Uncertainty and Competency Awareness for Assured Human-Autonomous System Interaction

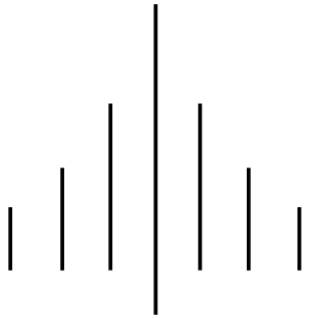
Nisar Ahmed
Assistant Professor

Ann and H.J. Smead Aerospace Engineering Sciences
University of Colorado at Boulder

2019 CCC Assured Autonomy Workshop

Arlington, VA

October 17, 2019



COHRINT
The Cooperative
Human-Robot Intelligence Lab

Technical Practices/Ideas/Attitudes That Create Issues

1. Human-autonomy interaction considered afterthought or band aid/last resort
   (i.e. the "regrettable but necessary evil" for assurance and safety…)

Why wait until it's too late? No built-in/by-design affordances?

2. Theorist/programmer/system designer need for "clean" I/O models of humans:

$$human.behavior = f(x;\theta)$$

(where .behavior is bounded, and specific f, x, θ will come from papers by very smart human factors people)

Uncertainty: context, task, environs, lack of data,… → theorems/algorithms/systems break

3. Intelligent competent machine = loner know-it-all (programmed with all right answers at start)

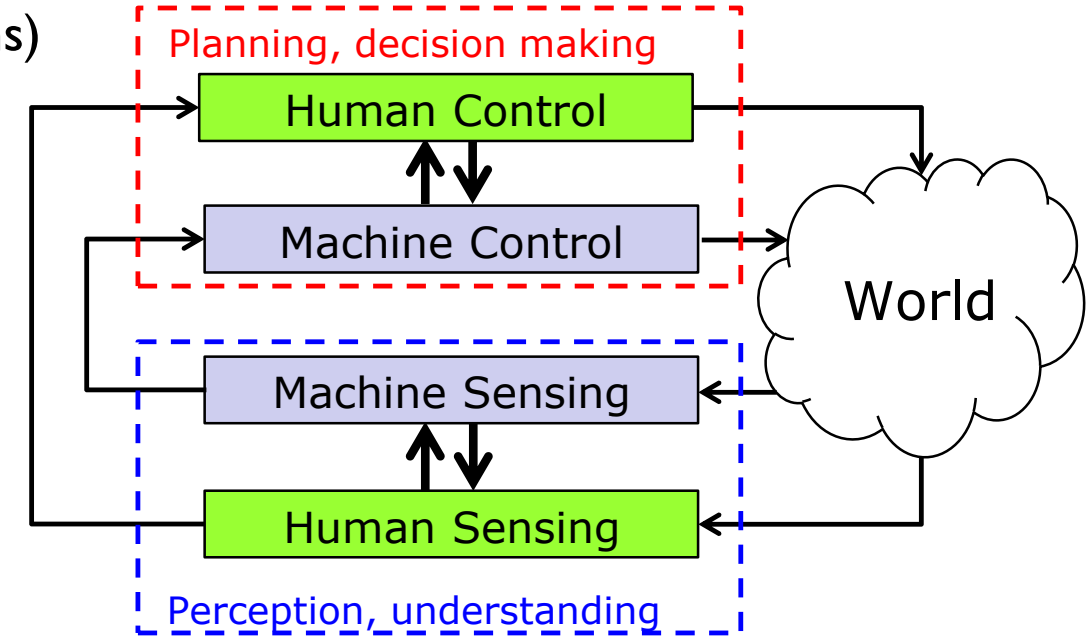
Smart competent (and autonomous) people ask questions of themselves and others
--- yet "smart autonomous machines" don't/can't?

Human-Machine Collaboration Under Uncertainty

- Example: use human sensing/perception to help machines make sense of dynamic world
  - fuse "hard data" (machines) with **"soft data"** (humans)

"Feedforward" commands [Arkin, et al 2015; Tellex, et al 2012; Huang, et al 2010]

