Deep Goal Reasoning: An Analysis

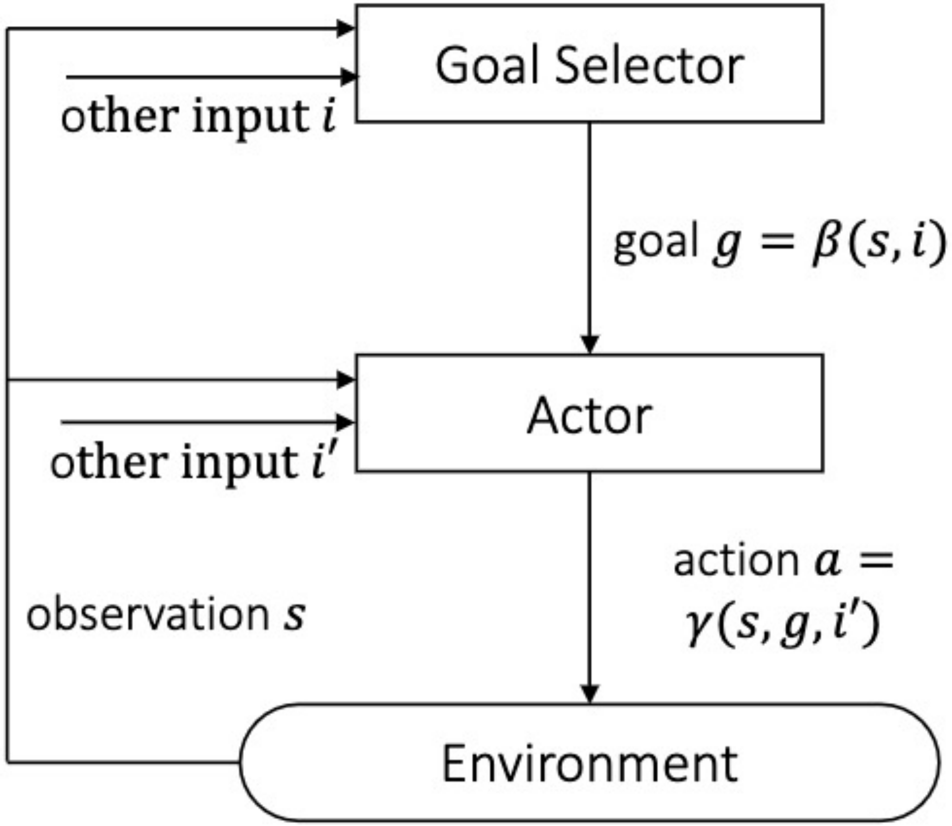
Weihang Yuan

Hector Munoz-Avila

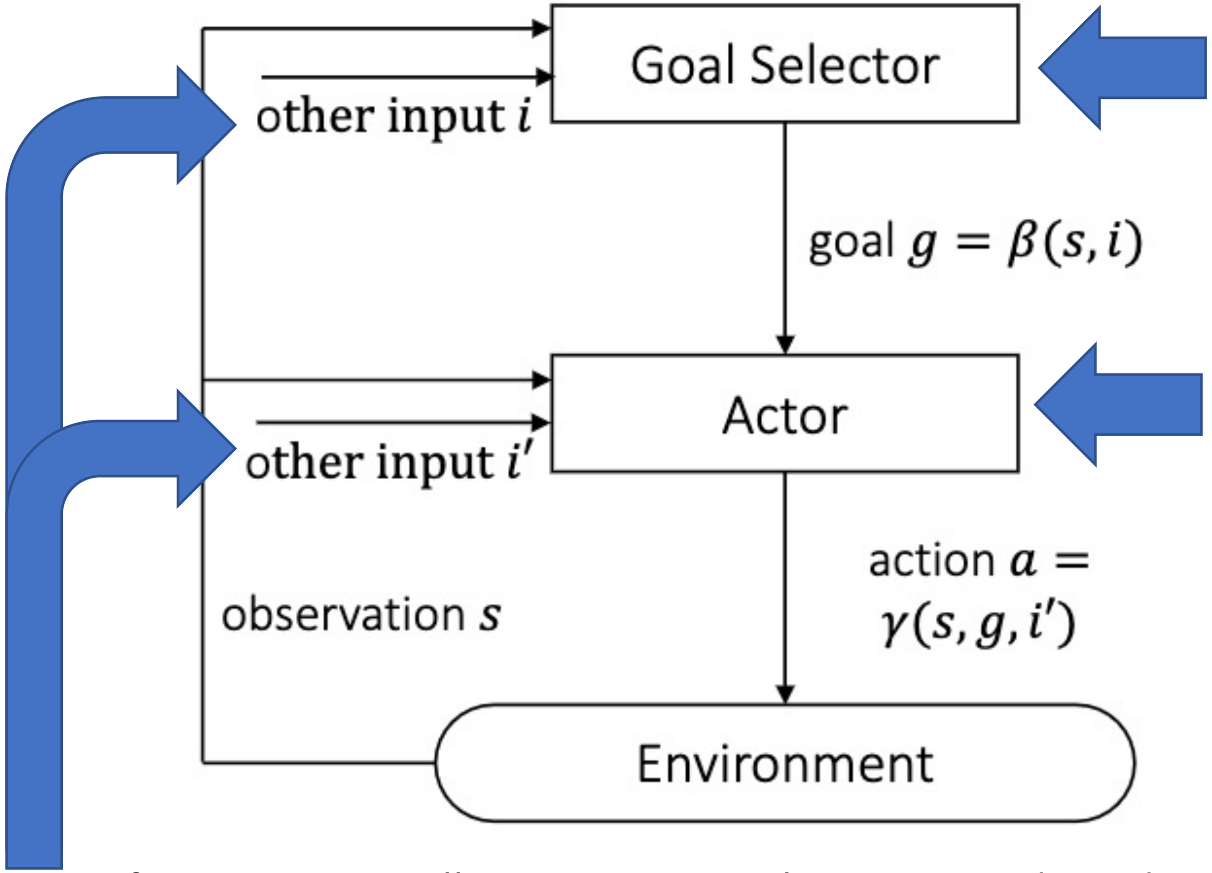
Disclaimer: Any opinion, findings and conclusions or recommendations in this presentation are those of the authors and are not necessarily those of NSF

