Redesigning the ICARUS Architecture to Model Social Cognition

Pat Langley
Institute for the Study of Learning and Expertise

Award Period 2/1/12–6/30/15

ONR Cognitive Science and Human-Robot Interaction

Program Review

August 4–7, 2015

Research Objectives

We aim to develop a unified theory of the human cognitive architecture that supports:

- Representing and reasoning about others' mental states
- Flexible inference and problem solving in this context
- Structural learning that supports these processes

The research project's significance lies in its potential to:

- Improve accounts of human reasoning and learning
- Support agents/robots that interact effectively with humans

We have included ideas from the earlier ICARUS architecture but addressed some of its key limitations.

Recent Accomplishments

During the past year, our team's accomplishments have included:

- Extending our model of incremental abductive inference to:
  - Use a more distributed, reified encoding of relations and unifications
  - Incorporate knowledge and beliefs about norms and anomalies
- Implementing / testing a complementary model of questing answering
- Extending our account of flexible problem solving to:
  - Organize search through an OR tree with alternative branches
  - Encode each node as an elaboration of its parent and inherit the rest
  - Handle problems stated as goals, tasks, or their combination
- Improving our model of flexible execution and interleaving to:
  - Operate over the new encoding for hierarchical plans
  - Support more effective revision of these plans when needed

Together, these support our aims to produce a more complete account of human cognitive abilities.

Abductive Inference in UMBRA

In previous years, we have developed UMBRA, an account of everyday reasoning that assumes inference:

- Draws conclusions not only about the environment but also about other agents' *mental states;*
- Involves *abductive* generation of *explanations* via introduction of default assumptions;
- Operates in an *incremental* fashion to process observations that arrive sequentially; and
- Proceeds in a *data-driven* manner because understanding arises from observations about agents' activities.

We have tested UMBRA on a variety of plan recognition tasks, some involving social interaction.

Meadows, Langley, & Emery, PAIR 2013

Extensions to UMBRA

In the previous year, we extended UMBRA to support embedded structures used in reasoning about mental states.

In the past year, we redesigned and reimplemented UMBRA from scratch to incorporate:

- A more distributed (Davidsonian) notation for relations / events
- Knowledge about norms that do not require any explanation
- Marking of assumed norms and detected anomalies
- Reified encoding of unifications to track equality constraints

These changes should lead the system to make fewer incorrect default assumptions and recover from others.

Answering Questions with Explanations

We have also developed a complementary module – PHOS – for answering questions that:

- Operates over the *same explanatory structures* as UMBRA
- Also relies on *incremental abductive inference* with defaults
- Proceeds in a *query-driven* rather than a data-driven manner
- May require *multiple attempts* before generating a response
- *Alters the explanation* and influences answers to later questions

Experiments revealed *depth-of-processing* effects; more effort by UMBRA reduced the effort PHOS required.

Meadows, Heald, & Langley, CogSci 2015

Flexible Problem Solving / Execution

At the previous review, we reported modules for problem solving and execution that:

- Operate over hierarchical decompositions of problems
- Cycle through stages that inspect / manipulate these structures
- Encode strategies as domain-independent control rules
- Including ones for interleaving planning and execution