"Feedback" reporting

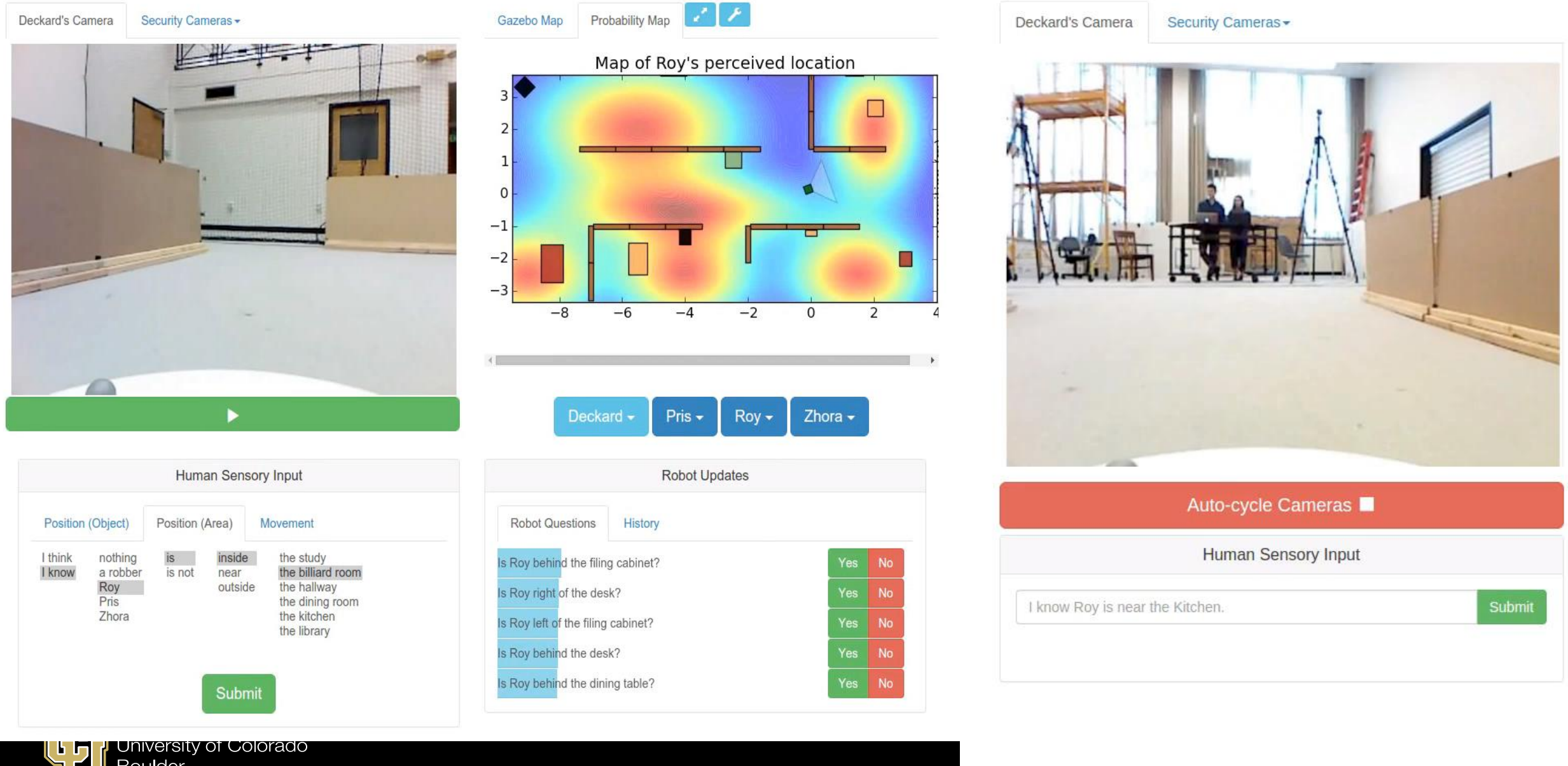


Humans as smart "soft sensors":

-**human perception already solved**: how to exploit?
-**"fill in gaps"** over large amounts of space, time
-**autonomous machines still decide** how to use/request info
-multi-layering, **multi-tasking** [Lewis, et al. 2009]

