Advances in Cognitive Systems
♠ 18 November 2021

NSF GRANT #1849131

ONR GRANT #N00014-18-1-2009

Computational Metacognition

MICHAEL COX, ZAHIDUDDIN MOHAMMAD, SRAVYA KONDRAKUNTA, VENTAKSAMPATH GOGINENI*

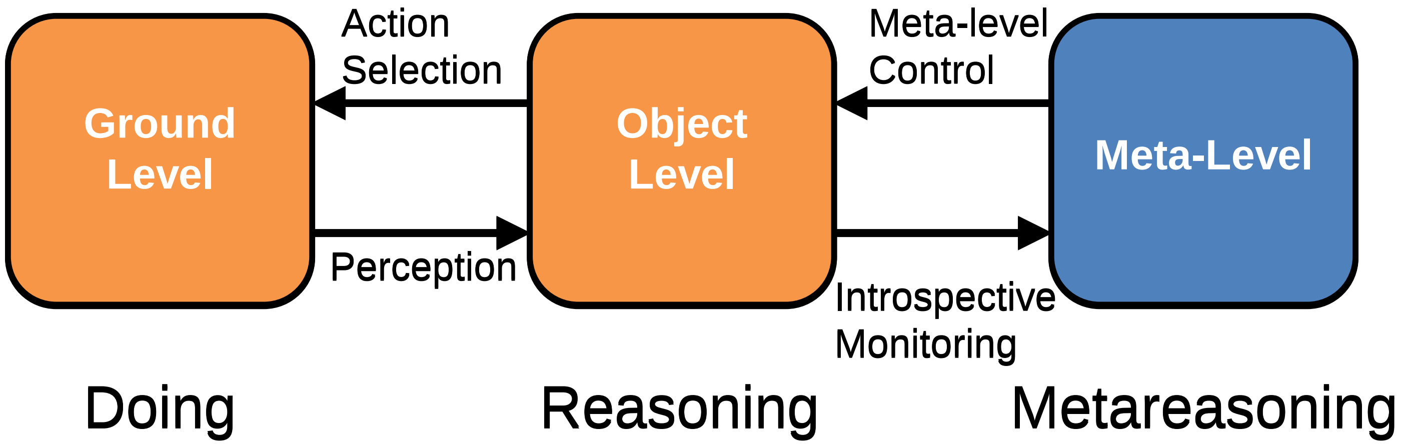
DUSTIN DANNENHAUER AND OTHALIA LARUE

PARALLAX ADVANCED RESEARCH, BEAVERCREEK, OH

*DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, WRIGHT STATE UNIVERSITY, DAYTON, OH

MCOX.ORG ♠ MIDCA-ARCH.ORG ♠ #PAPER01-COX

Model of Computational Metacognition



from Cox & Raja (2011)

Types of Computational Metacognition

Explanatory	Immediate	Anticipatory
Past	Present	Future
Hindsight	Insight	Foresight
Retrospective	Introspective	Predictive

Outline

Introduction

Basic Formalisms

The MIDCA Architecture

Problem Domain and Example

Experiments

Conclusion

Basic Formalisms

COGNITIVE AND METACOGNITIVE

Explanatory Metacognition

Expectations

- Expectation failure when $s_e \neq s_c$

Metacognitive Expectation Failure

- Mental state $s^M = (v_1, v_2, \dots, v_n)$
- Mental action α^M
- Metacognitive expectation $(s_i^M, \alpha_i^M, s_{i+1}^M)$
- Example $(g_c \in s_i^M, Plan, \pi \in s_{i+1}^M)$

Potential Responses g_c^M

- Change reasoning method
- Change the goals
- Learn new knowledge

Formalisms (Cognitive)

State Transition System: possible states and actions, successor function

$$\Sigma \ = \ (S, A, \gamma)$$

Successor Function: returns next state given current state and action

$$\gamma : S \times A \longrightarrow S$$

Problem Solution: a sequence of actions (plan)

$$\pi = \alpha_1 \mid \pi[2 \ldots n] = \langle \alpha_1, \alpha_2 \ldots \alpha_n \rangle$$

