



CAS Special Session: ATTRACTOR

Autonomy Teaming and TRAjectories for Complex
Trusted Operational Reliability

Building a Basis of Certification for Trust and
Trustworthiness of Autonomous Systems:

Ownship Self-Assessment of Success for
Autonomous Search and Rescue Operations

B Danette Allen, PhD
NASA Langley Research Center

AIAA Aviation 2019

20 June 2019

Overarching Problem

UAM – a target environment



Environment:
Safety-critical
Time-critical
Certification-dependent

Unprecedented complexity



Density + heterogeneity + autonomy + non-cooperative agents $\Rightarrow$
 System complexity increases / uncertainty grows / safety decreases $\Rightarrow$
 Control must transition to human $\cup$ machine, with increasing machine authority
 Autonomous System $\equiv$ cyber-physical-human system with capacity for independent decision making and a degree of authority to act on these decisions;
Machine authority is a major source of uncertainty



Authority: Automation & Autonomy

Framework for the Design and Evaluation of Autonomous Systems

Level	Observe	Originate	Plan	Execute
8	The computer gathers, filters, and prioritizes data without displaying any information to the human.	The computer perceives and integrates data which is not displayed to the human.		
7	The computer gathers, filters, and prioritizes data without displaying any information to the human. Though, a "program functioning" flag is displayed.	The computer analyzes and interprets, and integrates result which is only human if result fits context (context displays summaries).		
6	The computer gathers, filters, and prioritizes information displayed to the human.	The computer overlays predictions with analysis and integrates the data.	The computer performs ranking tasks and displays a reduced set of ranked data.	Computer executes automatically; informs the human for contingencies.

Credit: Chad Frost, ARC

Automated Systems

- Follows a script; stops and waits in unanticipated situations

Autonomous Systems

- Resolves choices on its own, locally

"Intelligent" Autonomous Systems

- Capable of more creative solutions to ambiguous problems

- Fully preprogrammed and act repeatedly and independently of external influence or control

- Self-directed toward a goal governed by strategies that direct their behavior. directed in unpredictable situations.

Sheridan Scale

- Computer offers recommendations
- Computer offers advice
- Computer narrows choices
- Computer suggests actions
- Computer executes actions
- Computer allows human to override
- Computer executes human commands
- Computer informs human of status
- Computer informs human of intent
- Computer decides

Credit: MITRE

Human

Supervisory Control

Fully Automatic

Autonomy

We Are Here

Coact

Associate

Chasm

- Algorithms respond to input data without human input or possibility of Override
- Take predetermined Action
- Control Theory - Feedback loops

- Non-deterministic
- Judgment
- Emergent behavior
- Perception & Reasoning



Decision-Making

- **Literally every action at every scale** in the system (moving from point A to point B, taking direction from another agent,...) is a result of a decision-making process / algorithm, regardless of who/what makes the decision.
- **When can we trust a decision?**
 - When it is trustworthy
 - It is trustworthy when it leads to a predictable outcome
- **ATTRACTOR** is developing an understanding of decision trustworthiness in autonomous systems



ATTRACTOR

Autonomy, Teaming & TRAjectories for Complex Trusted Operational Reliability

Reliable Multi-agent System for
Mission Planning and Execution

=

Trustworthiness:

- Attribute of the cyber-physical system
- Requires context-driven metrics
- Achieved through *analyzable trajectories* and behaviors, supported by effective *teaming, learning* and *training*
- *Requires persistent M&S*

+

Trust:

- Attribute of the users/observers/controllers of the system
- Explainable AI (XAI) is critical in providing *clarity in decisions* and forensics, establishing *justifiable trust*, and defining dynamic roles and responsibilities

For human agents, natural interaction for mission planning and execution, supported by XAI, fosters trust and validates reliability