“Cops and Robots” Indoor Experimental Testbed

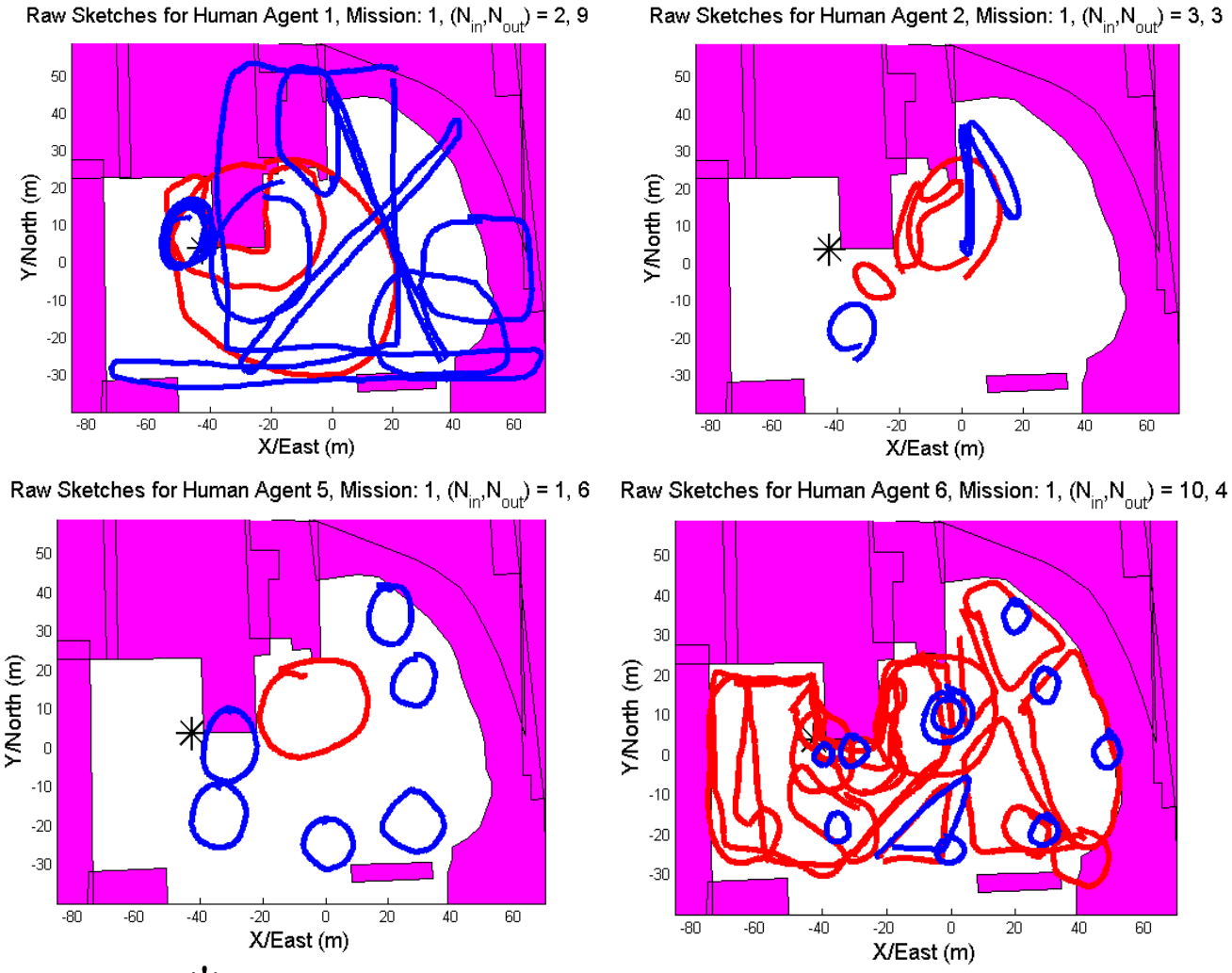
University of Colorado Boulder



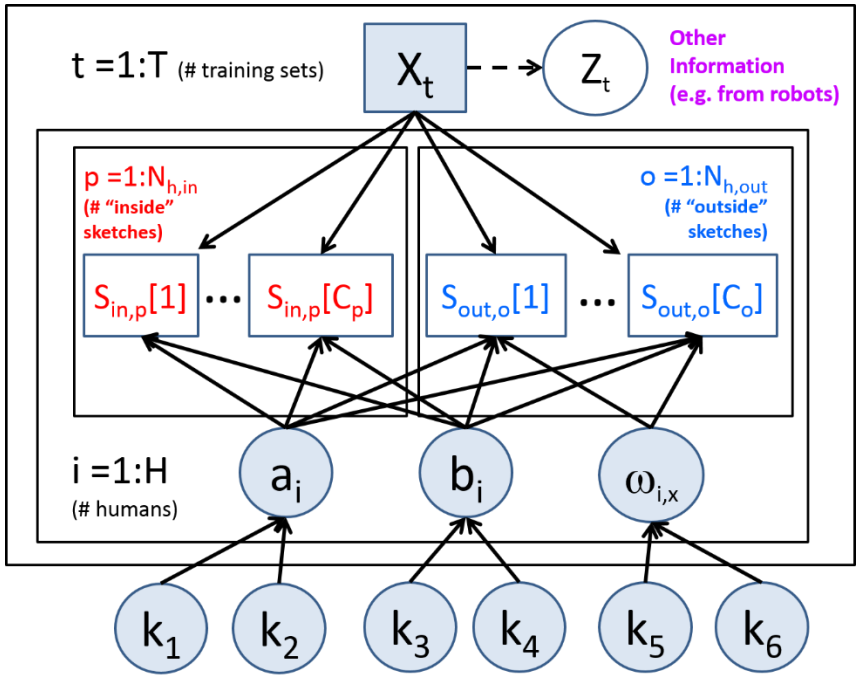
Human-only "Easter egg hunt" Using Sketch Interfaces