Plan Execution: starting from the initial state (s_0) results in the goal state (s_g)

$$\gamma(s_0, \pi) = \gamma(\gamma(s_0, \alpha_1), \pi[2 \ldots n]) \rightarrow s_g \vDash g$$

Formalisms (Metacognitive)

Self-Model

- $\Omega = (S^M, A^M, \omega)$

Cognitive Transition Function

- $\omega: S^M \times A^M \rightarrow S^M$

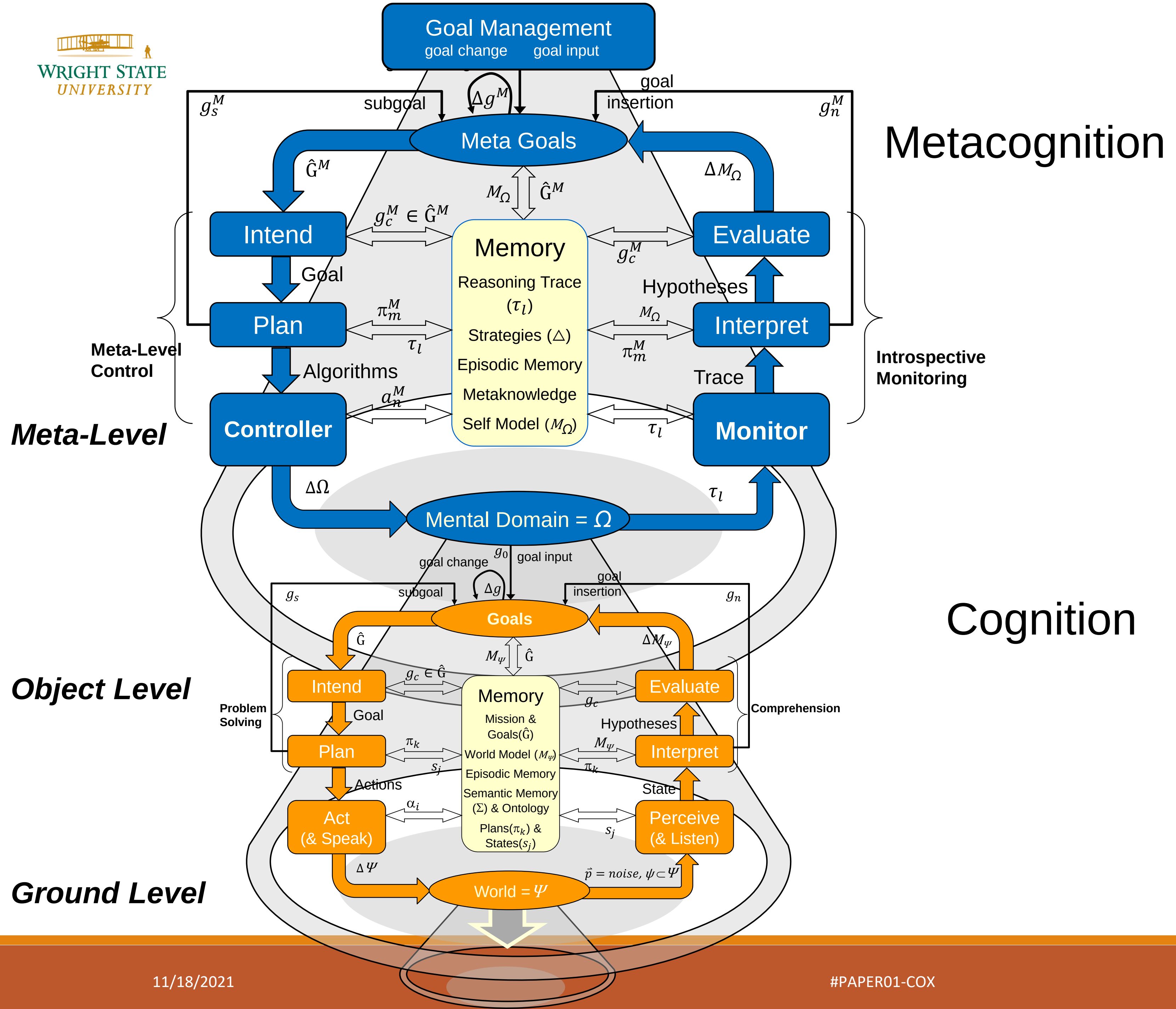
Cognitive Trace

- $\tau = \langle s_0^M, \alpha_1^M, s_1^M, \alpha_2^M, \ldots, \alpha_n^M, s_n^M \rangle$