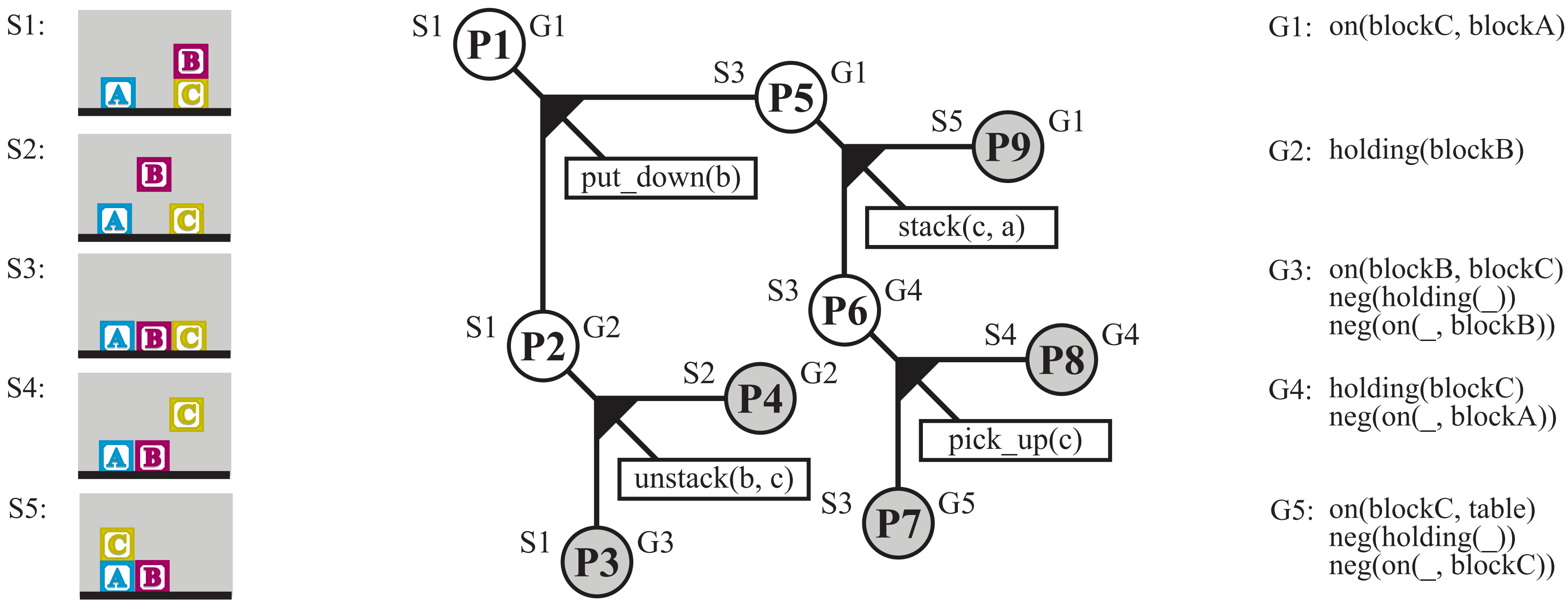
We demonstrated that both modules support a variety of familiar strategies from the literature.

We also tested them, and their combination, on a number of task domains, including ones that involved *social planning*.

Pearce, Meadows, Langley, & Barley, AAAI 2014

A Blocks World Example

A *problem* consists of a set of state literals and goal literals; a *solution* decomposes it into subproblems with operators.



In social settings, state and goal literals include descriptions of others' beliefs and goals.

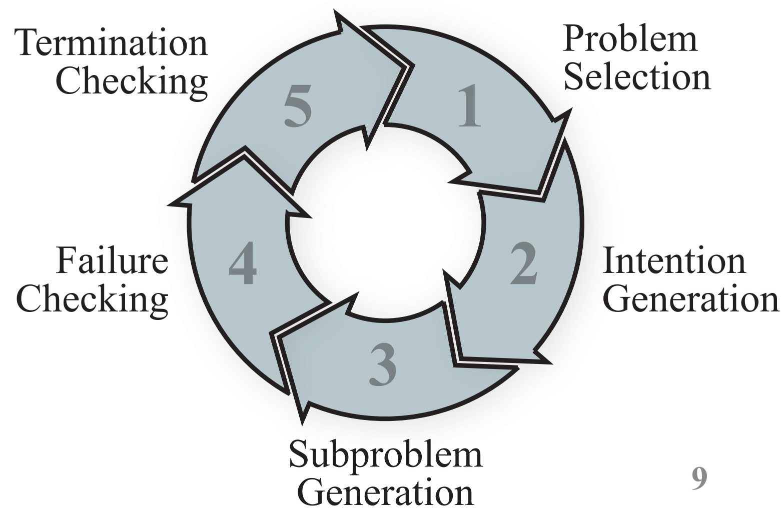
Extensions to Problem Solving

In the past year, we have revised the problem-solving module, now renamed HPS, so that it:

