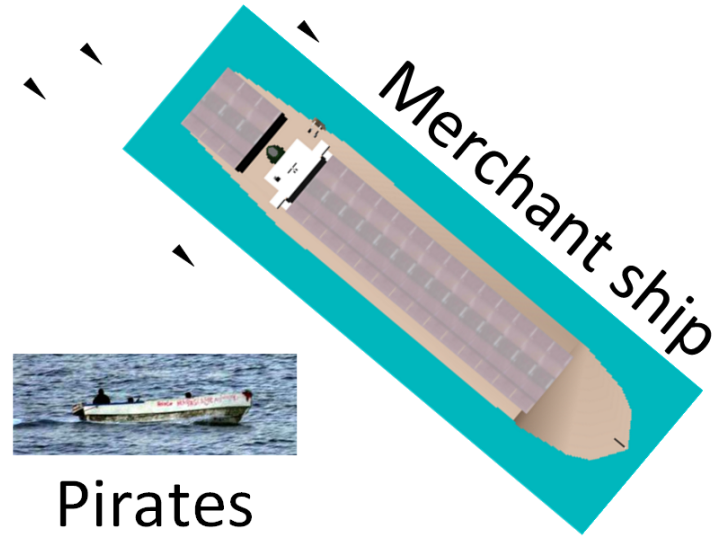
Life boat

Life-saving force: locate, track, communicate, beacon

Ethical control of unmanned systems is required for both lethal and lifesaving force if remote robots communicate intermittently, operating across lengthy time and distance.

Response dilemma for U.S. Navy ship

Lethal force: locate, warn, defend, threaten, attack



Pirates

Respond to one or both scenarios with USV/UAV assets to establish on-scene visibility and presence

OODA Loops for Ethical Control Canonical Missions

Ethical Control OODA Loops	Observe	Orient	Decide	Act
Sailor Overboard	Find Sailor	Report status	Avoid interference	Track sailor until rescued or relieved
Lifeboat Rescue	Find Lifeboat	Report status	Two-way communication	Track life raft until relieved
Pirate Seizure of Merchant Ship	Find merchant ship, pirate small boats	Identity Friend Foe Neutral Unknown (IFFNU) Issue warnings	Human commander authorization to use lethal force	Attack to defend ship if provoked, stay with merchant
Hospital Ship Swarm Attack	EM threat signals detected	(no orientation step in Sense Decide Act)	Reflex-response weapons attack	Mistaken attack on friendly = war crime
Hospital Ship Defense detects spoofing anti-pattern	EM threat signals detected	IFFNU including correlation	Human requirement for lethal force unmet, attack avoided	Report threat alert, commence search for hostile actors

Unmanned Maritime Autonomy Architecture (UMAA)

Richard R. Burgess, "Navy Requests Information for Unmanned Maritime Autonomy Architecture," SEA POWER, 20 FEB 19

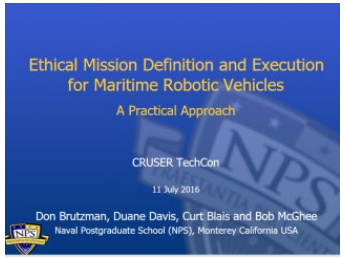
- "The intent of UMAA is to provide overarching standards that various UUVs and USVs can be built to in order to avoid creating multiple conflicting systems in the future"
- "The UMAA is being established to enable autonomy commonality and reduce acquisition costs across both surface and undersea unmanned vehicles."
- Topics of interest include Situational Awareness, Sensor and Effector Management, Processing Management, Communications Management, Vehicle Maneuver Management, Vehicle Engineering Management, Vehicle Computing Management, Support Operations

Multiple public NAVSEA documents refer to autonomy efforts and UMAA.

