Assured Autonomy: Self Confidence in Autonomous Systems

Matt Aitken[1,*], Nisar Ahmed[2,*], Eric Frew[3,*], Dale Lawrence[4,*], Brian Argrow[5,*]

Research and Engineering Center for Unmanned Vehicles

Motivation

- Autonomy = many layers of **approximations**
- Decisions based on **incomplete/uncertain info**
- **Certification standards** for autonomy: must account for non-determinism, complexity, adaptability, etc.
- **Trust**: integral part of certification for autonomy

Trust and Self-Confidence

User Trust Model

- **Trust**: willingness to depend on autonomy
  - Depends on situation and initial disposition/beliefs
  - Affects the user's actions
- **Assurance**: autonomy's ability to affect trust
  - Positive or negative
  - Ultimately affects user's actions

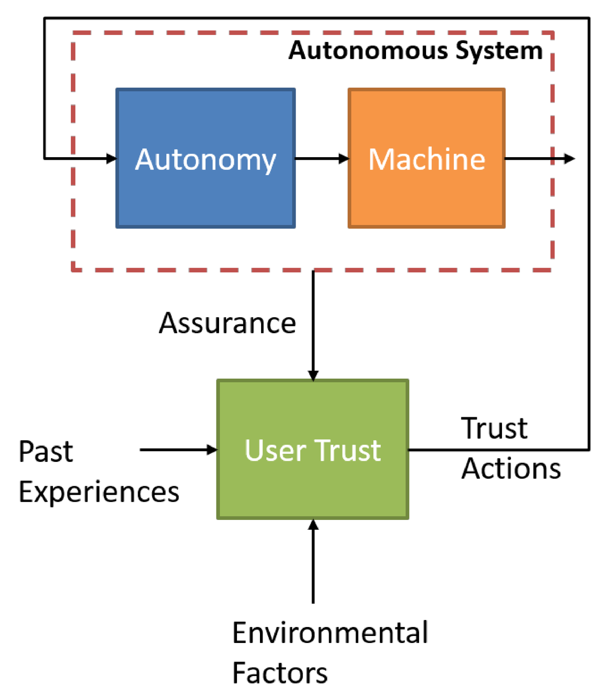

Fig. 1 – Trust Model

Self-Confidence as an Assurance

- Perceived ability to execute assigned tasks (within defined scope of autonomy) despite uncertainties in:
  - Its knowledge of world
  - Its knowledge of own self/state
  - Its reasoning and execution capabilities