The Goal Act Architecture

The Goal Act Architecture

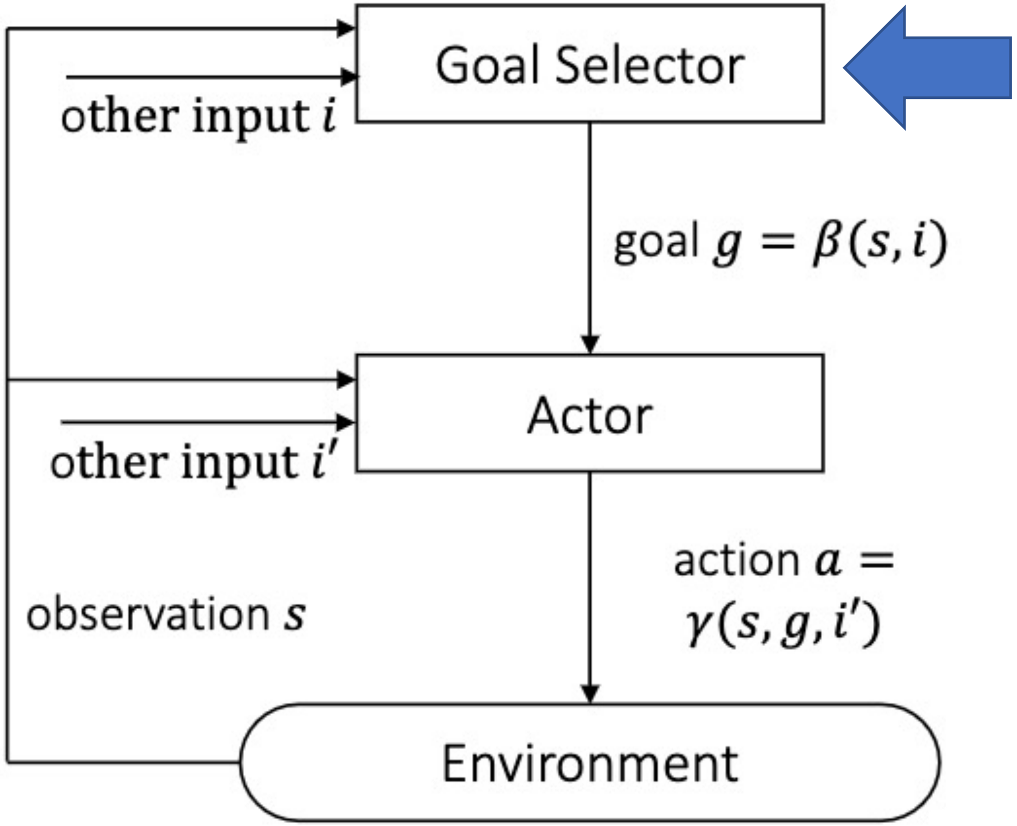


The Goal Act Architecture



A direct **extension** of HDQN -T. D. Kulkarni, K. Narasimhan, A. Saeedi, and J. Tenenbaum, “Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation,” in Advances in neural information processing systems, 2016, pp. 3675–3683