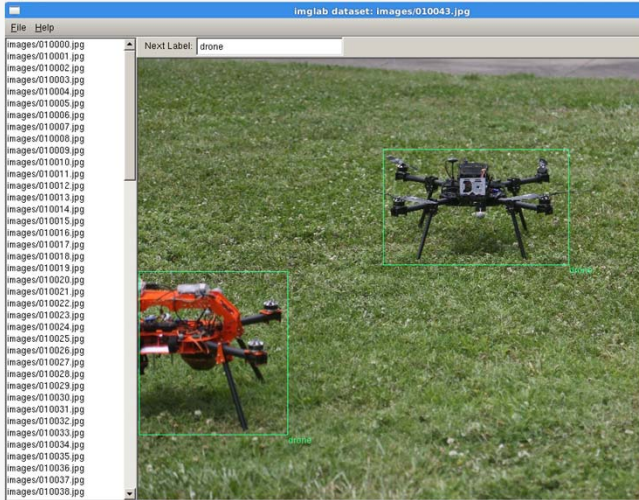
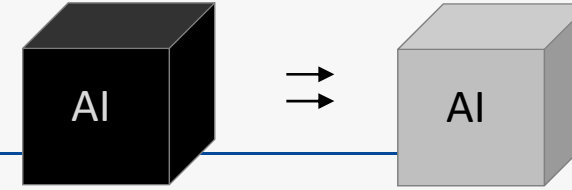


ATTRACTOR

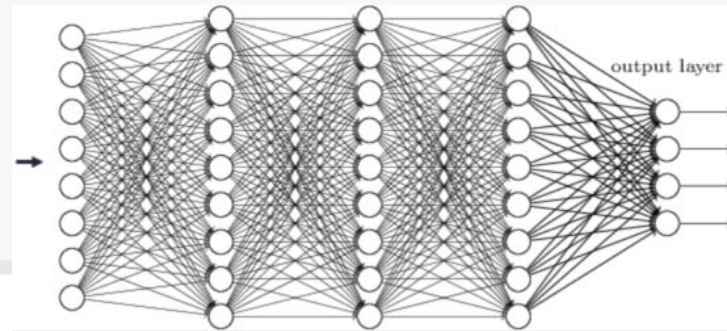
- The goal of ATTRACTOR is make progress towards building a basis of certification for autonomous systems
 - via establishing metrics for trustworthiness and trust
 - in single- and multi-agent team interactions
 - using eXplainable AI (XAI) and
 - and analyzable trajectories
- in the operational context of mission-critical planning and execution.
- A persistent modeling and simulation environment for a-priori, real-time and a-posteriori interaction for test and evaluation is the foundation upon which these goals rest.



eXplainable AI (XAI)



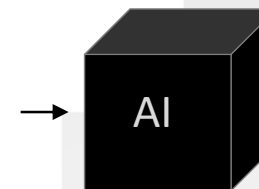
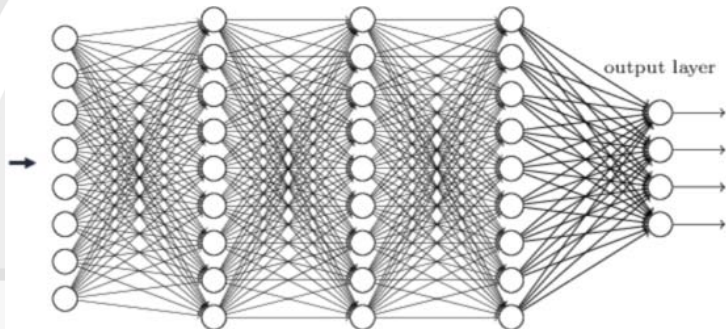
Machine Learning System



Two
Quads

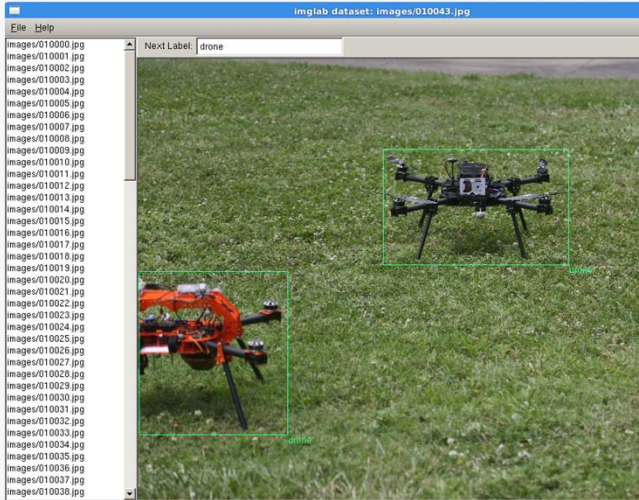
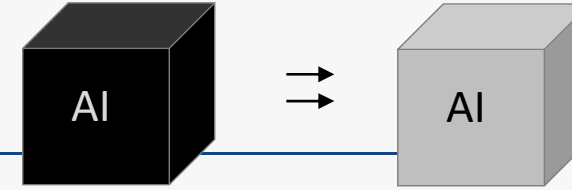
xAI explanation: These are two UAVs. They are quadrotors. They have four propellers. They have four legs (landing gear). Etc.