- NAVSEA PMS 406 is Program Office for Unmanned Maritime Systems
- NAVSEA Fact Sheet: Unmanned Maritime Systems Program Office (PMS 406) https://www.navsea.navy.mil/Portals/103/Documents/Exhibits/SNA2019/UnmannedMaritimeSys-Small.pdf
- Automated Management of Maritime Navigation Safety
  Navy SBIR 2020.1 - Topic N201-059, https://www.navysbir.com/n20_1/N201-059.htm

Proposed Critical Path Forward

Key Insights regarding Human Ethical Control



1. **Humans in military units are able to deal with moral challenges without ethical quandaries,**
   - **by using formally qualified experience, and by following mission orders that comply with Rules of Engagement (ROE) and Laws of Armed Conflict (LOAC).**

2. Ethical behaviors don't define the mission plan. Instead, **ethical constraints inform the mission plan.**

3. **Naval forces can only command mission orders that are**
   - **Understandable by (legally culpable) humans, then**
   - **Reliably and safely executed by robots.**

Reference: CRUSER TechCon Overview 2016
https://gitlab.nps.edu/Savage/EthicalControl/tree/master/documents/presentations

Assessment of Current Thinking

- Human supervision of potentially lethal autonomous systems is a matter of serious global importance.
- Wide consensus is emerging on principles, aspects of the problem, elements of solutions, and need to achieve better capabilities.
- Much philosophical concern, flurry of top-level interest in AI ethics, but few concrete activities are yet evident… hoping to change that.
- Cross disciplinary engagement: establishing an NPS Ethics Center can help

Ethical Control of Unmanned Systems project appears to provide a needed path towards practice, with the historic role of warfighting professionals more central than ever as weapons autonomy grows.

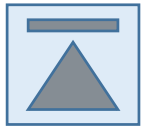
Possible domains and thesis topics

- Plain language and Semantic Web mission representations
  - Develop scenario for Warfare Innovation Continuum (WIC)
- Modeling, Simulation, Virtual Environments
  - UMAA RAIL SecDevOps, SPIDERS3D Virtual Sand Table
- Expeditionary missions, tri-service strategy USN-USMC-USCG
  - Show military impact, compare with ethical control and without
- Legal relationships
  - DoDD 3000.09, NATO Targeting guidance, Laws of Armed Conflict (LOAC), Law of Sea (UNCLOS), International Humanitarian Law (IHL)
- Your question here ??

Conclusions

- Human supervision is required for any unmanned systems holding potential for lethal force.
  - Cannot push “big red shiny AI button” and hope for best – immoral, unlawful.
  - Similar imperatives exist for supervising systems holding life-saving potential.
- Human control of unmanned systems is possible at long ranges of time-duration and distance through well-defined mission orders.
  - Readable and sharable by both humans and unmanned systems.
  - Validatable syntax and semantics through understandable logical constraints.
  - Testable and confirmable using simulation, visualization, perhaps qualification.
- Coherent human-system team approach is feasible and repeatable.
  - Semantic Web confirmation can ensure orders are comprehensive, consistent.
  - Human role remains essential for life-saving and potentially lethal scenarios.

Contact



Don Brutzman

brutzman@nps.edu
http://faculty.nps.edu/brutzman

Code USW/Br, Naval Postgraduate School
Monterey California 93943-5000 USA
1.831.656.2149 work
1.831.402.4809 cell

ETHICAL MISSION DEFINITION AND EXECUTION FOR MARITIME ROBOTS UNDER HUMAN SUPERVISION

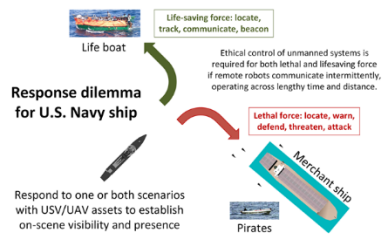
- Lethality requires ethical and legal basis, supervised by military teams.
- Executable robot tasking can resemble tactical tasking of humans afloat.
- Careful application of goal constraints makes ethical control feasible.
- Robot missions then complement and extend naval operation orders.
- Semantic Web logic can confirm ethical correctness and completeness.
- Next steps: continue 2 decades of work with realistic scenario testing.