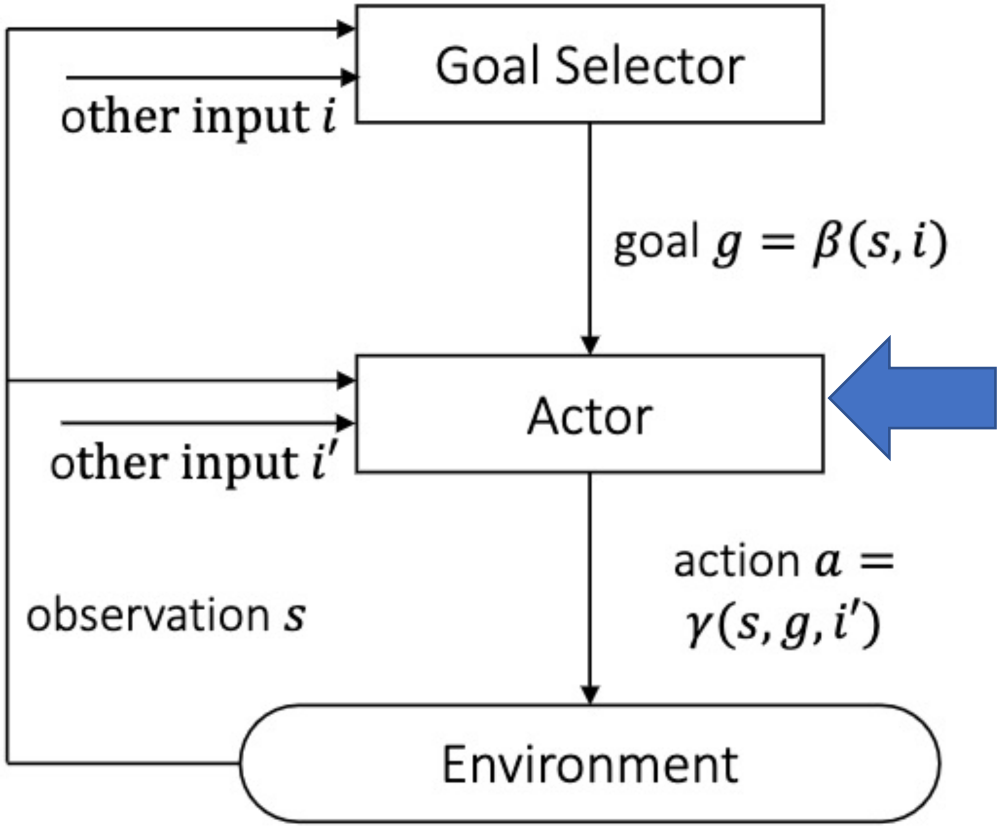
The Goal Act Architecture



The goal selector

- receives as input the state s and additional inputs i
- generates the next goal g

The Goal Act Architecture



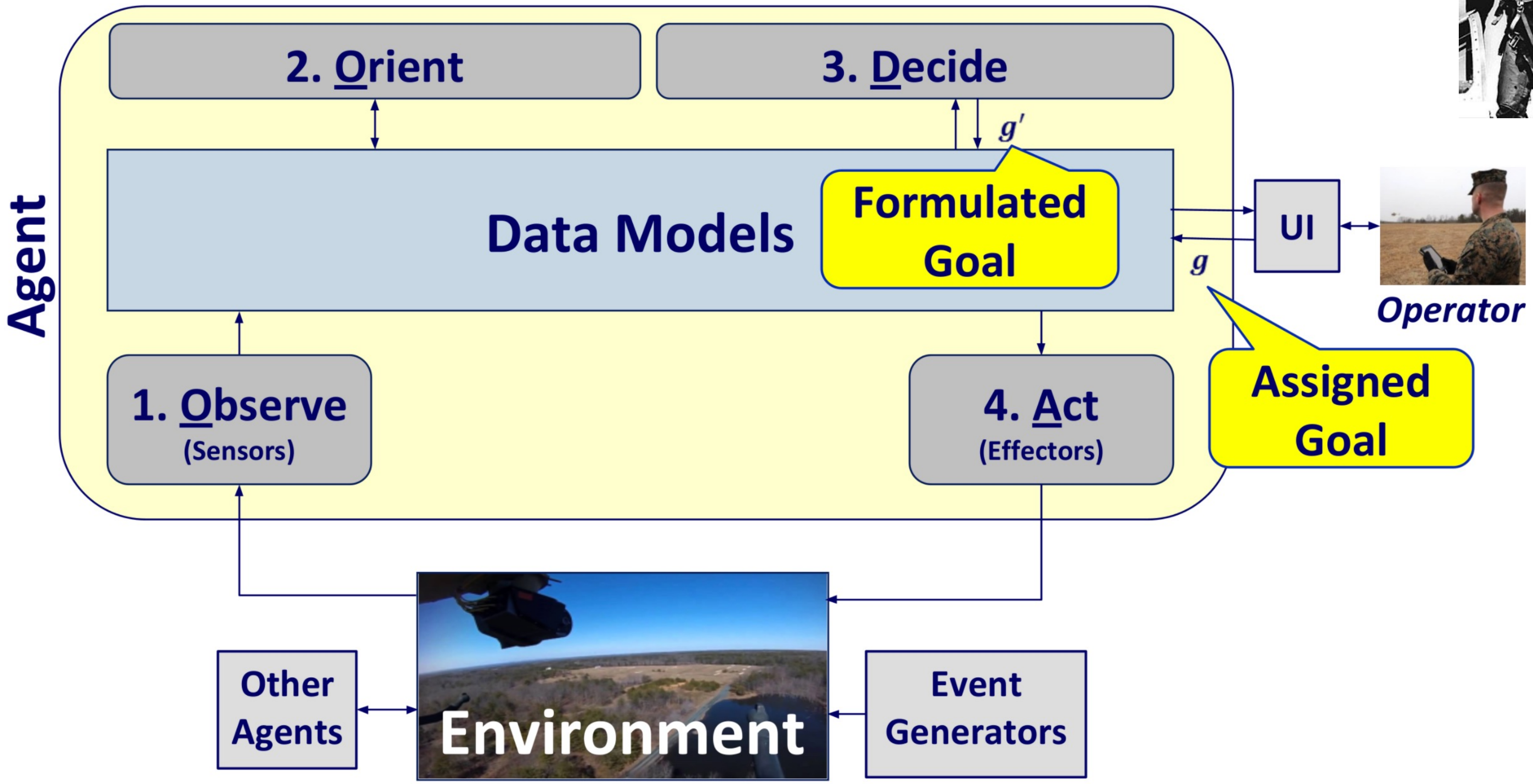
The Actor takes action based on:

- the goal g
- the state s
- possibly other inputs i’.

How is the GoalAct architecture situated relative to the goal reasoning paradigm?

