What do you really want to do? Towards a Theory of Intentions for Human-Robot Collaboration

Rocio Gomez[1] , **Mohan Sridharan**[2] , **Heather Riley**[3]
[1]Department of Electrical and Computer Engineering, The University of Auckland, NZ
[2]School of Computer Science, University of Birmingham, UK
[3]Department of Electrical and Computer Engineering, The University of Auckland, NZ
rociogomezbardon@gmail.com, m.sridharan@bham.ac.uk, hril230@aucklanduni.ac.nz

Consider a wheeled robot delivering objects to particular places or people, or a robot with manipulators stacking objects in desired configurations on a tabletop. Such robots that are deployed to assist humans in dynamic domains have to reason with different descriptions of uncertainty and incomplete domain knowledge. Information about the domain often includes commonsense knowledge, especially default knowledge that holds in all but a few exceptional circumstances. For instance, the robot may be told that "books are usually in the library, but cookbooks may be in the kitchen". The robot also extracts information from sensor inputs using algorithms that quantify uncertainty probabilistically, e.g., "I am 95% certain the robotics book is on the table". Although it is difficult to equip robots with comprehensive domain knowledge or provide elaborate supervision, reasoning with incomplete or incorrect information can lead to incorrect or suboptimal outcomes, especially when the robot is faced with unexpected success or failure. For example, a robot may be asked to move two books from the office to the library in a domain with four rooms. If this robot can only grasp one object at a time, it will plan to move one book at a time from the office to the library. After moving the first book, if the robot observes the second book in the library, or in another room on the way back to the office, it should stop executing the current plan because this plan will no longer achieve the desired goal. Instead, it should reason about this unexpected observation and compute a new plan if necessary. One way to achieve this behavior with a traditional planning system is to reason about all observations of domain objects and events during plan execution, but this approach is computationally unfeasible in complex domains. The architecture described in this paper, on the other hand, achieves the desired behavior by equipping a robot pursuing a particular goal with an *adapted theory of intentions* (Gomez, Sridharan, and Riley 2020). This theory builds on the fundamental principles of non-procrastination and persistence in the pursuit of a desired goal. It enables the robot to reason about mental actions and states, automatically identifying and considering the domain observations relevant to the current action and the goal during planning and execution. We refer to actions in such plans as *intentional* actions. Figure 1 provides an overview of our architecture that can be viewed as a logician and an executor communicating through a controller. We describe the following characteristics of our architecture:

- The domain's transition diagrams at two different resolutions are described in an action language, with the fine-resolution transition diagram defined as a refinement of the coarse-resolution transition diagram. At the coarse resolution, non-monotonic logical reasoning with incomplete commonsense domain knowledge, which includes a theory of intentions, produces a sequence of intentional abstract actions for any given goal.
- Each intentional abstract action is implemented as a sequence of concrete actions by automatically zooming to and reasoning with the part of the fine-resolution system description relevant to the current coarse-resolution transition and the goal. Each concrete action in this sequence is executed using probabilistic models of uncertainty, and the observed and inferred outcomes are added to the appropriate coarse/fine-resolution history.

Action languages are formalisms used to model domain dynamics (i.e., action effects). Prior work in our group extended the action language $\mathcal{AL}_d$ (Gelfond and Inclezan 2013) to model non-Boolean fluents and non-deterministic causal laws (Sridharan et al. 2019); we use it here because it provides the desired expressive power for robotics domains. Also, we chose to translate our action language descriptions to programs in CR-Prolog, which extends Answer Set Prolog (ASP) by introducing *consistency restoring* rules (Balduccini and Gelfond 2003). CR-Prolog supports non-monotonic logical reasoning with incomplete commonsense knowledge in dynamic domains, which is a key desired capability in robotics. Furthermore, for the execution of each concrete action on a robot, we use algorithms that incorporate probabilistic models of the uncertainty in perception and actuation.

Our architecture builds on the complementary strengths of prior work on two different architectures. The first architecture used declarative programming and introduced a theory of intentions to enable an agent to reason about intended actions and achieve a given goal (Blount, Gelfond, and Balduccini 2015). The second architecture introduced step-wise refinement of tightly-coupled transition diagrams at two different resolutions to support non-monotonic logical reasoning and probabilistic reasoning for planning and diagnostics in robotics (Sridharan et al. 2019). Prior work on