Transparency vs. a Posteriori Explicability

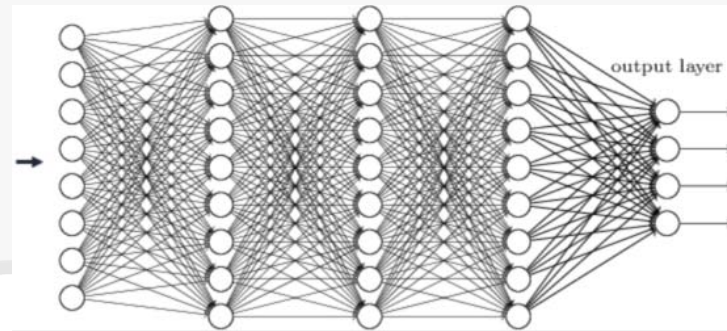


Current paradigm: Don't fly here
XAI: Don't fly here because risk is high because... OR ... because I don't know

eXplainable AI (XAI)



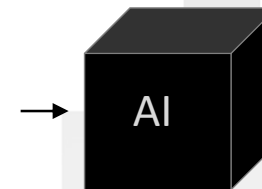
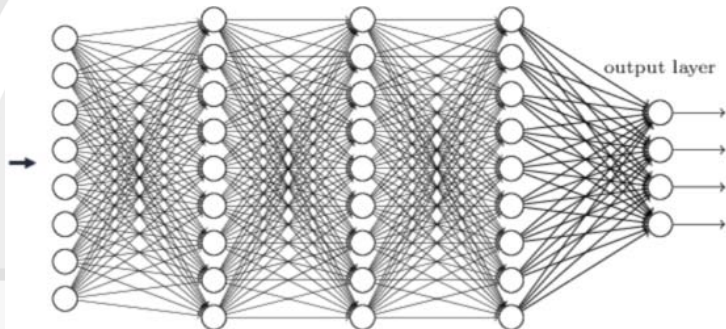
Machine Learning System



Two
Quads

xAI explanation: These are two UAVs. They are quadrotors. They have four propellers. They have four legs (landing gear). Etc.