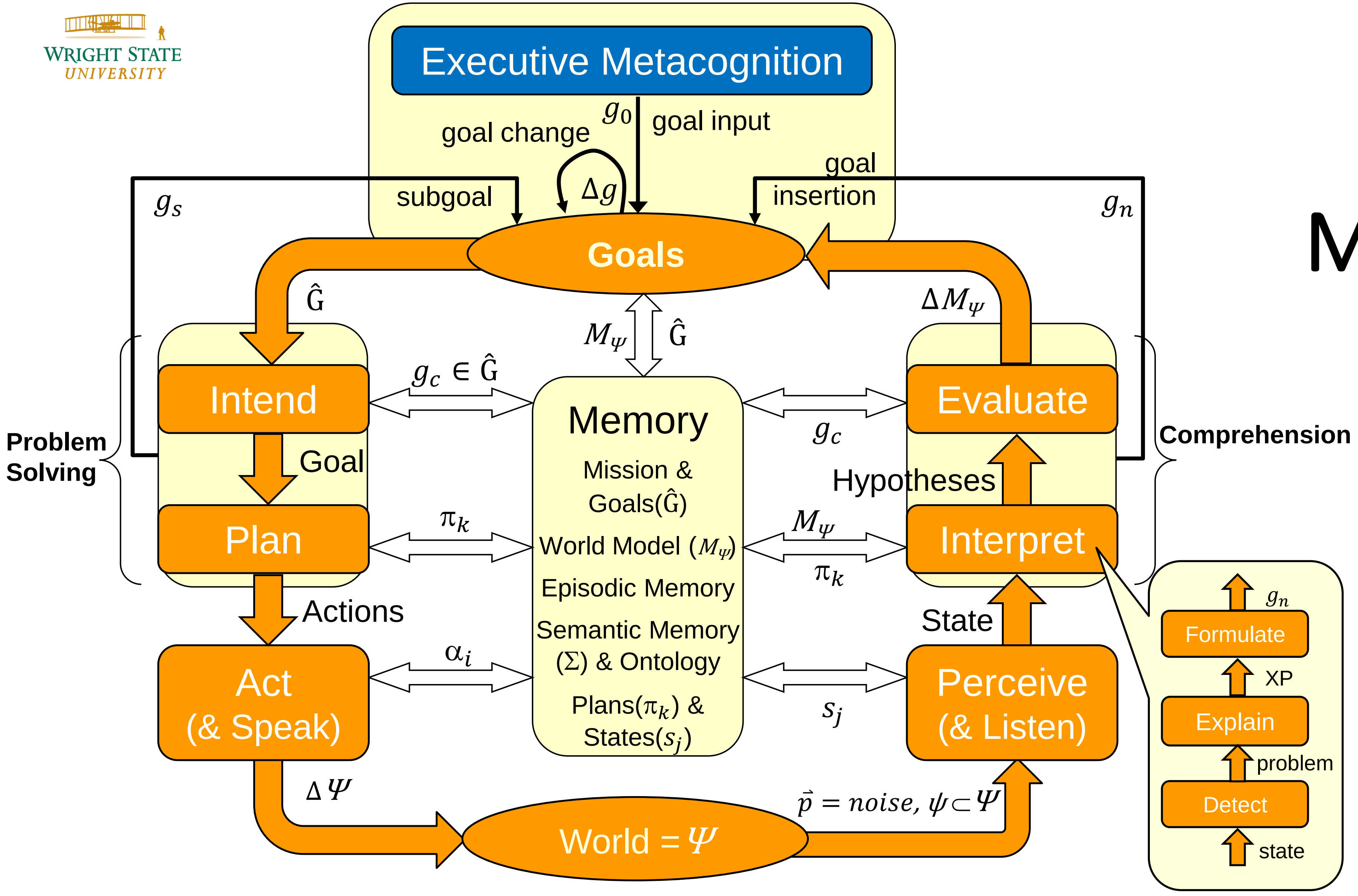
The MIDCA Architecture

THE METACOGNITIVE INTEGRATED DUAL-CYCLE ARCHITECTURE (MIDCA)

