Deep Models and Artificial Intelligence for Military Applications:
AAAI Technical Report FS-17-03

# Position Paper: Rational Behavior Model (RBM) and Human-Robot Ethical Constraints Using Mission Execution Ontology (MEO)

**Don Brutzman, Curtis Blais, Robert McGhee, Duane Davis**
Naval Postgraduate School (NPS), Monterey California USA 93943
brutzman@nps.edu clblais@nps.edu

**Abstract**

Autonomous systems can be ethically supervised by humans without constant communications. Adding constraints such as no-fly zones, time limitations, permission prerequisites etc. to mission orders allows operators to legally and ethically control mobile systems that have the potential for deliberate (or unintentional) lethal force. Ethical control can be practically achieved by providing parsable (and ethically validatable) orders to diverse unmanned systems.

## Ethical Control of Unmanned Systems

The authors have been engaged for several decades in academic research and military service relating to mission specification and execution for submarines, aircraft, and ground vehicles, both manned and robotic. We have increasingly focused on maritime robotics, especially with regard to execution of unmanned underwater vehicle missions under varying degrees of human oversight and ethical constraints. This work has included numerous simulations and also deployed experiments (Brutzman et al. 1998; Brutzman, McGhee, and Davis 2012; Brutzman, Davis, Blais, and McGhee 2016).

Common conclusions that treat ethical robots as an always-amoral philosophical conundrum or requiring undemonstrated morality-based artificial intelligence (AI) are simply not sensible or repeatable. For better or worse, actors around the world are rapidly designing and deploying mobile unmanned systems to augment human capabilities. Thus theory must meet practice. This contribution describes how mission orders can be specified in forms readable by both human operators and robot systems, including both syntactic validation of correctness and semantic validation of logical coherence. This approach has the potential to meet the moral requirements of international laws regarding human responsibility during armed conflict when unmanned systems are deployed.

The path to reach our present point of view has come from understanding that the Rational Behavior Model (RBM) software architecture used in our work has broader applicability than previously realized. This is because RBM depends upon a software construct called a Mission Execution Automaton (MEA) that is an extension of a Turing machine (TM) allowing the incorporation of arbitrary external agents. Such agents are presumed to be capable of interaction with their environment and returning one of a predetermined set of values to the finite state machine (FSM) portion of the MEA. Ternary logic is helpful, where allowable values returned from a query/command to an external agent have been limited to the set [success, failure, constraint]. Applied to a multiphase mission, concise mission-branch definition simplifies execution logic needed by remote systems. When a value constraint is returned, it can be taken to mean either that the current environment makes it unethical to complete the current phase, that a phase timeout has occurred, or that success or failure of the current phase cannot be determined, etc. Exception-handling steps can then occur.

Recent work has reached the next level, defining a Mission Execution Ontology (MEO) implemented in RDF/OWL that relates mission goals, task prerequisites and operating constraints with vehicle capabilities. MEO validation can be performed as part of tasking robotic or human external agents. This means that human-understandable orders to unmanned systems can further be semantically validated for logical correctness, ensuring that tasking of remote systems meets the same level of ethical rigor expected in human-to human orders. Confirmed validation makes authority meaningful for responsible humans.

In three-level robot architectures such as RBM, existing vehicle commands/missions can be incorporated as behaviors subject to overall regulation by rational (finite state) supervision of a mission-definition graph. Such tasking can be understood by qualified humans who are not computer professionals, much in the same way that military mission orders written in structured human language are understandable. Well-specified definitions of mission tasks and constraints can also be performed by a wide variety of diverse robot control code. With such a common understanding, it is possible to assign legal and moral responsibility for correct and ethical mission definition to a single (legally culpable) human individual. We believe that this kind of strong accountability is essential to military accountability, and may eventually become relevant for emerging robotic technologies affecting public safety.

Algorithms cannot replace human responsibility. Even so, a fully testable technology (such as that provided by the MEA and MEO formalisms) allows for the assignment of human accountability. Specifically, the MEA provides a mathematically rigorous mechanism for mission definition and execution as an exhaustively testable flow diagram. This approach ensures that accountable operators can fully understand all high-level task sequences before authorizing robot operations. The MEO employs description logics (DLs) and Semantic Web technologies to provide strong assurances that MEA mission definitions are semantically correct and fully executable by specific target vehicles. By applying the best strengths of human ethical responsibility, repeatable formal logic and directable unmanned systems together, these capabilities provide a practical framework for ethically grounded human supervision of unmanned systems.

## References


Brutzman, D., Healey, T., Marco, D., and McGhee, R. B. 1998. The Phoenix Autonomous Underwater Vehicle. *AI-Based Mobile Robots*. D. Kortenkamp, R. P. Bonasso, and R. Murphy, eds. Cambridge, Mass: AAAI Press: 323-360.

Brutzman, D. P., McGhee R. B., and Davis, D. T. 2012. An Implemented Universal Mission Controller with Run Time Ethics Checking for Autonomous Unmanned Vehicles--a UUV Example. In Proceedings of the IEEE Oceanic Engineering Society (OES) Autonomous Underwater Vehicles. Piscataway Township, NJ: Institute for Electrical and Electronic Engineers.

Brutzman, D. P., Davis, D. T., Blais, C. L., and McGhee, R. B. 2016. Ethical Mission Definition and Execution for Maritime and Naval Robotic Vehicles: A Practical Approach. In Proceedings of the OCEANS 2016 Conference. Piscataway Township, NJ: Institute for Electrical and Electronic Engineers.

Brutzman, D. Autonomous Unmanned Vehicle (AUV) Workbench, https://savage.nps.edu/AuvWorkbench

Brutzman, D. 2017. Ethical Control of Autonomous Unmanned Systems. *Network Optional Warfare (NOW) Blog*. https://wiki.nps.edu/display/NOW/2017/03/08/Ethical+Control+of+Autonomous+Unmanned+Systems