Transparency vs. a Posteriori Explicability



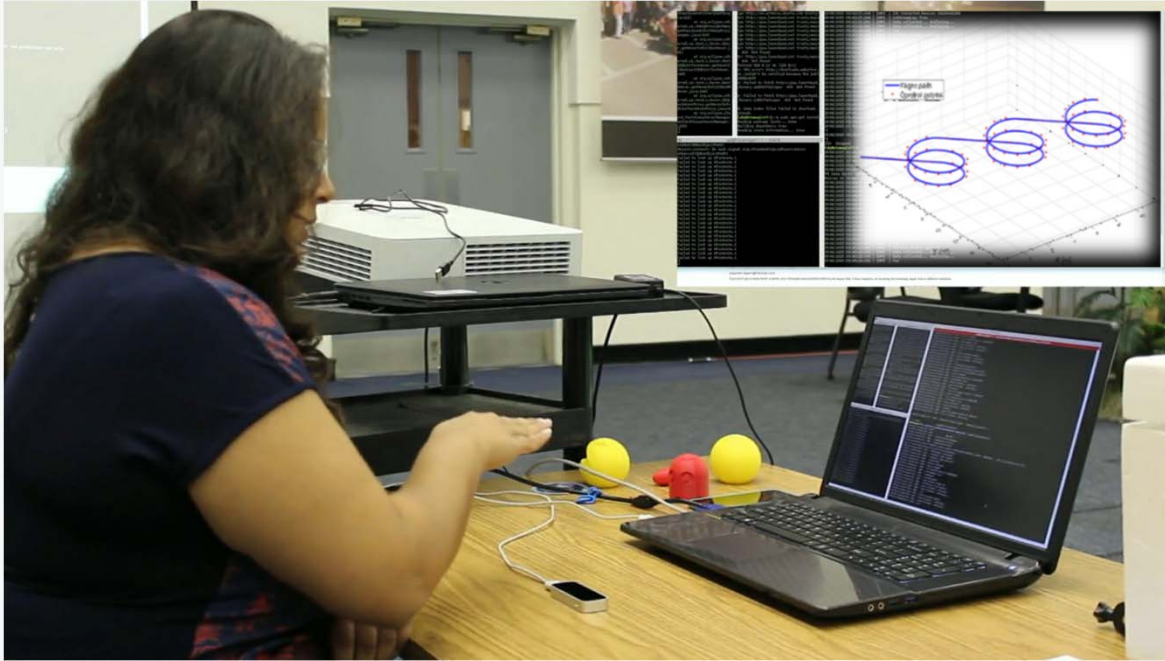
Current paradigm: Don't fly here

XAI: Don't fly here because risk is high because... OR ... because I don't know

Trust & Trustworthiness

- The How

- The What

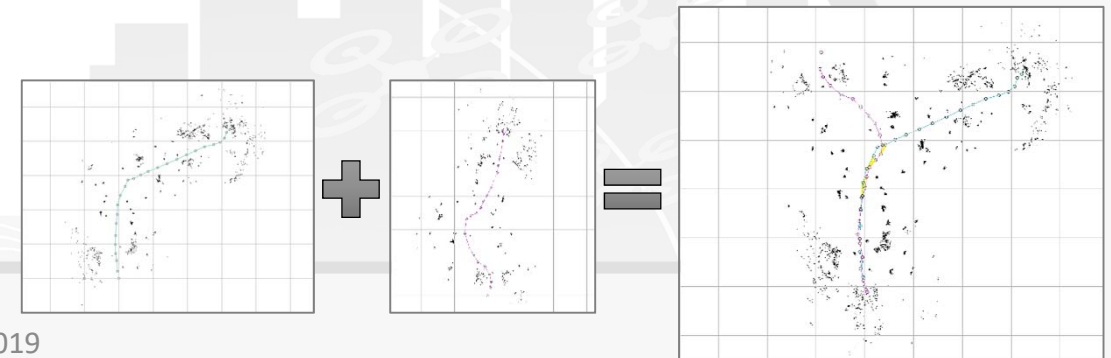


Natural Interaction



Search & Rescue Under The Canopy (w/ MIT)

“Put that there” [Bolt et al]
 “Fly this pattern”
 “Find xxx in this area”



Natural Interaction



Design Reference Mission (DRM)

- Search and Rescue (SAR) in challenging environments
 - Gives a focus for development
 - The platform is representative of all transport
 - The developed technologies are cross-cutting
 - SAR applications have intrinsic societal value
 - High complexity
- SAR Under the Canopy (SARUC)
 - An estimated 2000 people are lost in wooded areas each year (O(100) with fatal outcomes). Chances of survival diminish quickly. Topography and clutter limit the coverage and speed achievable by human search teams. The need for rapid SAR by autonomous vehicles grows.





ATTRACTOR DRM



A hiker is reported as lost, last seen at some location in the woods some hours ago.

A fleet of drones enters the forest to search for the hiker, using onboard LIDAR and IMUs to track their position and avoid obstacles.

The drones make a search plan based on how fast hikers can go, and where they are likely to go.

The drones periodically share a map of the forest whenever they have communication with each other



Required Capabilities

SARUC

Each vehicle must be able to:

- Estimate its position, attitude and velocity robustly from onboard sensors instead of GPS, since the tree canopy will interfere with GPS-based estimation;
- Locate the obstacles around it, such as trees and the forest floor;
- Decide where to search next and plan collision-free trajectories;
- Keep track of the vehicle energy state and maintain efficient egress paths;
- Share information and vehicle status with other vehicles, and subdivide the search and rescue tasks amongst the team appropriately.