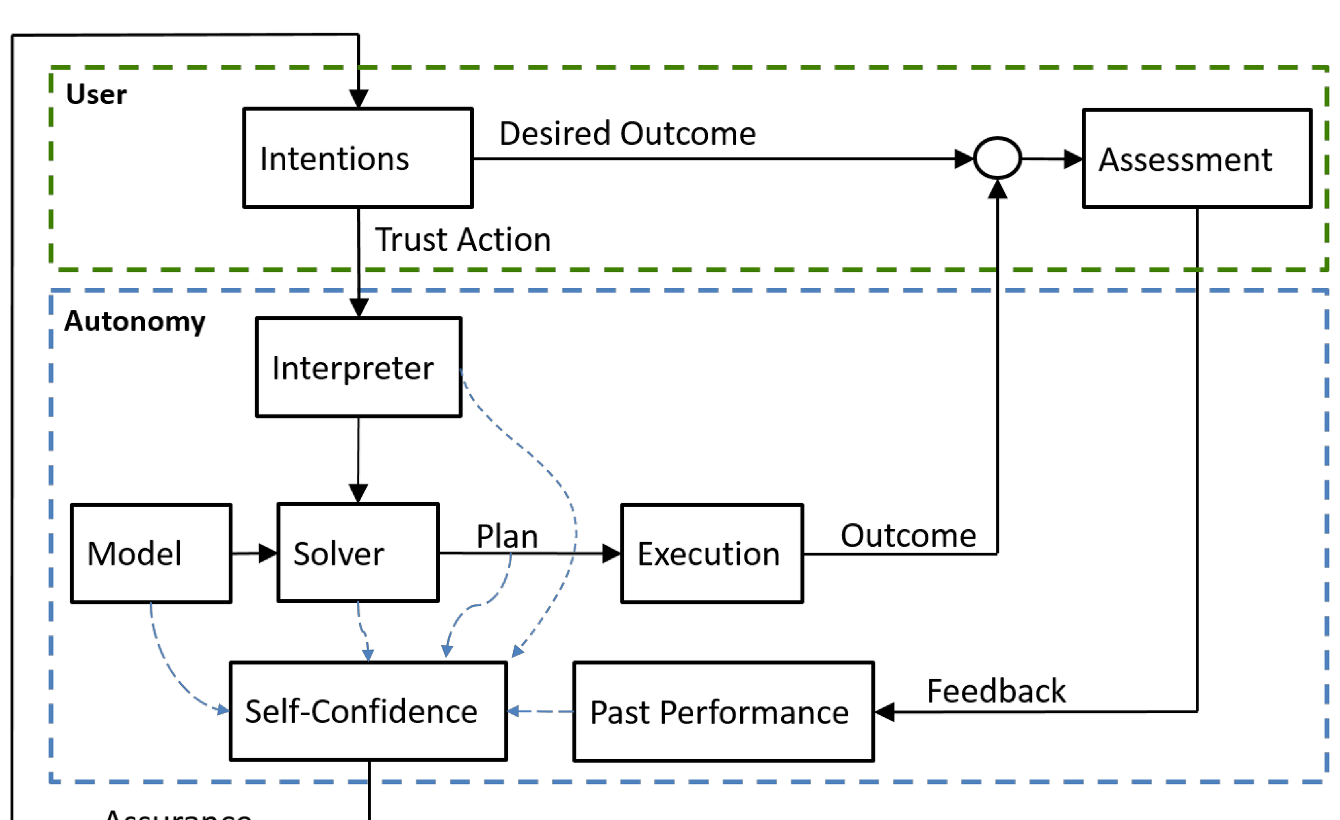

Fig. 2 – Self-Confidence Diagram

Scenario: Pursuit-Evasion on Road Network

Scenario Overview

- Unmanned ground vehicle (UGV) attempts to reach exit of a road network
- Pursuer attempts to capture the UGV
- UAV and unattended ground sensors (UGSs) gather data about pursuer's location
- Human supervisor interacts with UGV

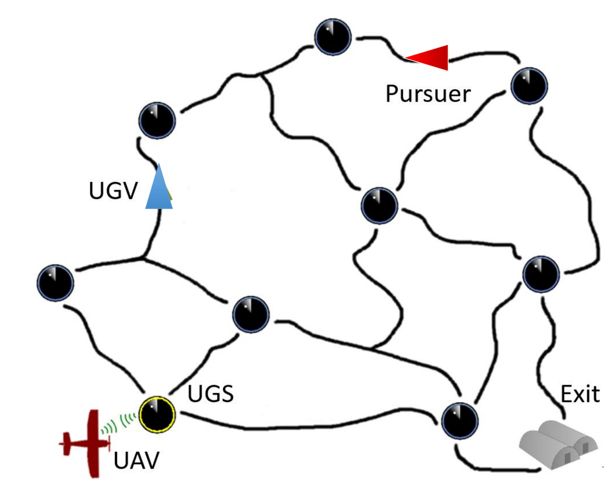

Fig. 3 – An example road network

Human-on-the-Loop

- No low level/expert system knowledge
- Comfortable with the problem
- Decisions influenced by trust
- Interrogate the autonomy
- Modify decision making via stance selection ('aggressive', 'defensive', etc.)
- Provide information, advice (identify intruder behavior, update map, etc.)