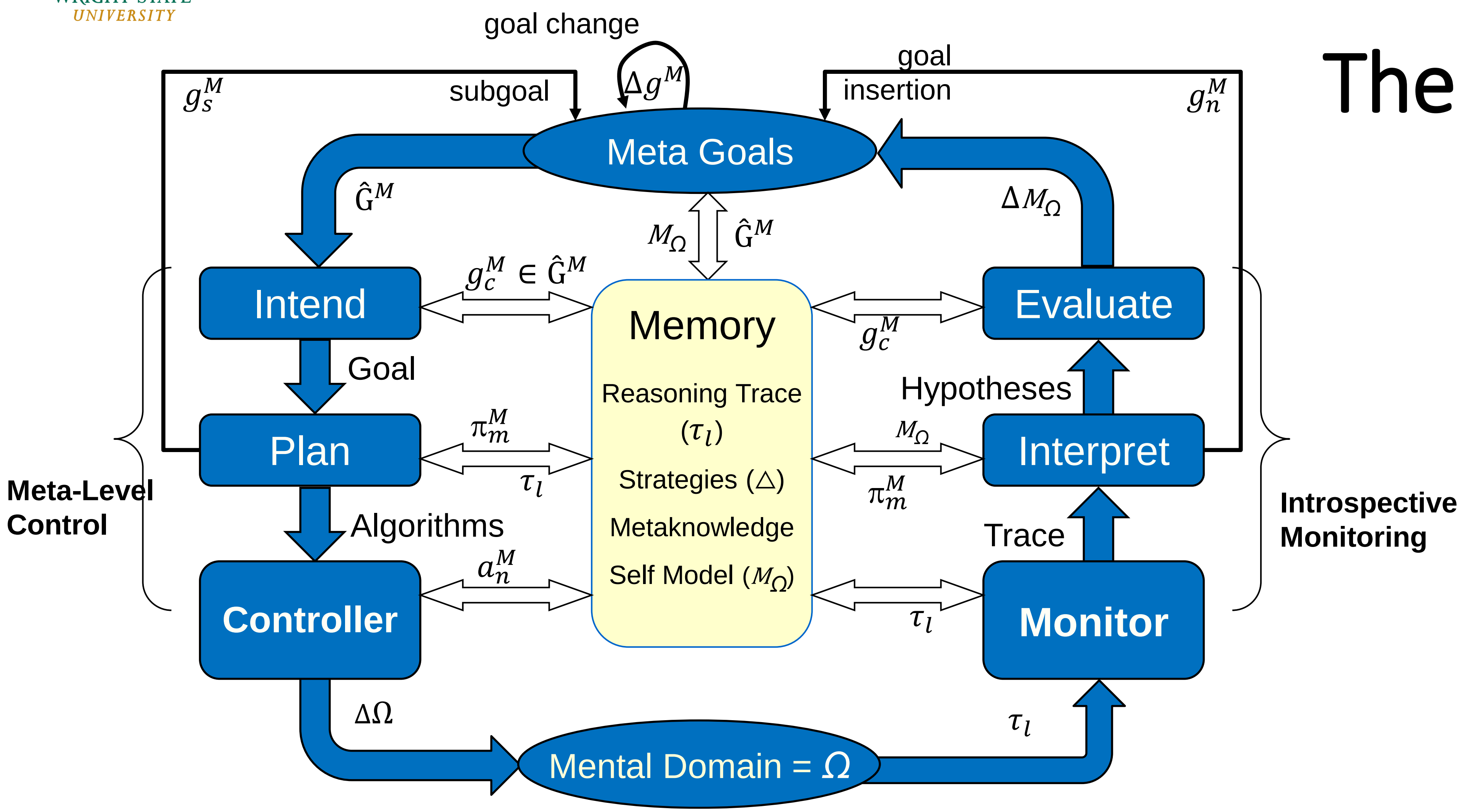
The Metacognitive Integrated Dual-Cycle Architecture (MIDCA)