- Organizes search as an OR tree with alternative branches
- Encodes each node as an elaboration on its parent
- Stores only altered elements and inherits the rest
- Retains dependency information to support plan revision
- Handles problems stated as goals, tasks, or in combination

These extensions let the system solve tasks that its predecessor could not.

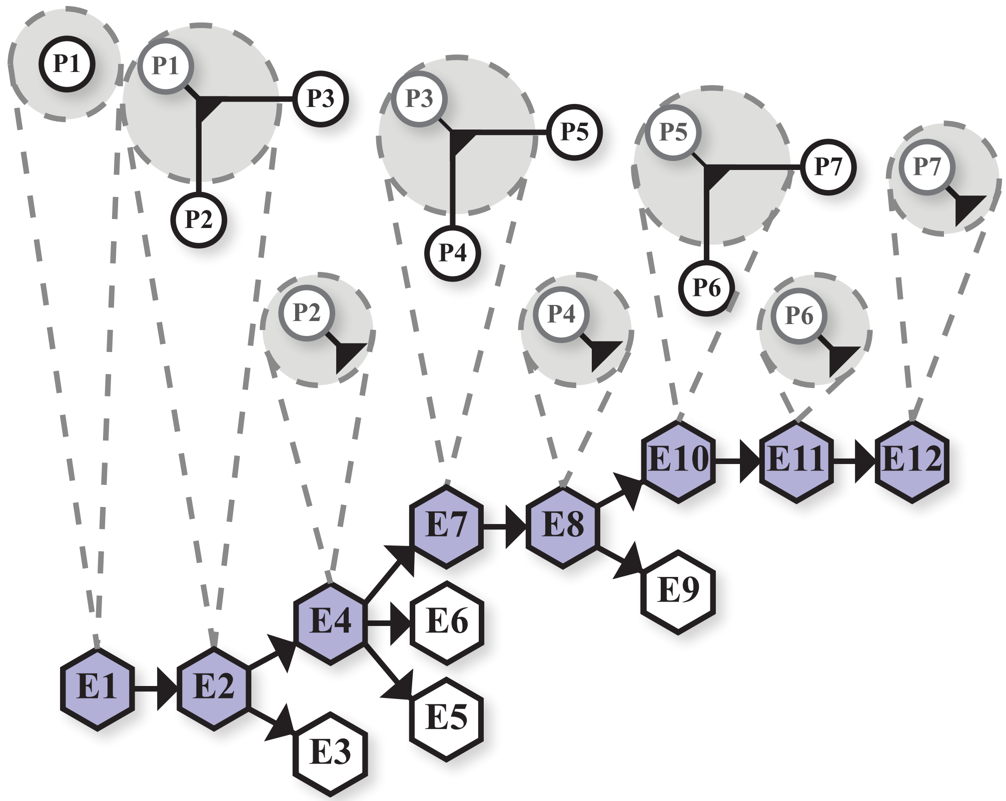
HPS retains the ability to mimic many different problem-solving strategies.



Pearce, Bai, Langley, & Worsfold, TR 2015

The Space of Partial Plans

HPS searches an OR space in which each node denotes a partial plan that elaborates on its parent.



Here the system has found a solution (E12) by following the highlighted path from E1.

