What do you really want to do? Towards a Theory of Intentions for Human-Robot Collaboration (Integrating Reasoning and Learning)

Rocio Gomez[1], **Mohan Sridharan**[2], and Heather Riley[1]
[1]The University of Auckland, NZ [2]University of Birmingham, UK m.sridharan@bham.ac.uk, https://www.cs.bham.ac.uk/~sridharm/

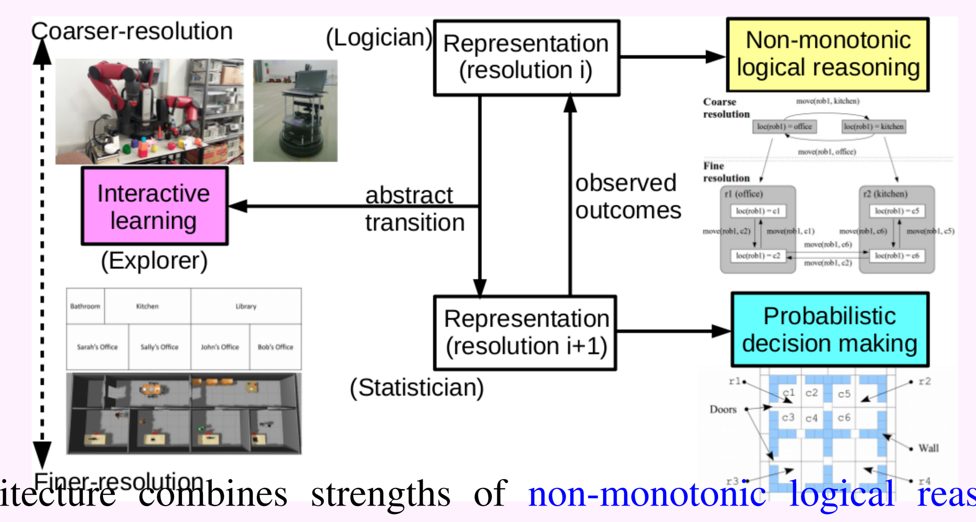
Research Questions

- How to best enable robots to represent and reason with qualitative and quantitative descriptions of incomplete knowledge and uncertainty?
  "Books are usually in the library"
  "I am 90% certain robotics book is in the library"
- How to best enable robots to learn interactively and cumulatively from sensor inputs and limited human feedback?
  Camera images, verbal cues, different surfaces
  "Robot with weak arm cannot lift heavy box"
- How to enable designers to understand the robots' behavior and to establish that it satisfies desirable properties?
  "Why did you go to the kitchen?"
  "How likely is it that the engineer is in the office?

Core Ideas and Inspiration

- Cognitive systems inspired by human cognition and motor control.
- Theories of intention, affordance, explanation, observation.
- Qualitative and quantitative reasoning at different abstractions; tight coupling between logician, statistician, and creative explorer.
- Interactive and cumulative learning of relevant concepts.

Architecture Overview

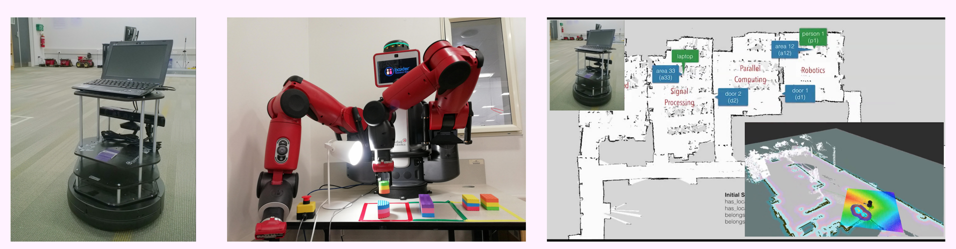


Architecture combines strengths of non-monotonic logical reasoning, probabilistic reasoning, and interactive learning.

Illustrative Domain

Robot Assistant (RA) domain:

- Find and move objects to places or people.
- Humans have *role* (engineer, manager); objects have attributes.
- Estimate occlusion of objects, stability of structures.
- Answer explanatory questions in simulated and real-world scenarios.