- 6 humans searching for partially buried object
- 8 search scenarios, 15 min each
  - vague "clues" about target location

Sample Sketch Data from Human Sensors: Whom to Trust??



Probabilistic model of information dependence:



Fusion = **simultaneous inference of sensor params + target state (solve via Gibbs sampling MCMC)**

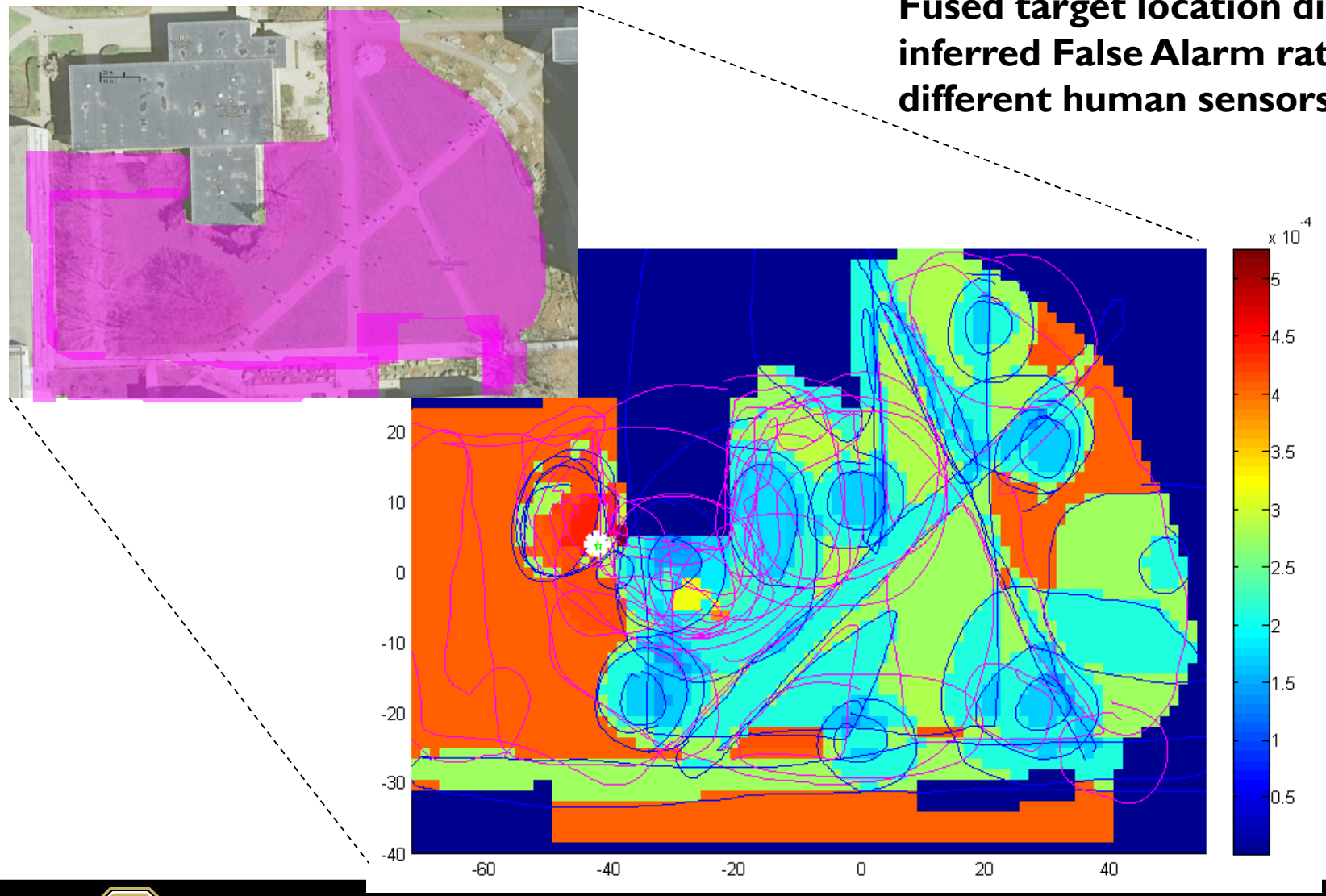
Unknown target location (latent variable)