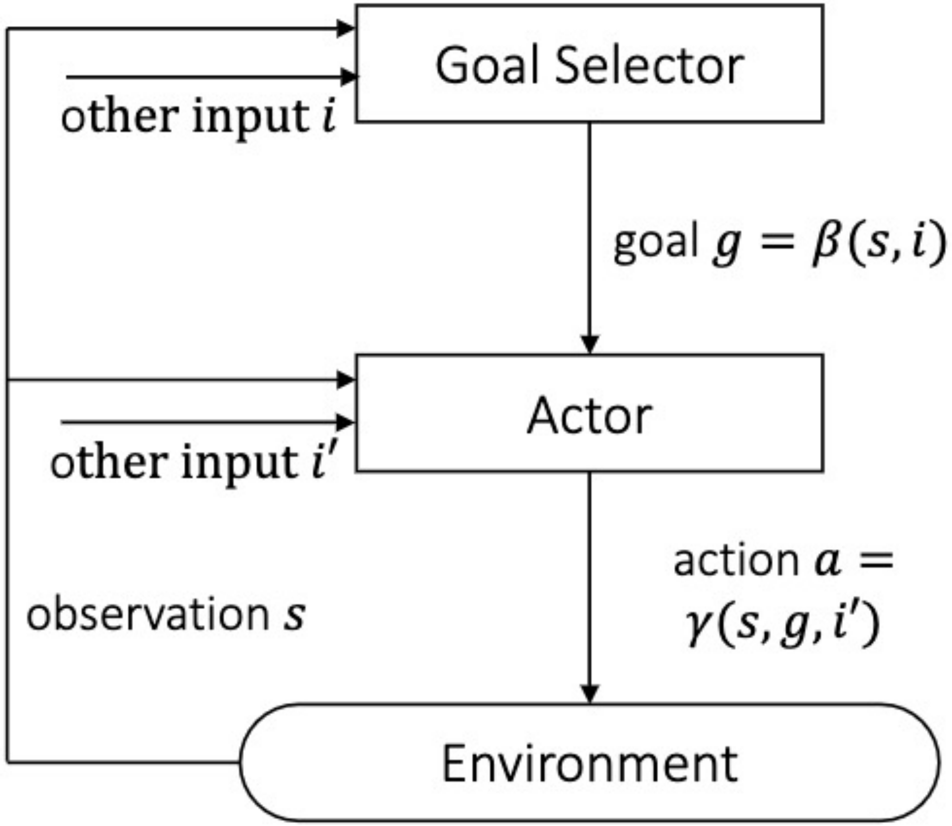
Goal Reasoning

John Boyd, USAF

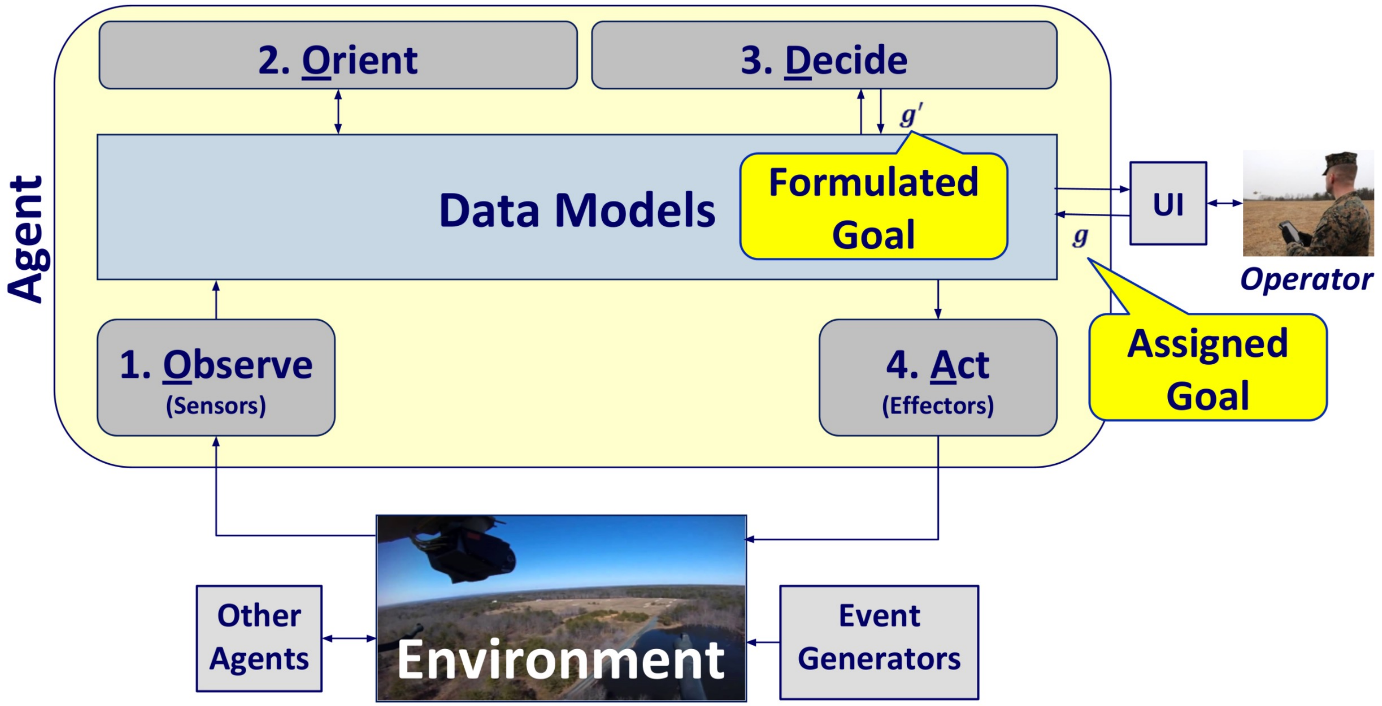


Aha, David W. "Goal reasoning: Foundations, emerging applications, and prospects." *AI Magazine* 39, no. 2 (2018): 3-24.