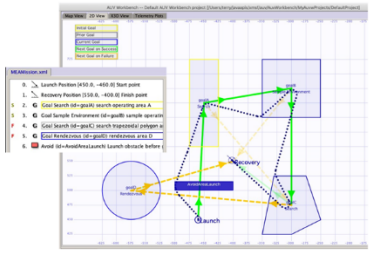
> "Ethical constraints on robot mission execution are possible today. There is no need to wait for future developments in Artificial Intelligence (AI). It is a moral imperative that ethical constraints in some form be introduced immediately into the software of all robots that are capable of inflicting unintended or deliberate harm to humans or property." Robert McGhee, April 2016

- IEEE Journal of Oceanic Engineering (JOE) paper along with online references.
- Authors Don Brutzman, Curtis Blais, Duane Davis and Robert McGhee, NPS.
- Feedback and recommendations always welcome. Contact: brutzman@nps.edu

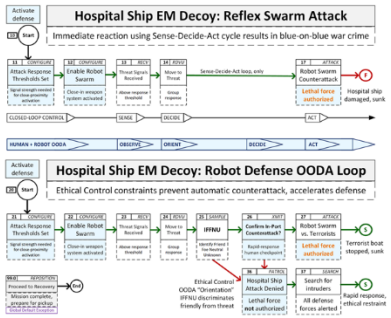
Ethical Control of Unmanned Systems:
Keeping Warfighters in Charge of Autonomy

CRUSER
Consortium for Robotics and Unmanned Systems Education and Research





Mission Execution Ontology (MEO) evolving



Milestones and Transitions

- CRUSER development led to first project selection under CRADA with Raytheon Missile Systems (RMS).
- Successful progress on test missions entering TRL 5 with simulation and Web-sharable 3D visualization.
- Expressing multiple robot mission plans consistently, coherently for diverse UAV, USV, UUV platforms.
- Use Semantic Web Standards to support warfighters.
- Evaluate NAVSEA Unmanned Maritime Autonomy Architecture (UMAA) evolution for robot qualification.

Why / Objectives

- Ethical control of unmanned systems can be accomplished through structured mission definitions that are trusted, consistently readable, validatable, repeatable and understandable by humans and robots.
- Orders must be lawful. Unmanned systems must behave ethically and comprehensibly if they are to support manned military units effectively.
- Well-structured mission orders can be tested and trusted to give human commanders confidence that offboard systems *will do what they are told to do*, and further *will not do what they are forbidden to do*.
- Demonstrate that no technological limitations exist that prevent applying the same kind of ethical constraints on robots and unmanned vehicles that already apply to humans, in lethal and life-saving scenarios.

https://savage.nps.edu/EthicalControl

What / Deliverables

- Update Mission Execution Ontology (MEO) concepts demonstrated in tests and simulation, building to perform field experimentation (FX).
- Supervise thesis work to explore canonical exemplar missions that are expected to utilize unmanned systems, looking across the full range of Naval warfare communities. Example scenarios include UAV for sailor overboard, UAV for refugee/lifeboat escort, and adept scouts. All must observe Law of Armed Conflict (LOAC), Rules of Engagement (ROE), and moral guidance of commanders despite long durations/distances.
- Define, simulate, and test combination of real-world goals and ethical constraints to robot mission tasking across set of canonical scenarios.
- Illustrate how human-robot teams meet moral and legal requirements if deploying unmanned systems with potential for lethal, life-saving force.

NPS PRAESTANTIA PER SCIENTIAM

Building on Concept Demonstrations

Principal Investigator: Don Brutzman brutzman@nps.edu 1.831.656.2149
Co-Investigator: Curtis Blais clblais@nps.edu 1.831.656.3215