Coarse Resolution Domain Representation

- System description $\mathcal{D}_C$ with sorted signature Σ_C and axioms.
- Σ_C has sorts, statics, and fluents. For RA domain:

$$next\_to(place, place),\ loc(thing) = place,\ stable(object),$$
$$in\_hand(robot, object), obj\_relation(relation, object, object)$$

- Σ_C has actions. For RA domain:

$$move(robot, place),\ pickup(robot, object),$$
$$putdown(robot, object), exo\_move(object, place)$$

- Axioms: constraints, causal laws, executability conditions.

$$move(rob_1, Pl) \textbf{ causes } loc(rob_1) = Pl$$
$$loc(O) = Pl \textbf{ if } loc(rob_1) = Pl,\ in\_hand(rob_1, O)$$
$$obj\_relation(above, A, B), I) \textbf{ if } obj\_relation(below, B, A), I)$$
$$\textbf{impossible } pickup(rob_1, Ob_1) \textbf{ if } obj\_relation(below, Ob_1, Ob_2)$$

- History $\mathcal{H}_C$ with prioritized defaults in initial state.

$$\textbf{initial default } loc(X) = library \textbf{ if } book(X)$$
$$\textbf{initial default } loc(X) = office \textbf{ if } book(X),\ loc(X) \neq library$$

- Compute answer sets of CR-Prolog program $\Pi(\mathcal{D}_C, \mathcal{H}_C)$.
- Non-monotonic logical reasoning essential for robotics+AI.

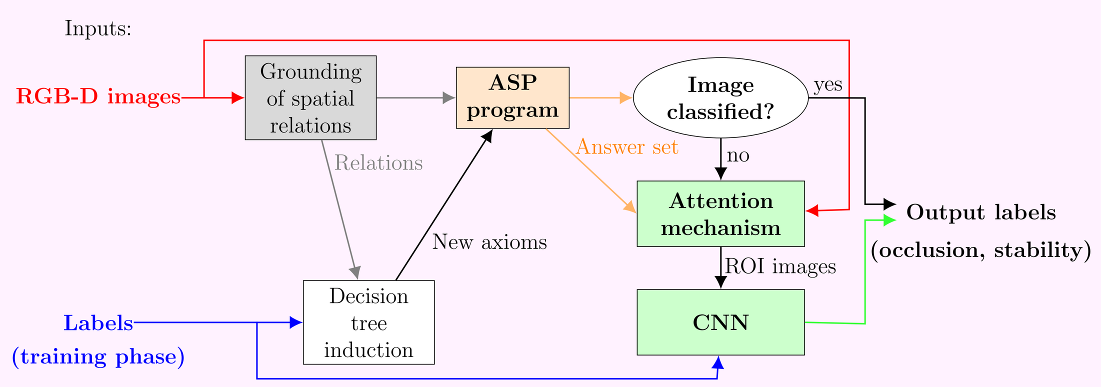
Theory of Affordances and Intentions

- Affordance: attributes of object(s)+agent(s) with reference to actions.
- Unexpected success/failure; model intentional actions, observations.
- Principles of persistence, non-procrastination, and relevance.
- Expand $\mathcal{D}_C$ and $\mathcal{H}_C$; mental fluents and actions; axioms for action effects, start/stop activities; model attempted actions.

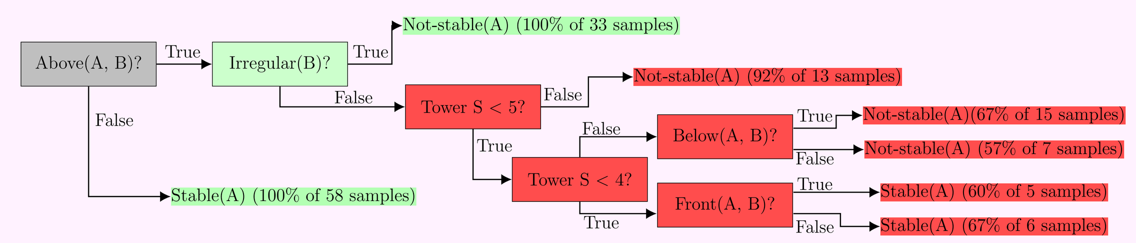
Fine Resolution Domain Representation

- Refinement: describe ($\mathcal{D}_C$) at finer resolution ($\mathcal{D}_F$).
- Theory of observation: knowledge fluents + actions.
- Randomize and zoom to $\mathcal{D}_{FR}(T)$ for $T = \langle \sigma_1, a^H, \sigma_2 \rangle$.
- Formal relationships between descriptions. Separation of concerns.
- Probabilistic model of uncertainty in sensing and actuation.
- Fine-resolution execution with $\mathcal{D}_{LR}(T)$ and probabilities, e.g., POMDP policy, probabilistic grasping. Add coarse-resolution outcomes to $\mathcal{H}$.