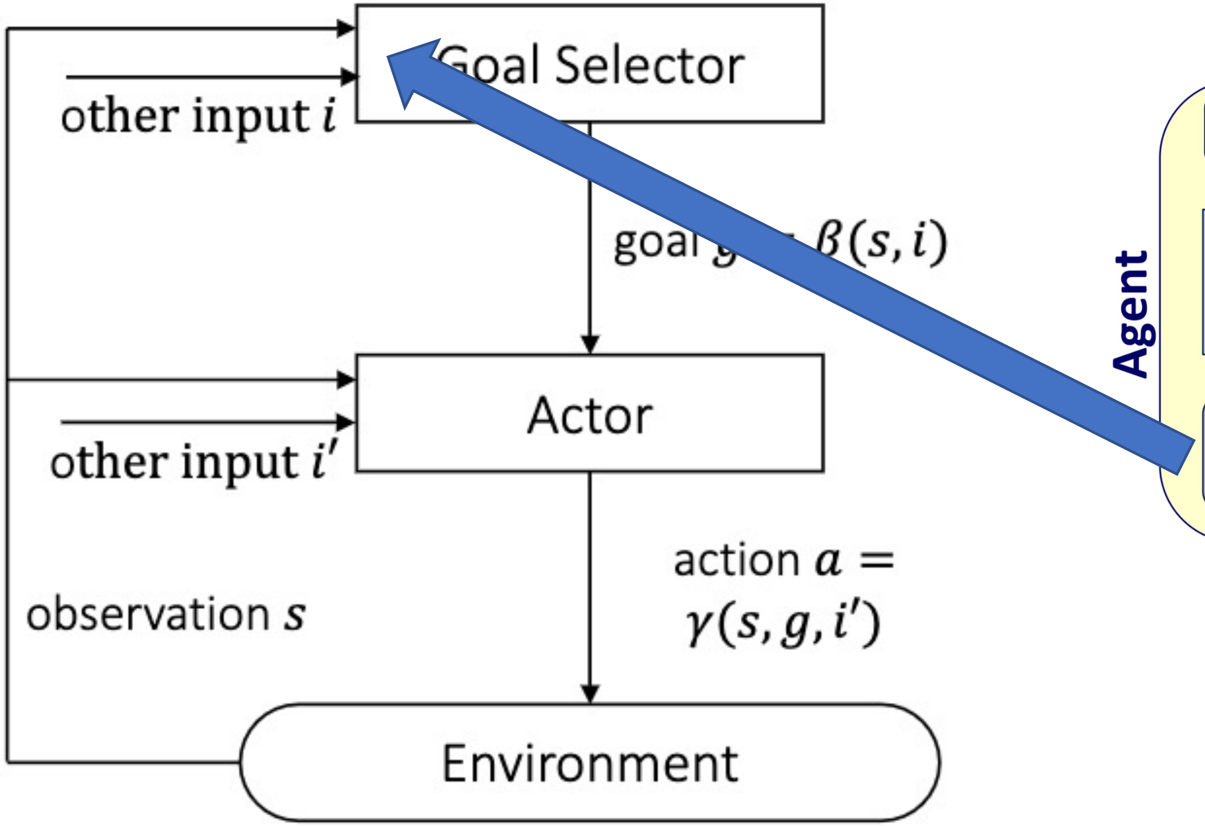
GoalAct Architecture



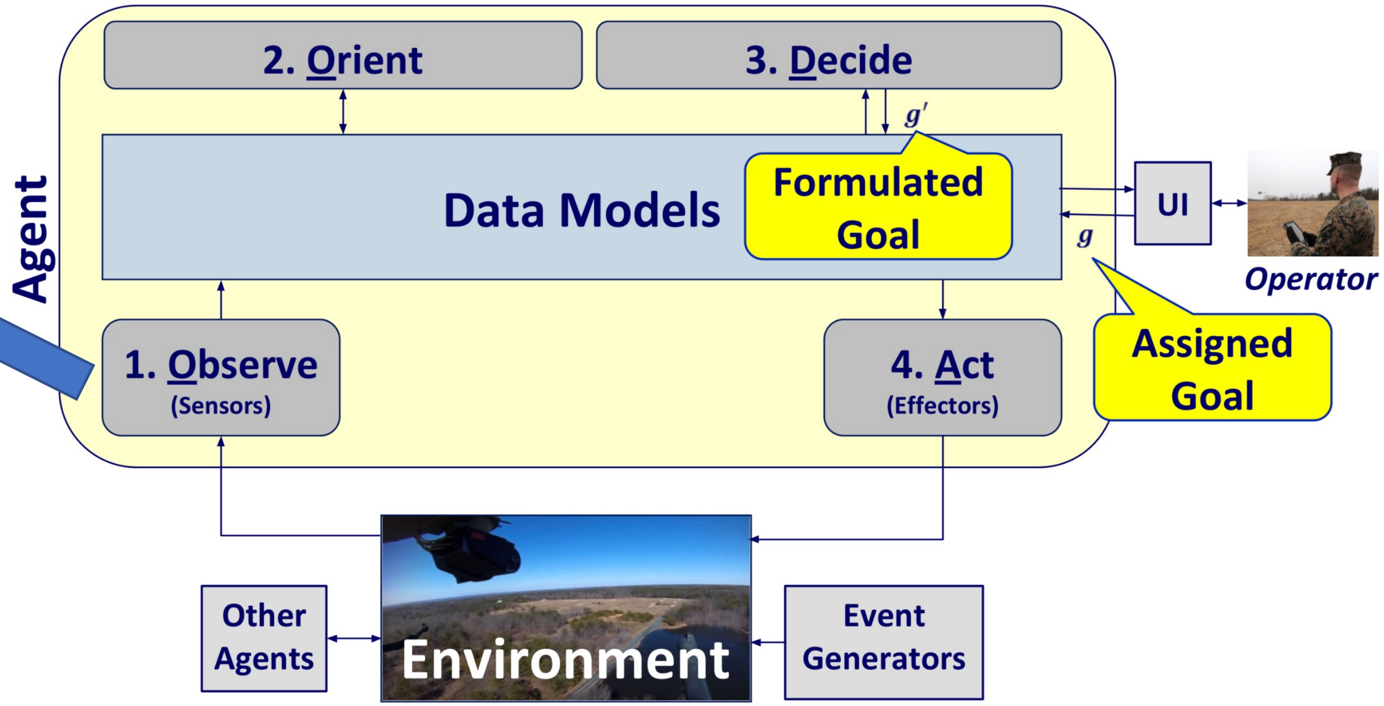
Goal Reasoning



H-DQN's Like Architectures

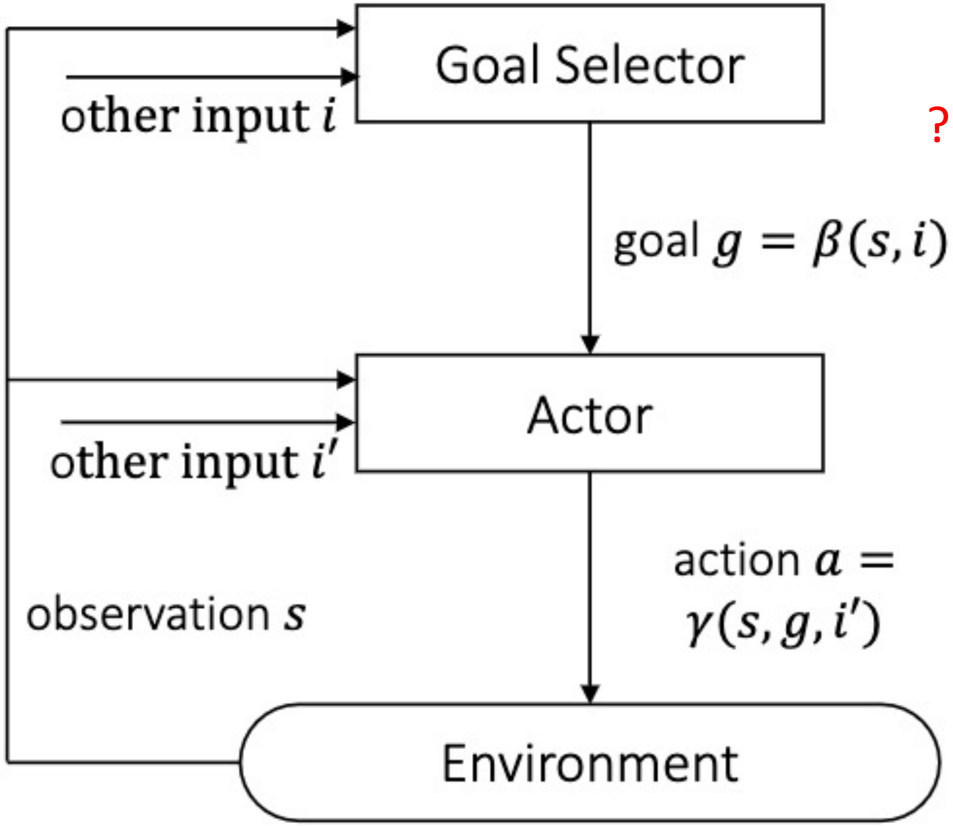


Goal Reasoning



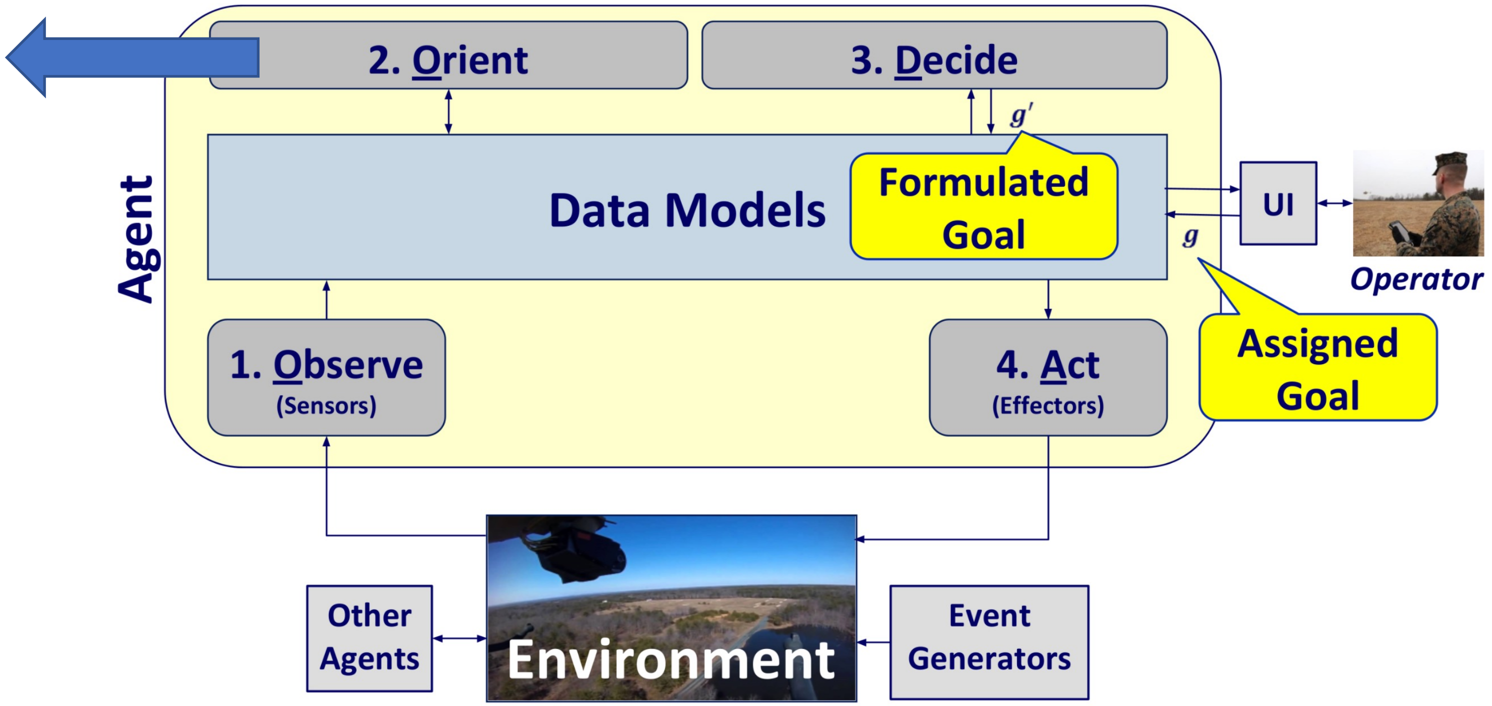
- Observe refers to raw sensor readings from the environment
- Represented in GoalAct as a state s, possibly partially observed
- The state can be an an snapshot image of what the agent sees or a simple vector representing a map annotated with the current location of the agent