The Metacognitive Integrated Dual-Cycle Architecture (MIDCA)



The Meta-Level of MIDCA



The Problem Domain

AND AN EXAMPLE

The Plant Protection Domain

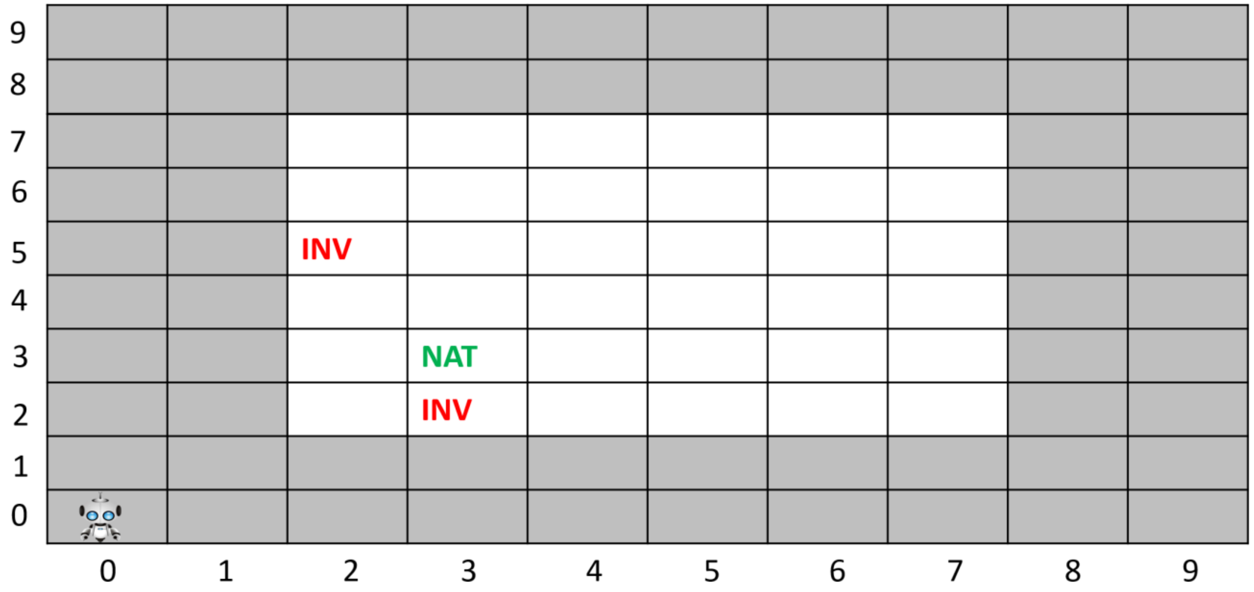
GOALS TO ACHIEVE

- Native Plants Preserved
- Invasives Dead

ACTION MODELS

- Move
- Spray
- Communicate

Plant Protection Example



MIDCA Specifics

Mental State is a 7-tuple

- $s_i^M = (g_c, \hat{G}, \pi_k, M_\Psi, D, E, \alpha_i)$

Mental Actions are from the set

- $A^M = \{Perceive, Detect\text{–}Discrepancies, Explain, Goal\text{–}Insertion, Evaluate, Intend, Plan, Act\}$