SARUC



Exploration and Detection
March 2018

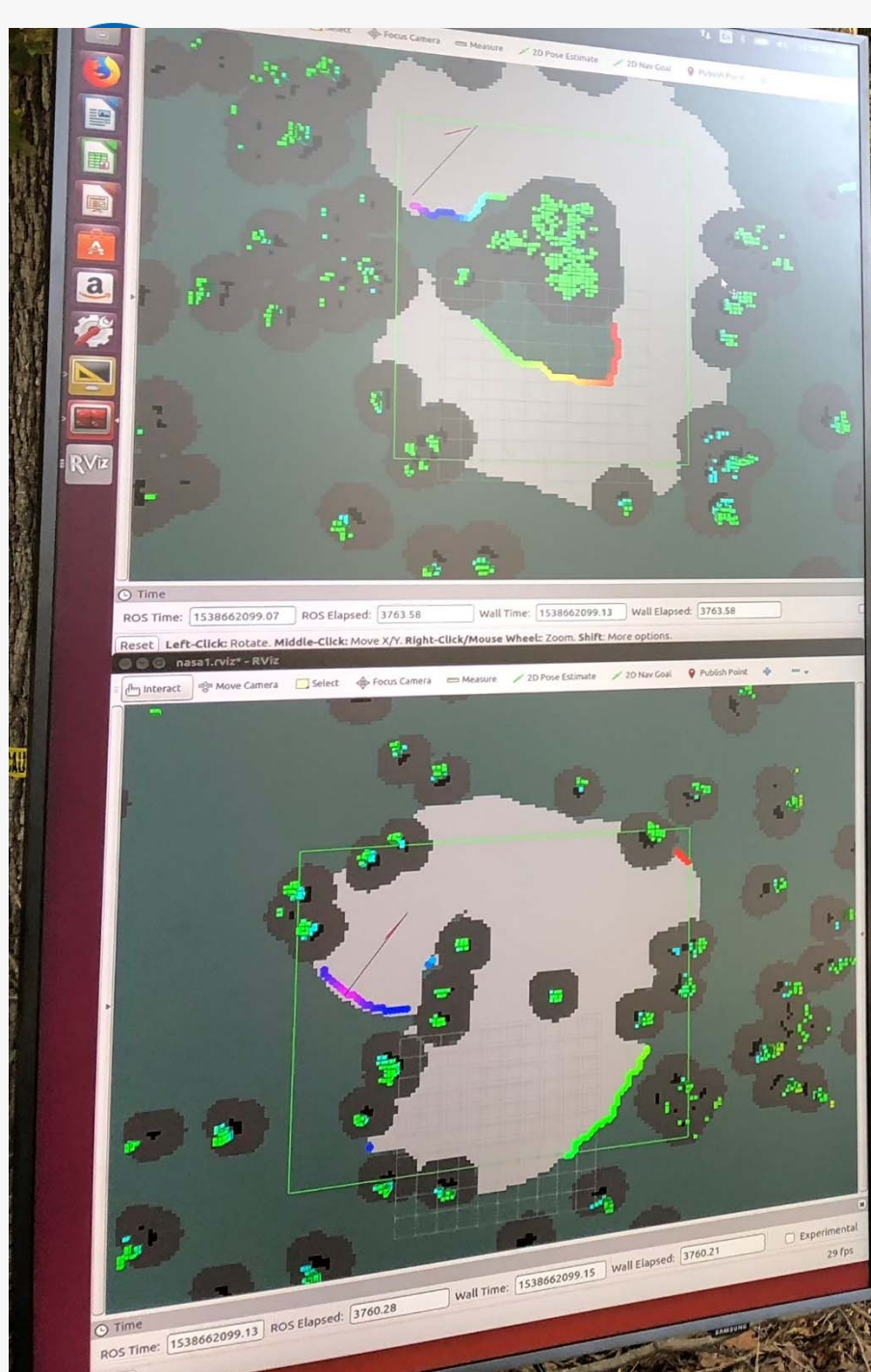
Collaboration with
Massachusetts Institute
of Technology (MIT)