H-DQN's Like Architectures



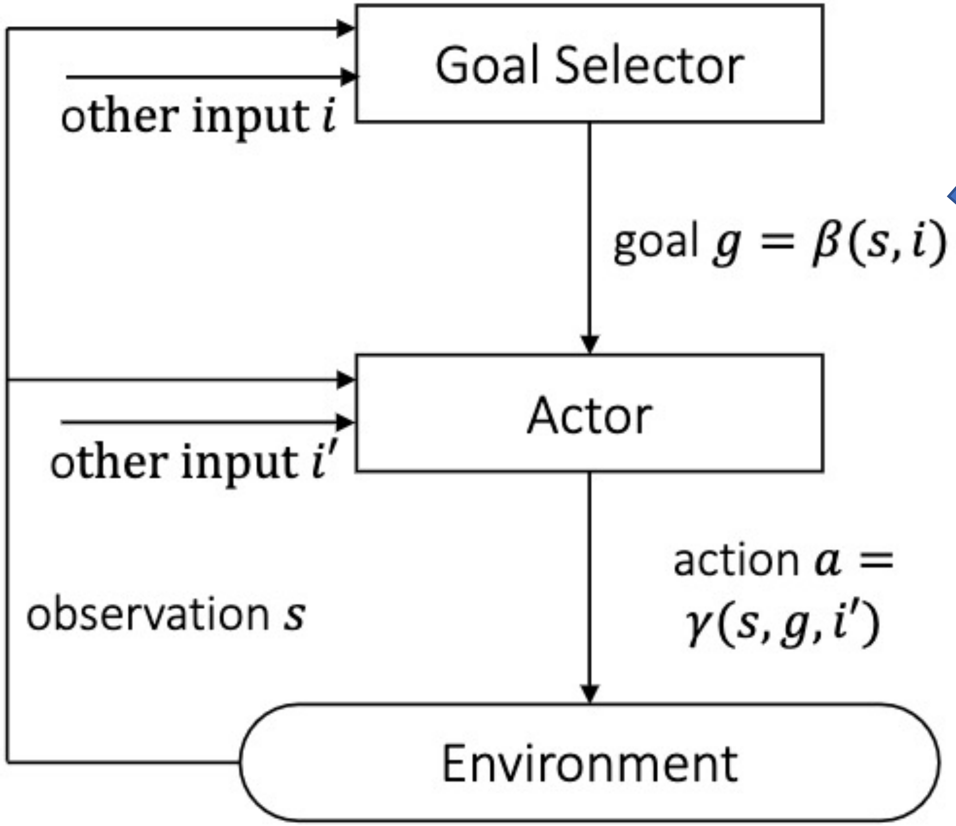
?