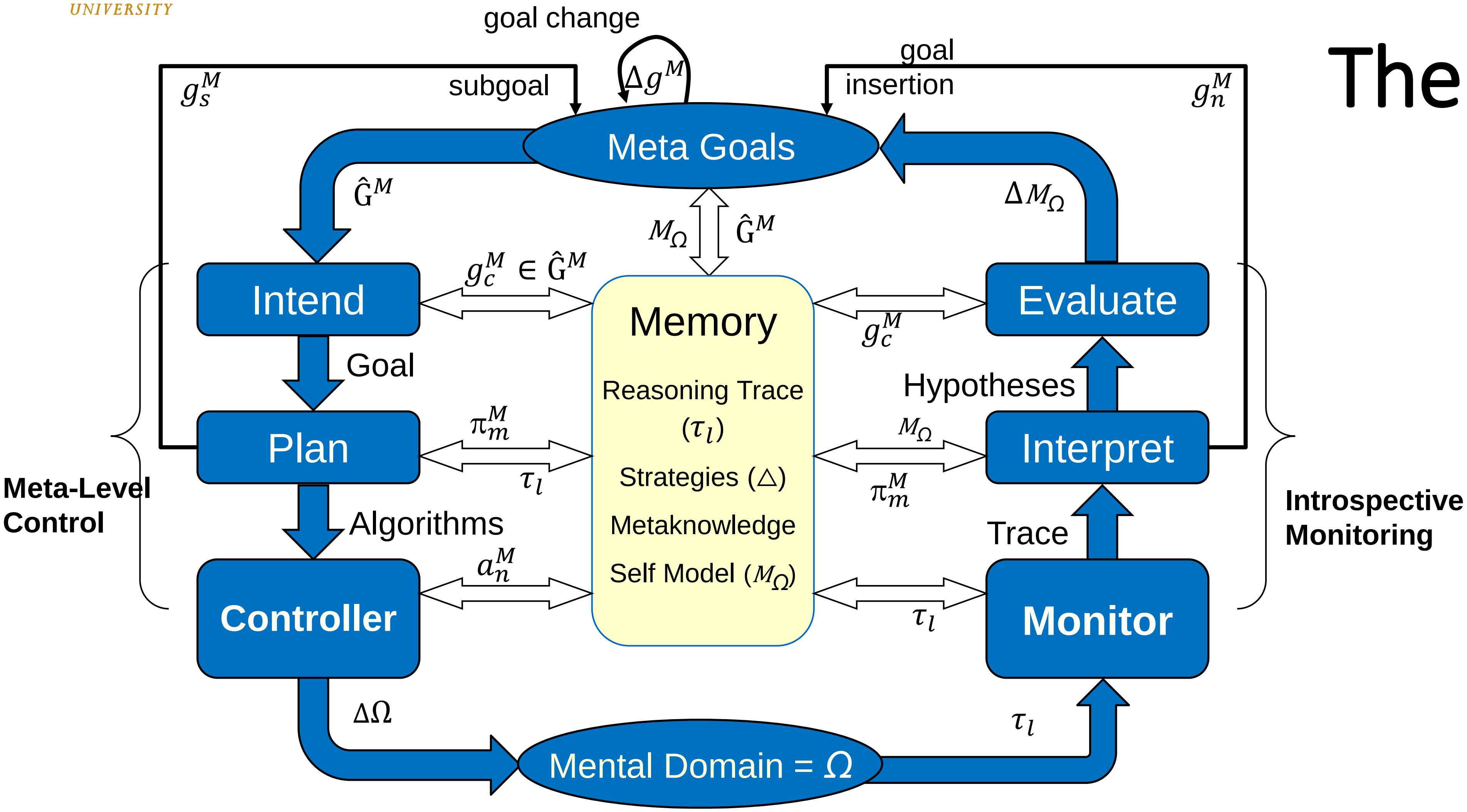
Metacognitive Expectation

- $(discrepancy \in s_i^M,\ \mathit{Explain}, explanation \in s_{i+1}^M)$

Learning Goal

- $(learned\ spray\ s_{i+1})$

The Meta-Level of MIDCA



Achieving a Learning Goal

FOIL learns a horn clause

- spray (pos1, time2) :- spray (pos0, time1), adj_time (time2, time1), adj_north (pos0, pos1)

Conditional Effect translated into spray operator

- (forall (?pos - mapgrid)
  (when (and (adj_north ?to ?pos))
  (and (not (native-at ?pos))(not (invasive-at ?pos)))))