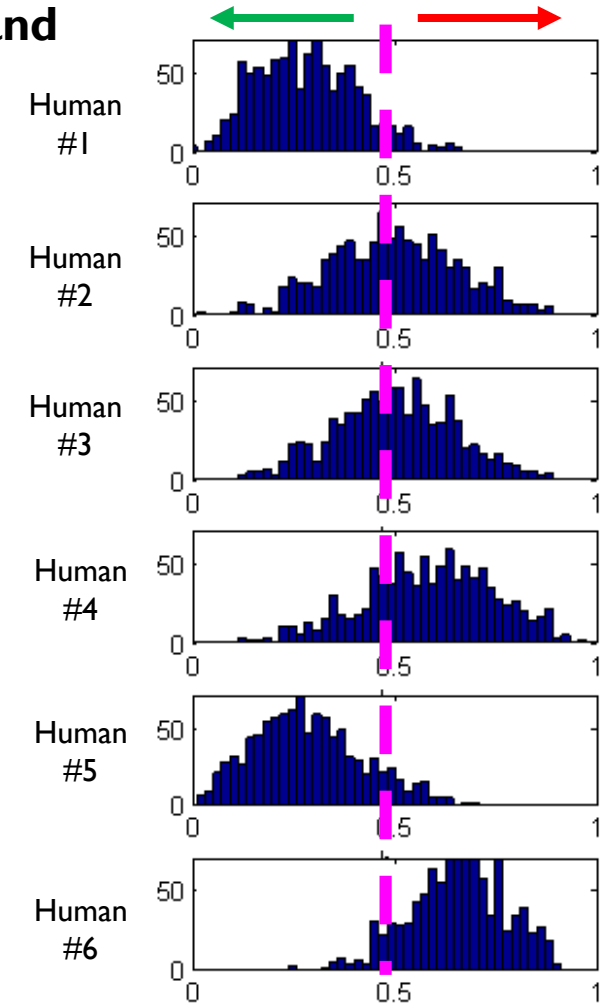
Raw data: Discretized labeled sketches

Meta-uncertainties: Unknown human sensor parameters: false alarm & true detection rates (+ priors) – latent variables

Sample Sketch Data from Human Sensors

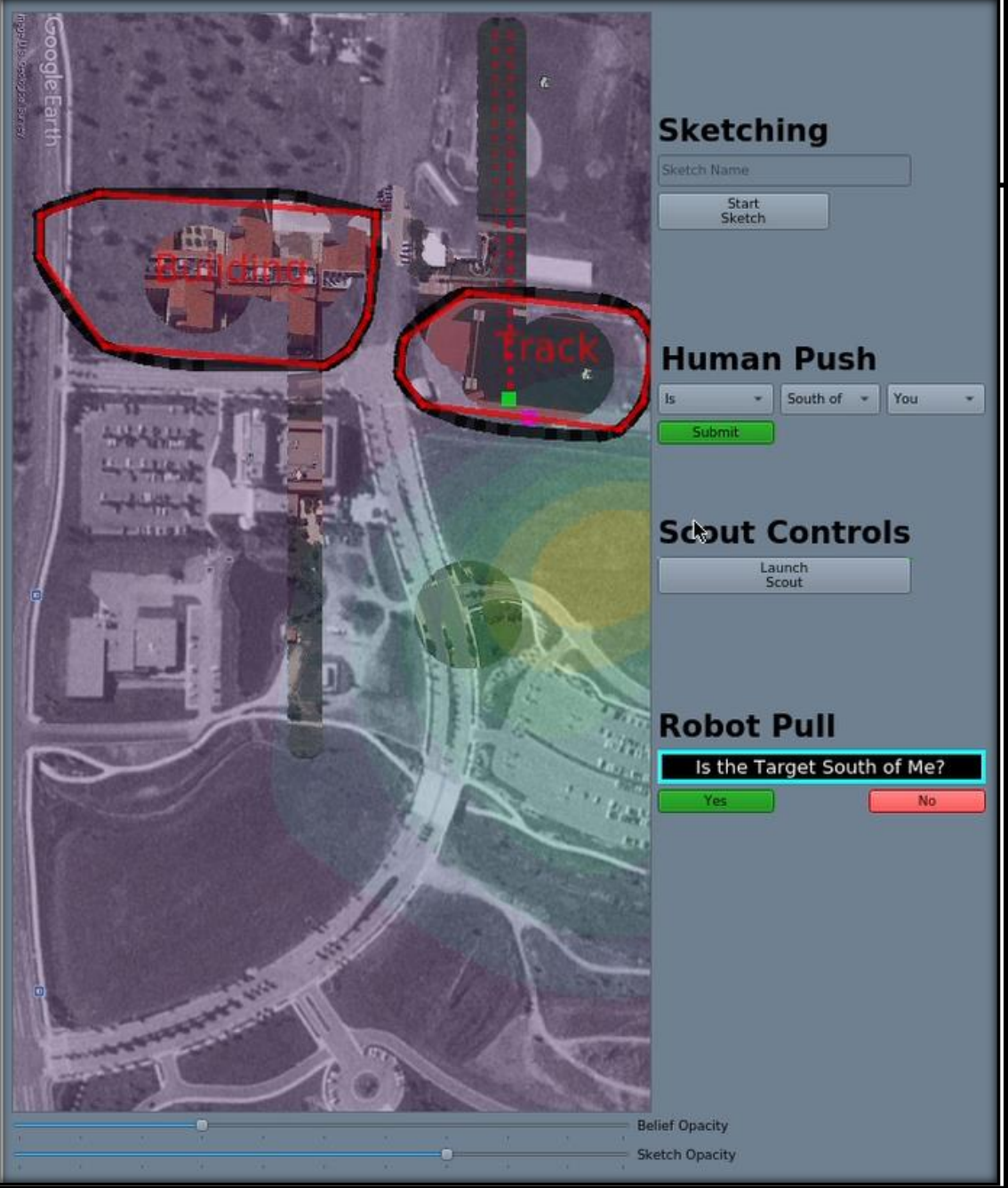
Fused target location distribution and inferred False Alarm rates for different human sensors