Fig. 4 – A UAS Operator

Mixed Observability Markov Decision Process

- States $\boldsymbol{s} = (\boldsymbol{x}, \boldsymbol{y})$: UGV and pursuer position
- Actions $\boldsymbol{a}$: valid UGV step directions
- Observations $\boldsymbol{o}$: $O = (o_1, o_2)$, $o_i \in \{\text{detect, no detect}\}$, $p(o_i|y)$ given
- Reward: $r(s,a) = \begin{cases} +1000 & \text{if } x = \text{exit} \wedge x \neq y, \\ -1000, & \text{if } x = y, \\ -1, & \text{o'wise} \end{cases}$
- Transition Probabilities: $p(y'|y),\ p(x'|x,a)$
- Belief $\boldsymbol{b}$: $b = [x, \vec{p}(y|O)]$
- Bayes' filter updates:

$$b' = \tau(b,a,O) \propto \sum_s p(s'|s,a)p(O|s')$$

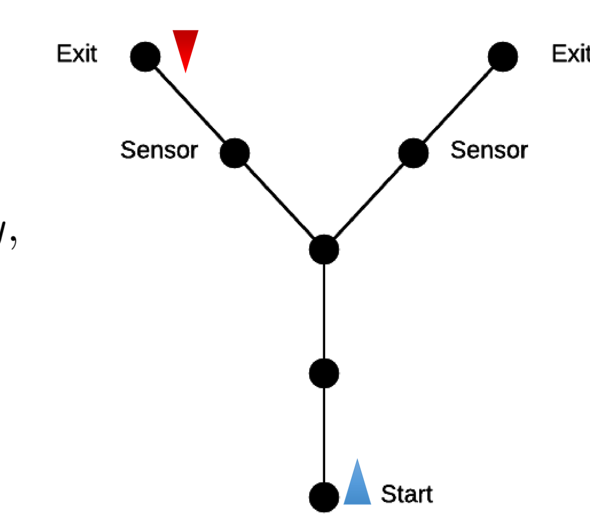

Fig. 5 – Simple Road Network Scenario

Policy Generation

- Goal: Maximize expected reward $\boldsymbol{V}$:

$$V^\pi(b_0) = \sum_{t=0}^{\infty} \gamma^t\, r(b_t, a_t) = \sum_{t=0}^{\infty} \gamma^t\, E[R(s_t, a_t)]$$

- Find optimal policy :

$$\pi^* = \underset{\pi}{\operatorname{argmax}}\ V^\pi(b_0) \ \text{ – solution intractable}$$

- Approximate using SARSOP and APPL software

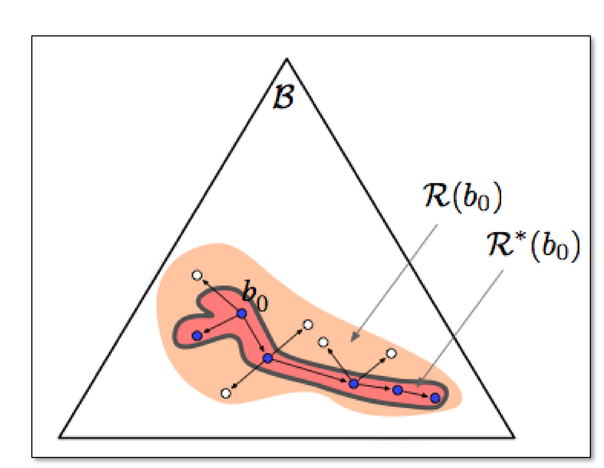

Fig. 6 – SARSOP algorithm [Kurniawati, et al 2008]

Self-Confidence Formulation

$$SC = f_{sc}(X_{sc}), \qquad X_{sc} = [x_1, x_2, x_3, x_4, x_5, \ldots]$$

- x_1 Command Interpretation
  - Are the autonomy and user 'on the same page'?
- x_2 Model Validity
  - How well does the autonomy's model reflect the real world?
- x_3 Solver Quality
  - How well can the solver use the model to generate policies?
- x_4 Outcome Assessment
  - How 'good' is the outcome distribution?
- x_5 Past Performance
  - How well has the autonomy done in similar circumstances?