Experiments

EMPIRICAL EVALUATION OF METACOGNITION

Experimental Design

Standard Planning Agent

- No metacognition, no learning, no goal reasoning

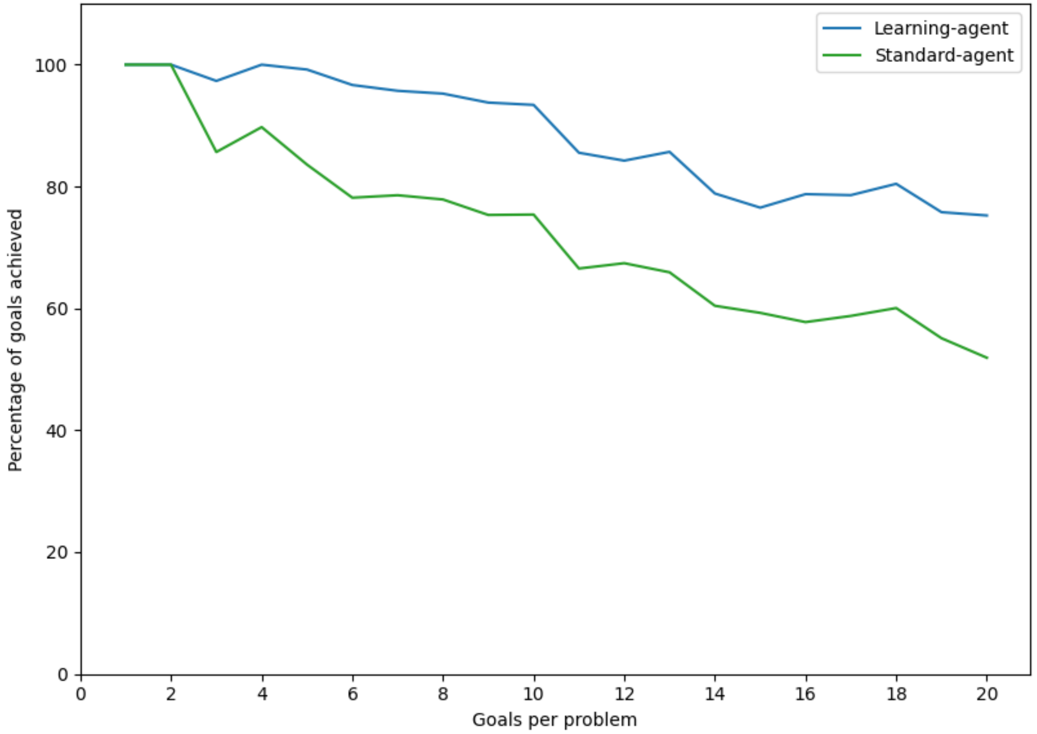
Metacognitive/Learning Agent

Plants Placed Randomly – only one plant per cell

Parameters

- *Ratio* - native:invasive held constant
- *Number of Goals* - range from 1 to 20
- *Number of Trials* - 100 for each (thus, 2000 per experiment)