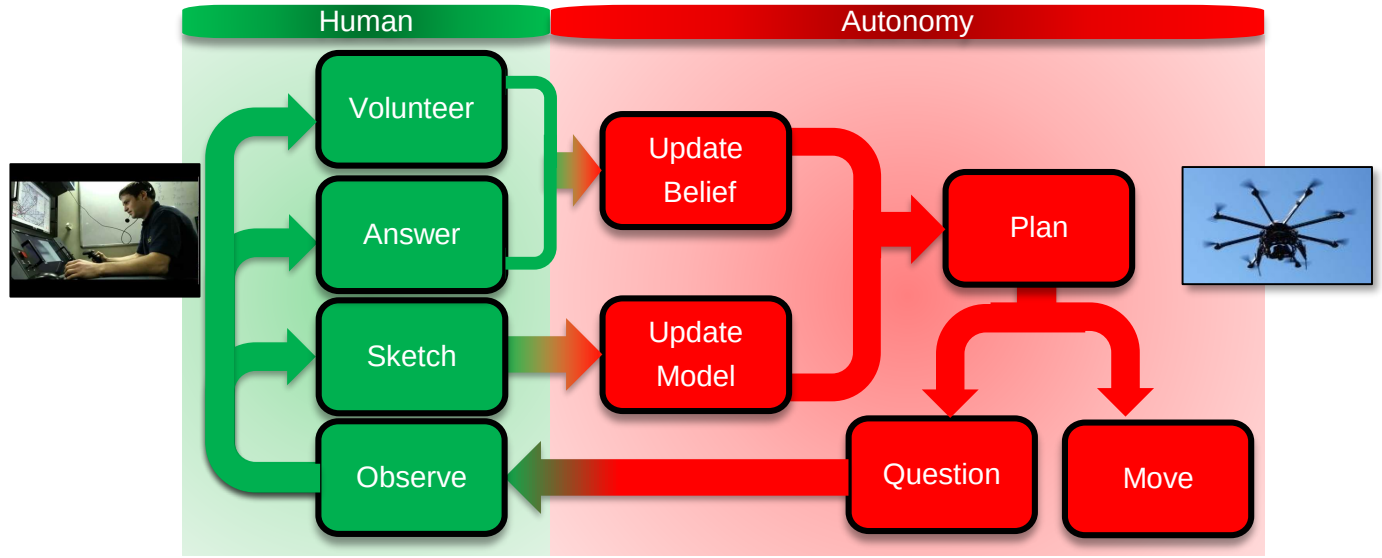


Posterior p(false alarm rate), $p(b_i \mid data)$

Ongoing Work: Sketch + Chat for Dynamically Observable Decision Processes

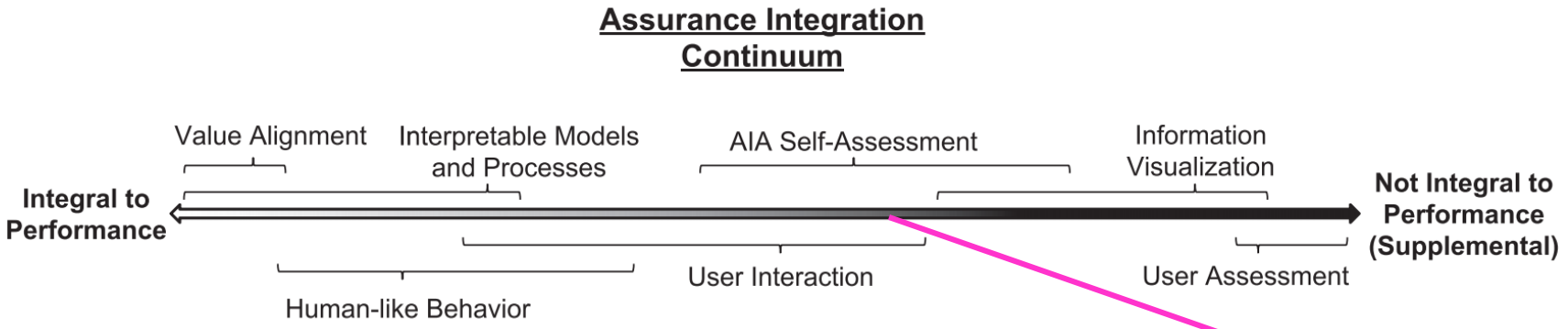


- How to search with **unknown/dynamic maps**?
- Use sketching to create **dynamic chat dictionaries**
- **Dynamically "rewire" POMDP policies online** to improve querying + search movements
- **Multi-level fusion:** query about target state as well as target state dynamics and environment



Algorithmic "Soft Assurances" and Self-Confidence for Autonomy