Example Calculation: Outcome Assessment

Logistic UPM/LPM Metric

- Measure of 'goodness' of a reward distribution $p(reward|\pi)$

$$x_4 = \frac{1}{1 - e^{-k(\log(\frac{UPM}{LPM}))}}, \qquad UPM/LPM = \frac{\int_{r^*}^{\infty}(r - r^*)\,p(r)dr}{\int_{-\infty}^{r^*}(r^* - r)\,p(r)dr}$$

- r^* is the minimal acceptable reward, x_4 is a logistic function with steepness k applied to the log of the UPM/LPM ratio

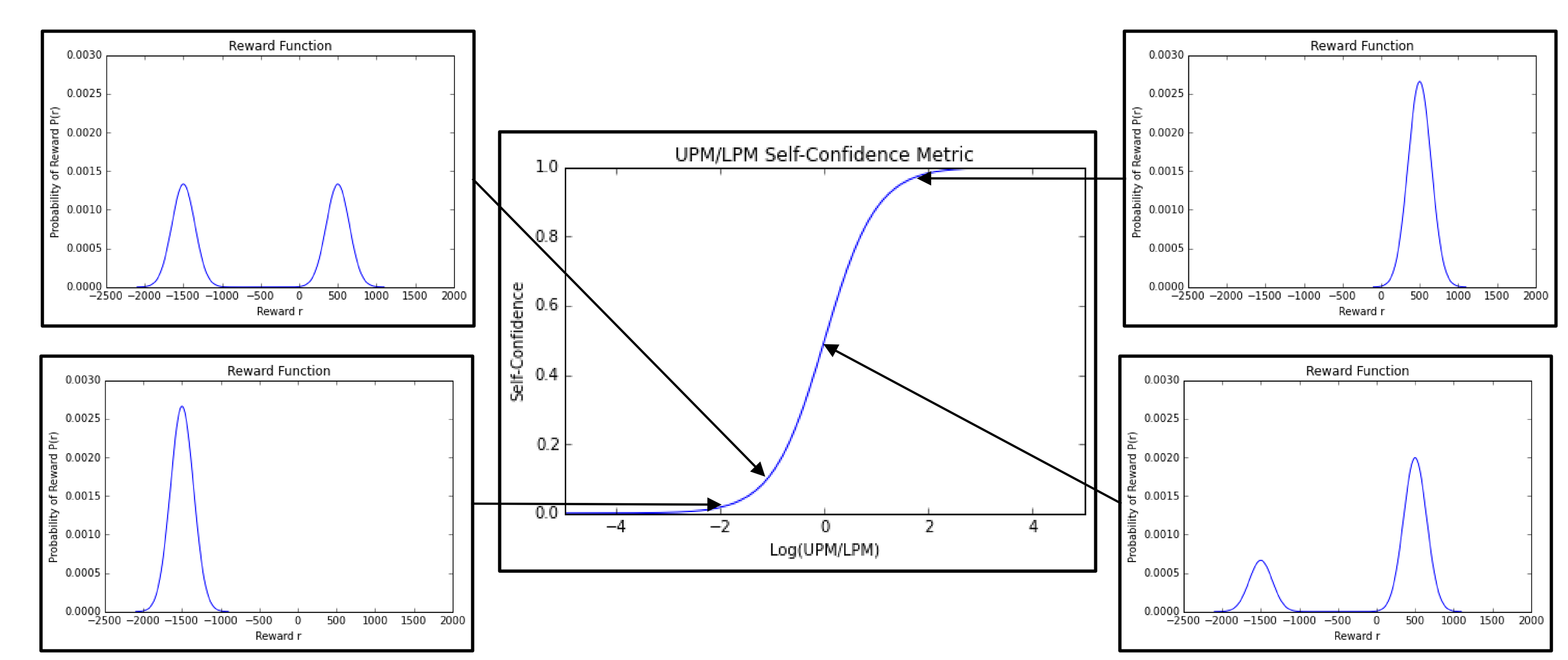

Fig. 7 – Outcome Assessment Self-Confidence ($r^* = 0, k = 1$)

Next Steps

- ROS/Gazebo scenario simulation
- Implement self-confidence reporting
- Design GUI and user study
  - First test: detect measurable difference in trust between users with and without self-confidence
- Run experimental user study

Fig. 8 – Husky from Clearpath Robotics

Research and Engineering Center for Unmanned Vehicles
UNIVERSITY OF COLORADO BOULDER

*Department of Aerospace Engineering Sciences
[1]matthew.s.aitken@colorado.edu [2]nisar.ahmed@colorado.edu [3]eric.frew@colorado.edu [4]dale.lawrence@colorado.edu [5]brian.argrow@colorado.edu