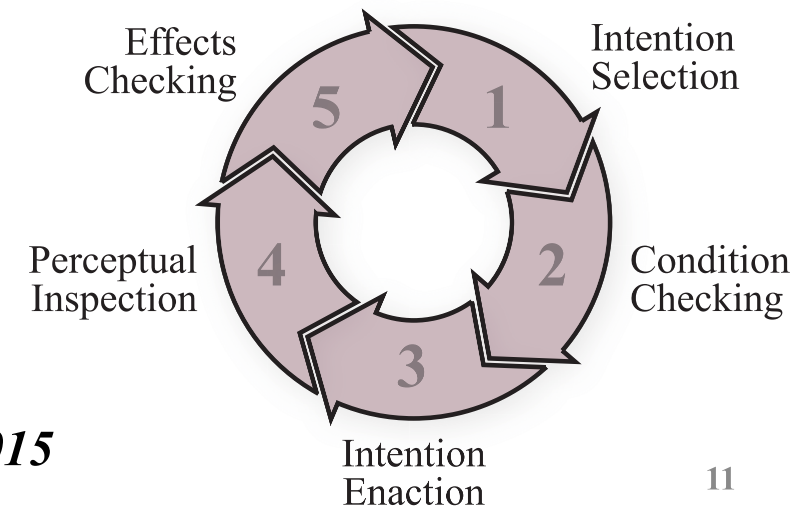
Extensions to Execution / Interleaving

We have also extended the plan execution module, now renamed HPE, so that it:

- Operates over the new encoding for hierarchical plans
- Supports more effective revision of plans when needed
- Including the ability to execute generic HTNs reactively

HPE retains capacity to reproduce many strategies for execution such as open-loop vs. closed-loop control.

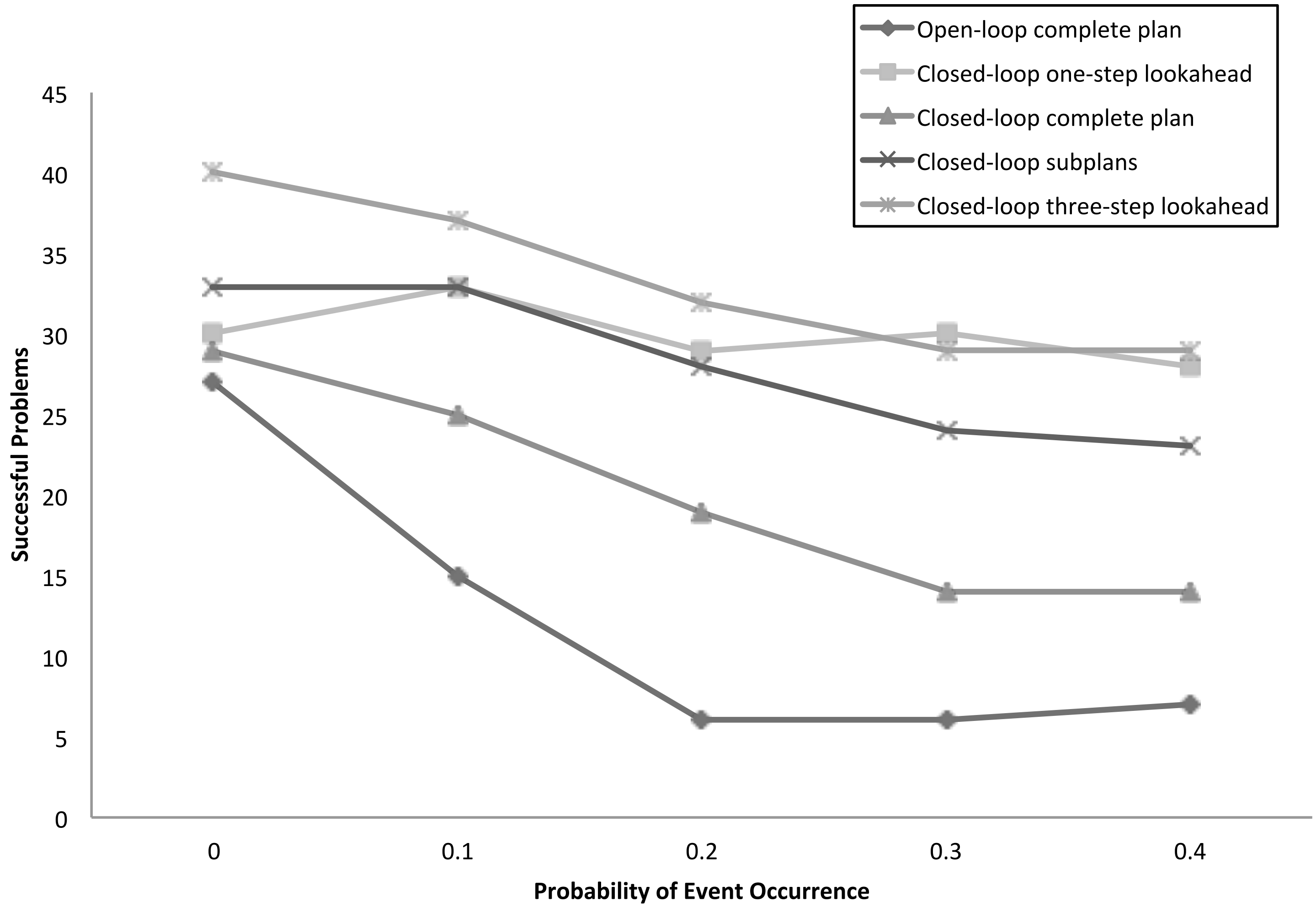
Together, HPS and HPE also support a variety of strategies for interleaving problem solving with execution.