Interactive (Deep) Learning

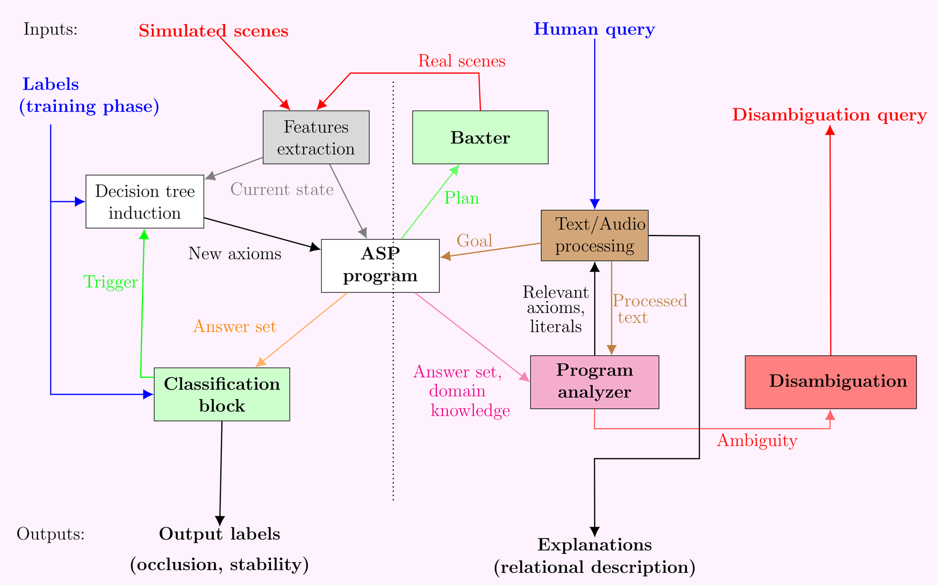


- Incomplete knowledge; unexpected or sub-optimal outcomes.
- Labeled samples; limited human time and expertise; delayed outcomes.
- Incrementally learn previously unknown actions, axioms.
- Generalize from human verbal descriptions: "Robot is labeling big textbook", "Robot labeled small fragile cup": $label(R, O)$ **causes** $labeled(O)$
- Relevance and relational inference guide active exploration or reactive execution with knowledge or reinforcement.
- Reason with knowledge for estimation tasks. If not successful, reasoning guides deep learning with automatically identified ROIs.
- Represent learned model's behavior in (decision tree); cumulative learning and construct new axioms.

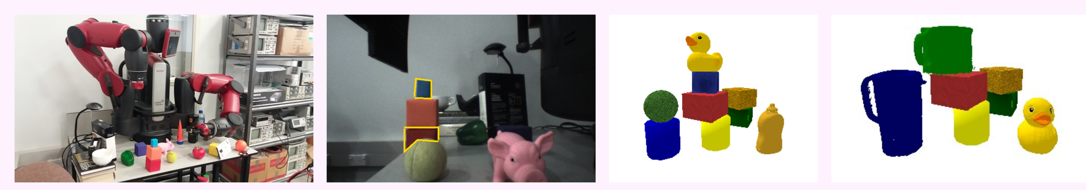


$$\neg stable(A) \textbf{ if } obj\_relation(above, A, B),\ surface(B, irregular)$$
$$\textbf{impossible } grasp(rob_1, C) \textbf{ if } weight(C, heavy), arm(rob_1, electro)$$

Theory of Explanations and VQA



- Characterize explanations: abstraction, specificity, verbosity.
- Methodology for constructing explanations interactively.
- Visual Question Answering (VQA).



- Complementary strengths of non-monotonic logical reasoning, deep learning, and inductive learning.