- How to actively "nudge" user's expectations of behavior/performance to align with actual abilities/competencies?



[Israelsen and Ahmed, "Dave…I can assure you…that it's going to be alright…," ACM Surveys, 2019]

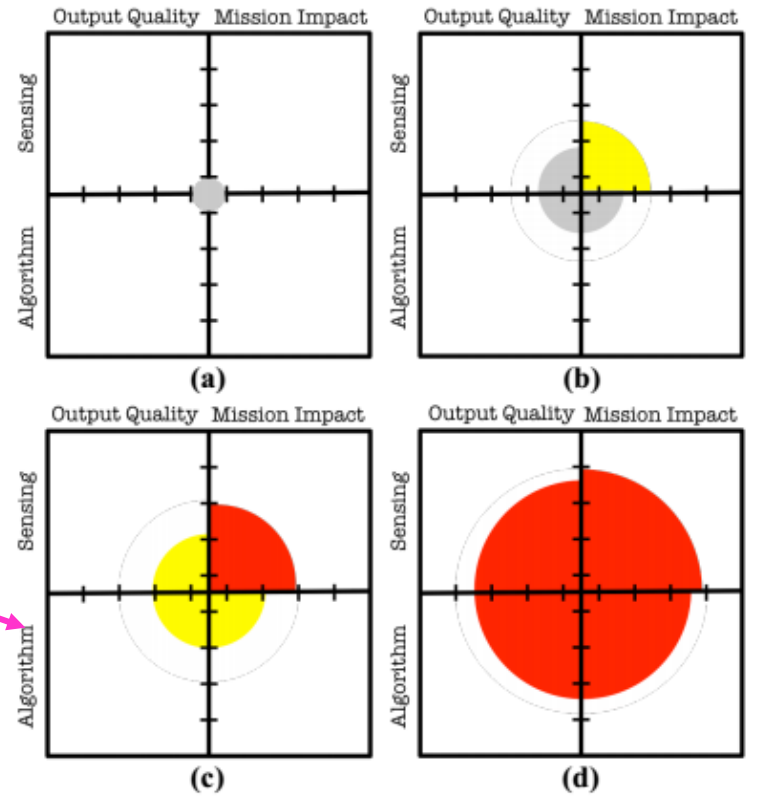


Figure 5: Uncertainty Level Indicator (ULI) with Increasing Uncertainties Proceeding From (a) to (d)

[Hutchins, Cummings, Draper and Hughes, HFES 2015]

Algorithmic "Soft Assurances" and Self-Confidence for Autonomy

(NSF CUAS; DARPA Competency Aware Machine Learning [Draper, CU Boulder, UT Austin])