Bai, Pearce, Langley, Barley, & Worsfold, ACS 2015

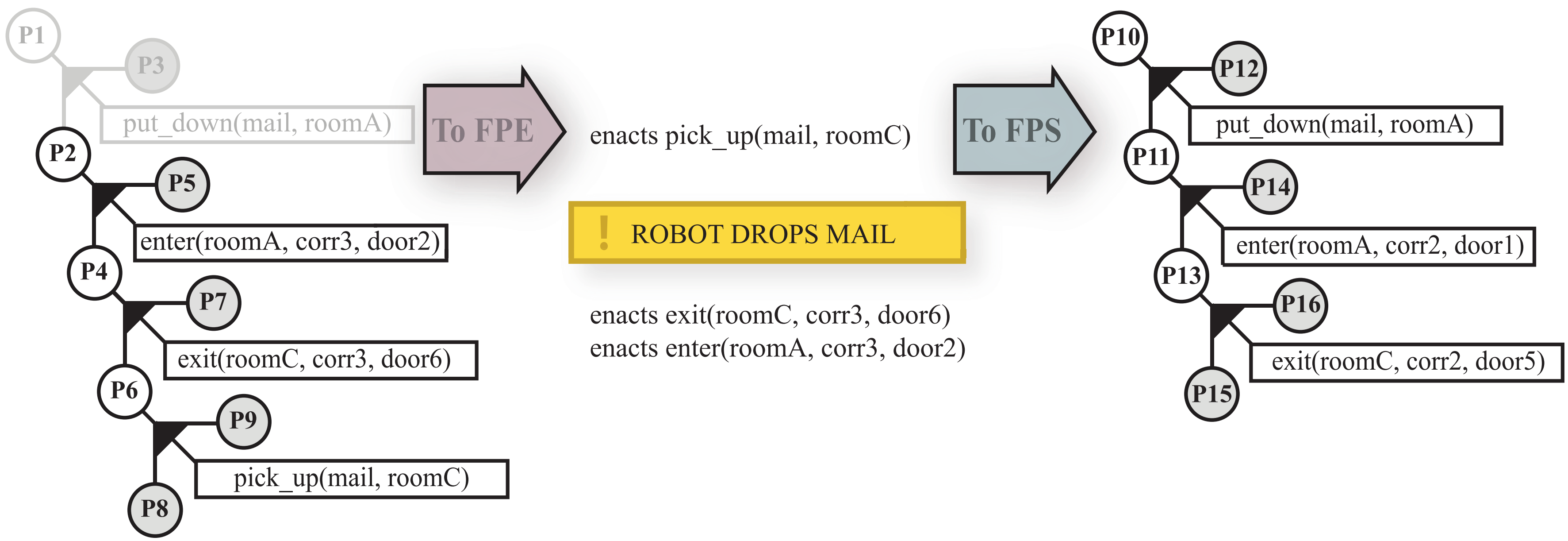
Interleaving Planning and Execution

Experiments with different interleaving strategies suggest their effectiveness varies with environmental characteristics.



Interrupting Execution to Replan

The problem solver, HPS, passes control to the plan executor, HPE, to carry out a (possibly partial) plan.



If HPE finds conditions unexpectedly violated during execution, it passes control back to HPS for replanning.

Plans for Future Research

Although this project has ended, a new effort will build on the results we have achieved to:

- Extend the architectural framework's representational power
- Modify its mechanisms to operate over these extended structures
- Make strategic knowledge conditional on situational features

Our emphasis will be on supporting greater *adaptability* and thus increased *autonomy* in intelligent agents.

This work will draw on our results with social cognition, but we do not plan to extend those abilities.

Summary Remarks

In this talk, I presented elements of a new architectural framework that addresses ICARUS' limitations by:

- Representing mental states in terms of embedded beliefs and goals
- Incorporating modules for incremental abduction that enable:
  - Data-driven generation of explanations for observed behavior
  - Query-driven use of these explanations to answer questions
- Including modules for problem solving and execution that:
  - Construct plans by decomposing problems into subproblems
  - Use meta-level rules to create and execute plans with different strategies
  - Support alternative strategies for interleaving planning with execution

These elements provide the building blocks for a new cognitive architecture that will support more autonomous agents.

Publications in the Past Year