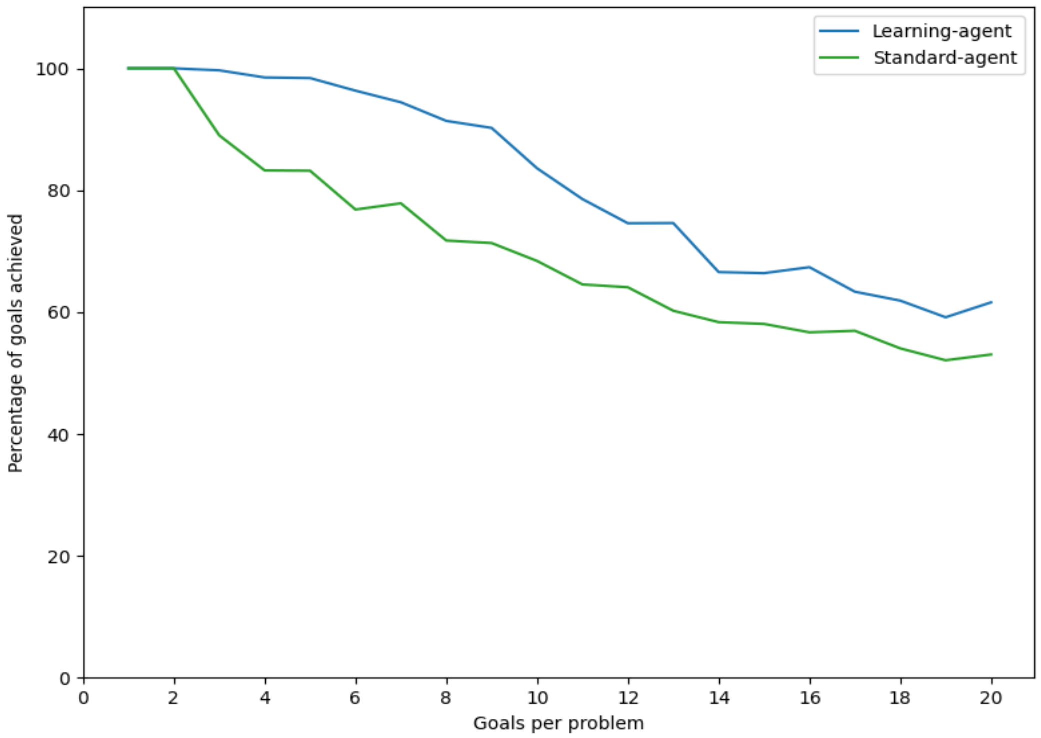
Empirical Results for Experiment 1



Performance As a Function of Problem Complexity

Ratio of native to invasive: ***75:25***

Empirical Results for Experiment 2



Performance As a Function of Problem Complexity

Ratio of native to invasive: ***60:40***

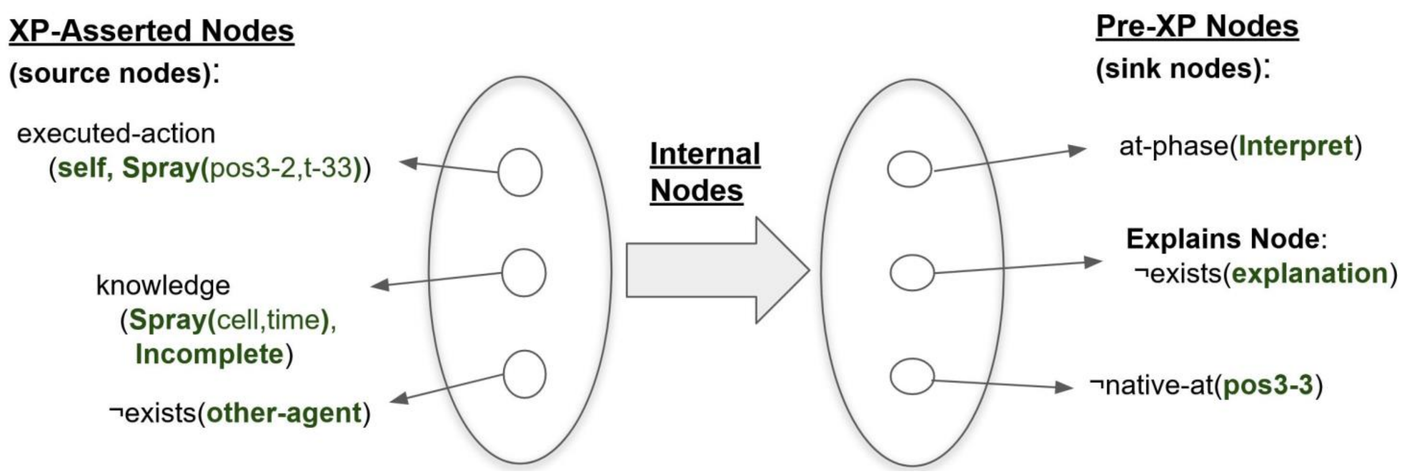
Conclusion

- **Open-source Code** available at `https://github.com/COLAB2/midca`
- **Integrating cognition and metacognition** is hard in any domain
  - Much still in preliminary stages, but exciting results are emerging
  - Previous results at the cognitive level and the state of the art in the planning community suggest that solutions to more complex metacognitive problems are feasible
- **The Future** is interesting

Backup Slides

INTERACTIONS OF GOAL OPS

Meta-level Explanation



Meta-level PDDL Operator

```
(:action perform-learning
   :parameters (?op - operator ?current-state - state)
   :precondition
   (and
      (has-discrepancy ?current-state)
      (outdated ?op)
      (caused_discrepancy ?op ))
   :effect
   (and
      (learned ?op ?current-state))
 )
```