Remote Operator /Analyst (High level tasking, "Go/No Go" calls)

$$R_k = \begin{cases} +2000 & \text{if } x_{UAV} = \text{Goal}, \\ -2000 & \text{if } x_{UAV} = x_{hostile} \\ -20 & \text{o'wise} \end{cases}$$

Task, System, Environment Parameters

Solver Plan/Policy π

Model-based Task Simulation

Autonomous UAV with RL-based Task Policy π

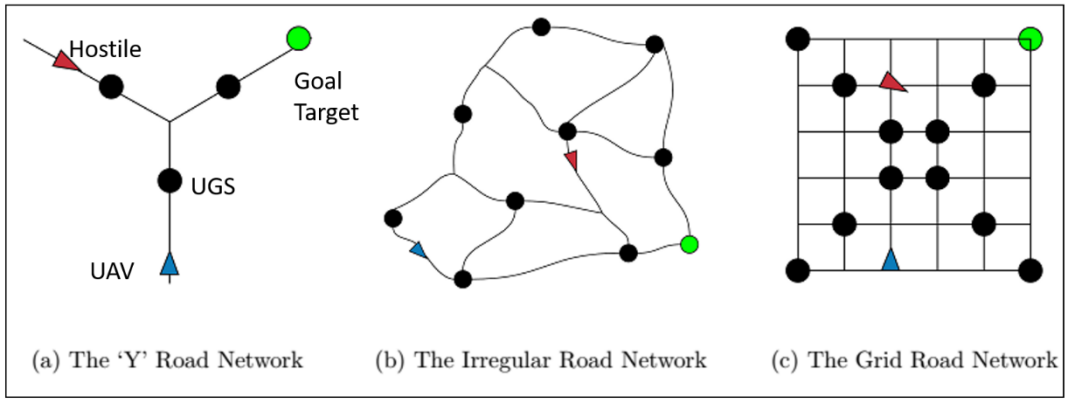
sample R_∞ outcome distribution via π rollout

UPM/LPM Ratio Calculation

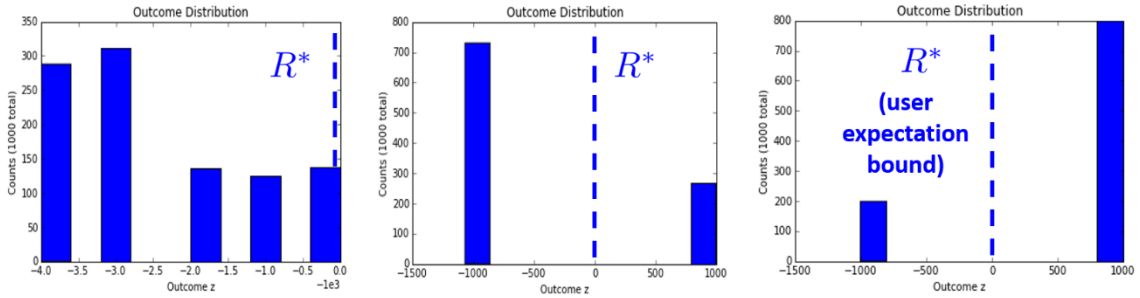
$$\frac{UPM}{LPM} = \frac{\int_{R^*}^{\infty} z p_\pi(z) dR_\infty}{\int_{-\infty}^{R^*} z p_\pi(z) dR_\infty}$$

Logistic Transform

Scenario (a) Scenario (b) Scenario (c)



(a) The 'Y' Road Network (b) The Irregular Road Network (c) The Grid Road Network



Outcome Score $= -1$
Total lack of confidence (impossible to reach goal and avoid hostile)

Outcome Score $= -0.54$
Fairly underconfident (can succeed under favorable conditions)

Outcome Score $= 0.61$
Fairly confident (many trajectories to ensure success)

Summary: Technical Issues and Opportunities

1. Human interaction considered afterthought or band aid/last resort (i.e. the "regrettable but necessary evil" for assurance and safety…)

→ Safety and assurance via "layers of defense": in what ways can humans be more than just the last resort layer? When/where is this appropriate?

2. Theorist/programmer/system designer need for "clean" I/O models of humans:

$$human.behaviors = f(x;\theta)$$

→ Models and reasoning under uncertainty: how to avoid curse of dim/data? How generalizable, transferable, interpretable, tractable, observable, certifiable, safe,…?

3. Intelligent competent machine = loner know-it-all

→ Foundations of introspection and "help seeking": principles for autonomous competency self-assessment and reporting? How to define information utilities? How to accommodate different context, uncertainties, users, systems, tasks…?