Exploration and Map Closure
October 2017



SARUC Technologies



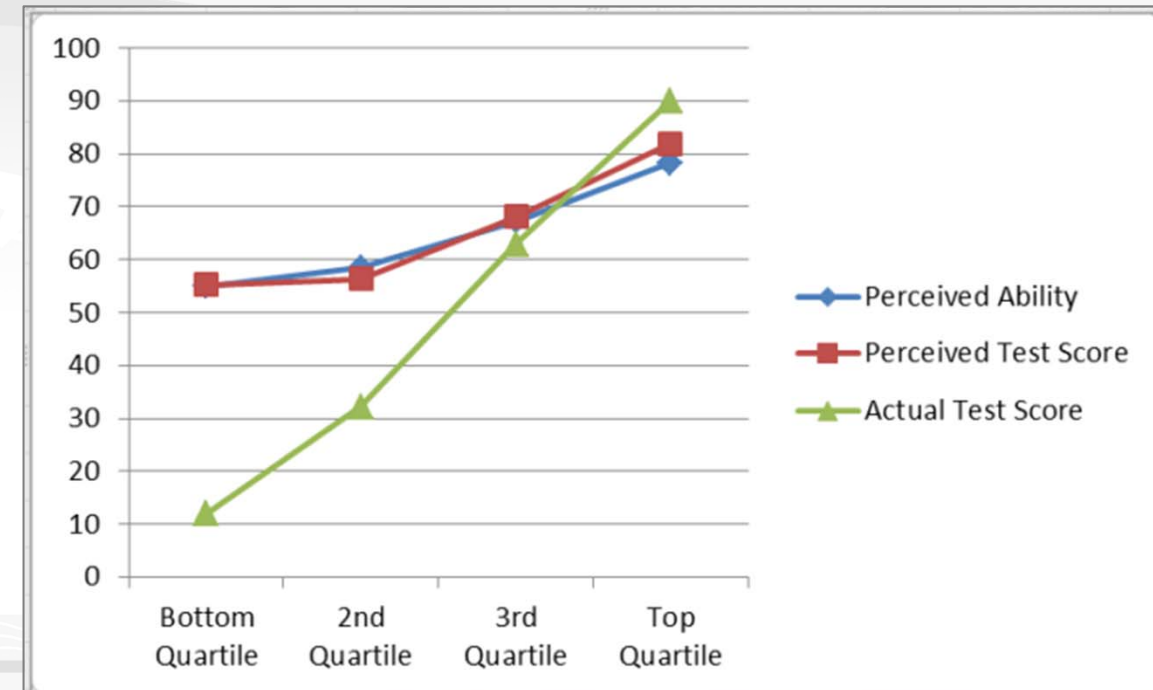


Self-Confidence

- In the field of psychology, the **Dunning–Kruger effect** is a cognitive bias in which people mistakenly assess their cognitive ability as greater than it is. It is related to the cognitive bias of illusory superiority and comes from the inability of people to recognize their lack of ability. Without the self-awareness of metacognition, people cannot objectively evaluate their competence or incompetence.

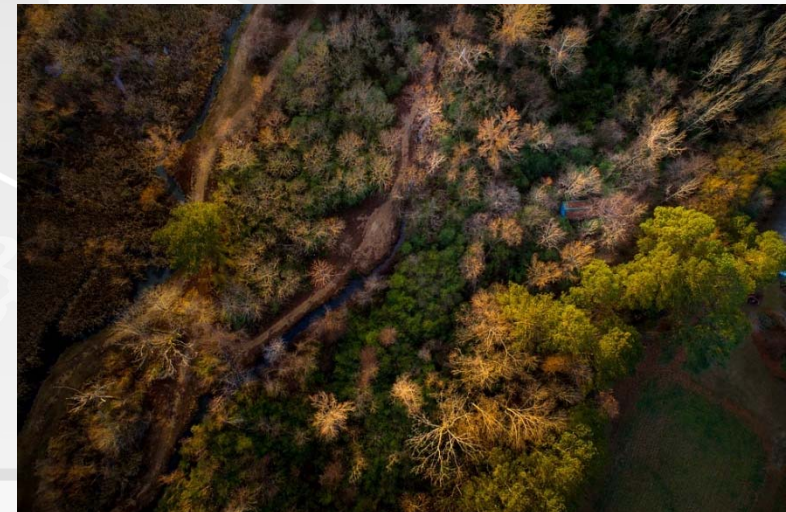
Kruger, Justin; Dunning, David (1999). "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments". *Journal of Personality and Social Psychology*. 77 (6): 1121–1134. CiteSeerX 10.1.1.64.2655. doi:10.1037/0022-3514.77.6.1121.

- Is this exclusive to people?
What about autonomous systems?
- Can a machine decide when it should or shouldn't take an action?



Ownship Self-Assessment

- What if that action is entering a wooded area?
Or continuing to search the path ahead?
- How can an agent self-assess its ability to proceed?
 - Clutter Treshold: Obstacles are relative
 - Size
 - Agility
 - Resilience
 - Clutter Assessment: Sensing and Perception
 - Edge Detection
 - Depth Map
- Multi-agent cooperation/collaboration





Mixed Reality SARUC



AUTONOMOUS
MULTI-MODAL
MULTI-AGENT
SEARCH & RESCUE

Questions?

- Check out our blog autonomyincubator.blogspot.com
- Join us on Twitter [@AutonomyIncub8r](https://twitter.com/AutonomyIncub8r)