Meadows, B., Heald, R., & Langley, P. (2015). An integrated account of explanation and question answering. *Proceedings of the Thirty-Seventh Annual Meeting of the Cognitive Science Society*. Pasadena, CA.

Pearce, C., Bai, Y., Langley, P., & Worsfold, C. (2015). *Problem solving through successive decomposition*. Technical Report, Department of Computer Science, University of Auckland, Auckland, NZ.

Bai, Y., Pearce, C., Langley, P., Barley, M., & Worsfold, C. (2015). An architecture for flexibly interleaving planning and execution. *Poster Collection: The Third Annual Conference on Advances in Cognitive Systems*. Atlanta, GA.

Gabaldon, A., Meadows, B., & Langley, P. (in press). Knowledge-guided interpretation and generation of task-oriented dialogue. In A. Raux, W. Minker, & I. Lane (Eds.), *Situated dialog in speech-based human-computer interaction*. Berlin: Springer.

Langley, P., Meadows, B., Gabaldon, A., & Heald, R. (2014). Abductive understanding of dialogues about joint activities. *Interaction Studies*, *15*, 426–454.

Langley, P. (2014). Four research challenges for cognitive systems. *Advances in Cognitive Systems*, *3*, 3–11.

Cooperative Development

Our research on this project has benefited from results produced on a number of other efforts:

- Commitments to hierarchical concepts / skills borrowed from initial ICARUS architecture developed under ONR funding;
- Representation of mental states developed jointly with ONR MURI project at CMU;
- Ideas on abductive inference co-developed with W. Bridewell in ONR MURI work at Stanford.

These efforts have let us make more rapid progress than would have been possible otherwise.

Project Budget

The research project's budget, by federal fiscal year, is:

- FY2012: $118K
- FY2013: $179K
- FY2014: $182K
- FY2015: $ 60K

No DURIP were awarded in relation to this project.

End of Presentation