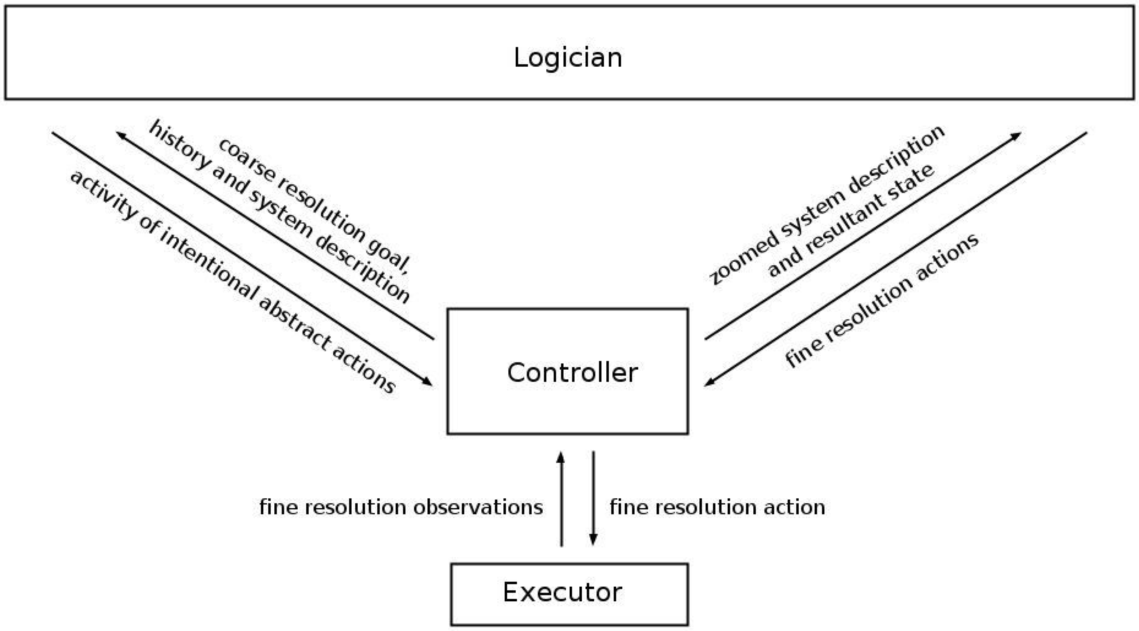

Figure 1: Architecture represents and reasons with intentions and beliefs using tightly-coupled transition diagrams at two resolutions. It combines the complementary strengths of non-monotonic logical reasoning and probabilistic reasoning, and it may be viewed as a logician and an executor communicating through a controller.

the refinement-based architecture did not include a theory of intentions. Also, prior work on the theory of intentions did not consider the uncertainty in sensing and actuation, and did not scale to complex domains. The key contributions of our architecture are thus to:

- enable planning with intentional abstract actions, and the associated mental states, actions, and beliefs, in the presence of incomplete domain knowledge, partial observability, and non-deterministic action outcomes; and
- support scalability to larger domains by automatically restricting fine-resolution reasoning to knowledge and observations relevant to the goal or the coarse-resolution abstract action at hand, and by using probabilistic models of the uncertainty in sensing and actuation only when executing concrete actions.

We demonstrate the applicability of our architecture in the context of a: (i) simulated robot assisting humans in an office domain; (ii) physical robot (Baxter) manipulating objects on a tabletop; and (iii) wheeled robot (Turtlebot) moving target objects to desired locations or people in an office domain. We show that our architecture improves reliability and computational efficiency in comparison with a baseline architecture that does not reason about intentional actions and beliefs at different resolutions, and with a baseline architecture that does not limit reasoning to the relevant part of the domain.

Acknowledgments

This work was supported in part by the Asian Office of Aerospace Research and Development award FA2386-16-1-4071, and the US Office of Naval Research Science of Autonomy award N00014-17-1-2434. Opinions and conclusions are those of the authors.

References

Balduccini, M., and Gelfond, M. 2003. Logic Programs with Consistency-Restoring Rules. In *AAAI Spring Symposium on Logical Formalization of Commonsense Reasoning*.

Blount, J.; Gelfond, M.; and Balduccini, M. 2015. A theory of intentions for intelligent agents. In *International Conference on Logic Programming and Nonmonotonic Reasoning*, 134–142. Springer.

Gelfond, M., and Inclezan, D. 2013. Some Properties of System Descriptions of AL_d. *Journal of Applied Non-Classical Logics, Special Issue on Equilibrium Logic and Answer Set Programming* 23(1-2):105–120.

Gomez, R.; Sridharan, M.; and Riley, H. 2020. What do you really want to do? Towards a Theory of Intentions for Human-Robot Collaboration. *Annals of Mathematics and Artificial Intelligence, special issue on commonsense reasoning*. https://rdcu.be/b3dUf.

Sridharan, M.; Gelfond, M.; Zhang, S.; and Wyatt, J. 2019. REBA: A Refinement-Based Architecture for Knowledge Representation and Reasoning in Robotics. *Journal of Artificial Intelligence Research* 65:87–180.