Experimental Results

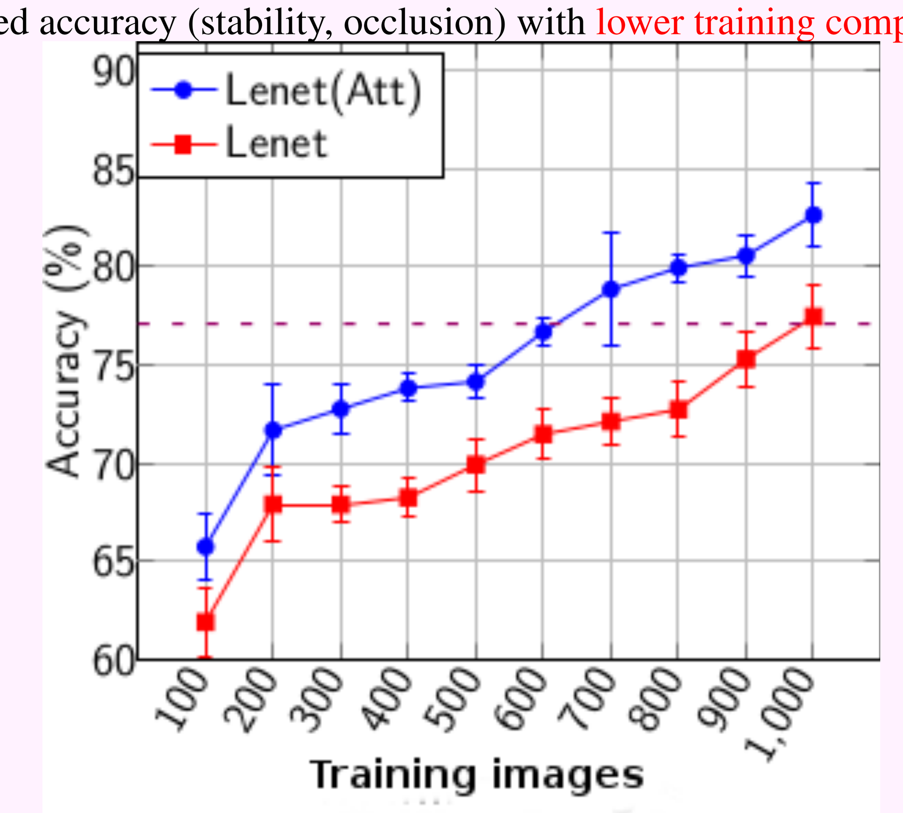
1. Actions and axioms learned with high precision and recall.

Missing Axioms	Precision	Recall
Strict	69.2%	78.3%
Relaxed	96%	95.1%

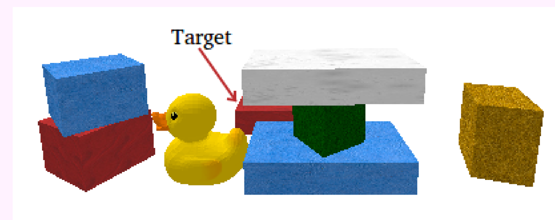
2. Precision and recall of retrieving relevant literals for explanations with and without learned axioms.

	Precision		Recall	
Query Type	Without	With	Without	With
Plan description	78.54%	100%	67.52%	100%
Why X?	76.29%	95.25%	66.75%	95.25%
Why not X?	96.61%	96.55%	64.04%	100%
Belief	96.67%	99.02%	95.6%	100%

3. Desired accuracy (stability, occlusion) with lower training complexity.



4. Minimal and correct plans with learned knowledge.



Conclusions + Future Work

- Step-wise refinement simplifies design and implementation, increases confidence in behavior, promotes scalability.
- Precise relationship between descriptions at different resolutions.
- Reasoning directs interactive learning of domain dynamics.
- Explanations at desired level of abstraction.
- Explore interplay between reasoning and learning in other domains.

Acknowledgements

Research threads pursued in collaboration with Ben Meadows, Tiago Mota, Heather Riley, Rocio Gomez, Michael Gelfond, Shiqi Zhang, and Jeremy Wyatt, with support from the U.S. ONR awards N00014-13-1-0766, N00014-17-1-2434 and N00014-20-1-2390, the AOARD award FA2386-16-1-4071, and the U.K. EPSRC award EP/S032487/1.

References

1. Mohan Sridharan, Michael Gelfond, Shiqi Zhang and Jeremy Wyatt. **REBA: Refinement-based Architecture for Knowledge Representation and Reasoning in Robotics.** In *Journal of Artificial Intelligence Research*, 65:87-180, 2019.
2. Tiago Mota, Mohan Sridharan and Ales Leonardis. **Integrated Commonsense Reasoning and Deep Learning for Transparent Decision Making in Robotics.** In *Springer Nature Computer Science* Journal, 2(4):1-18, 2021.
3. Mohan Sridharan and Ben Meadows. **Knowledge Representation and Interactive Learning of Domain Knowledge for Human-Robot Interaction.** In *Advances in Cognitive Systems Journal*, 7:77-96, 2018.
4. Heather Riley and Mohan Sridharan. **Integrating Non-monotonic Logical Reasoning and Inductive Learning With Deep Learning for Explainable Visual Question Answering.** In *Frontiers in Robotics and AI, special issue on Combining Symbolic Reasoning and Data-Driven Learning for Decision-Making*, Volume 6, 2019.
5. Mohan Sridharan and Ben Meadows. **Theory of Explanation for Human-Robot Collaboration.** In the *Knstliche Intelligenz Journal*, 33(4):331-342, December 2019.