Goal Reasoning

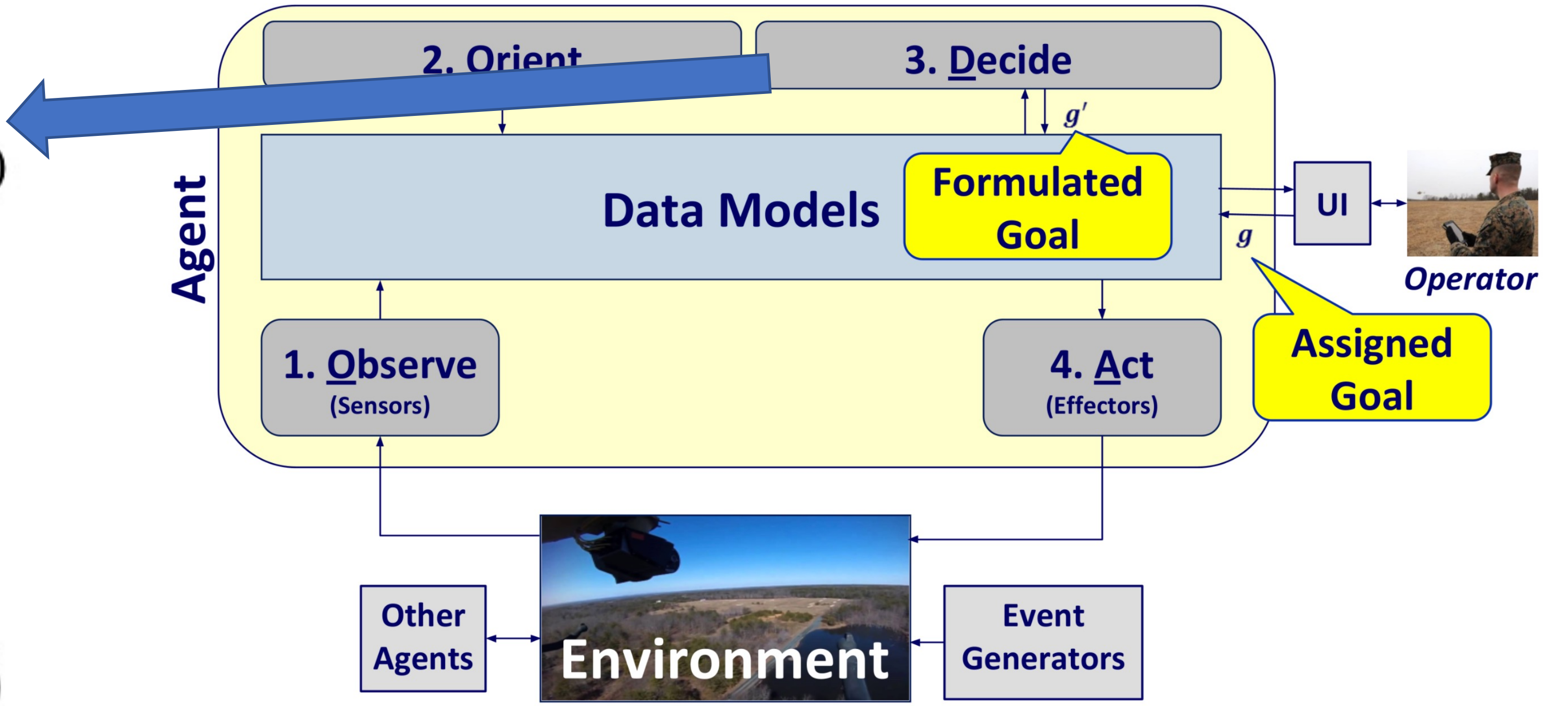


- The orient steps focuses the agent's attention. For example:
  - Goal-driven autonomy generates expectations X of the outcome of its actions
  - These expectations are then matched against the observed state, o(s) (e.g., X = o(S))
  - This determines if the goals are to be changed
- The orient step is completely bypassed in existing implementations of GoalAct.

H-DQN's Like Architectures

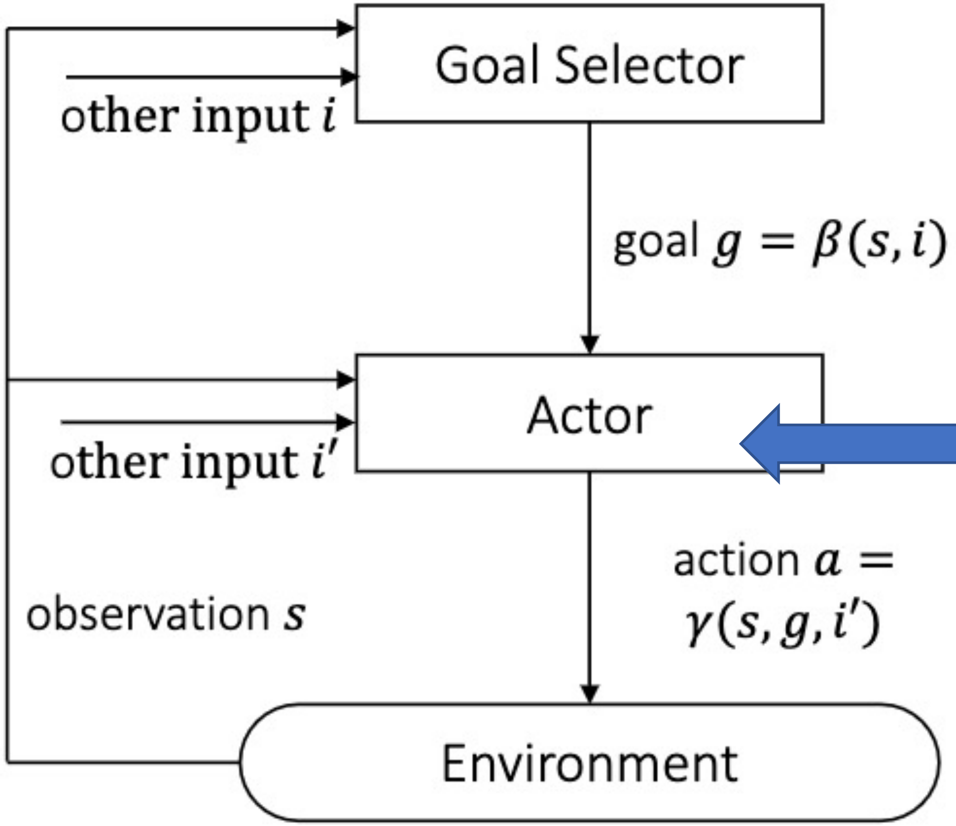


Goal Reasoning

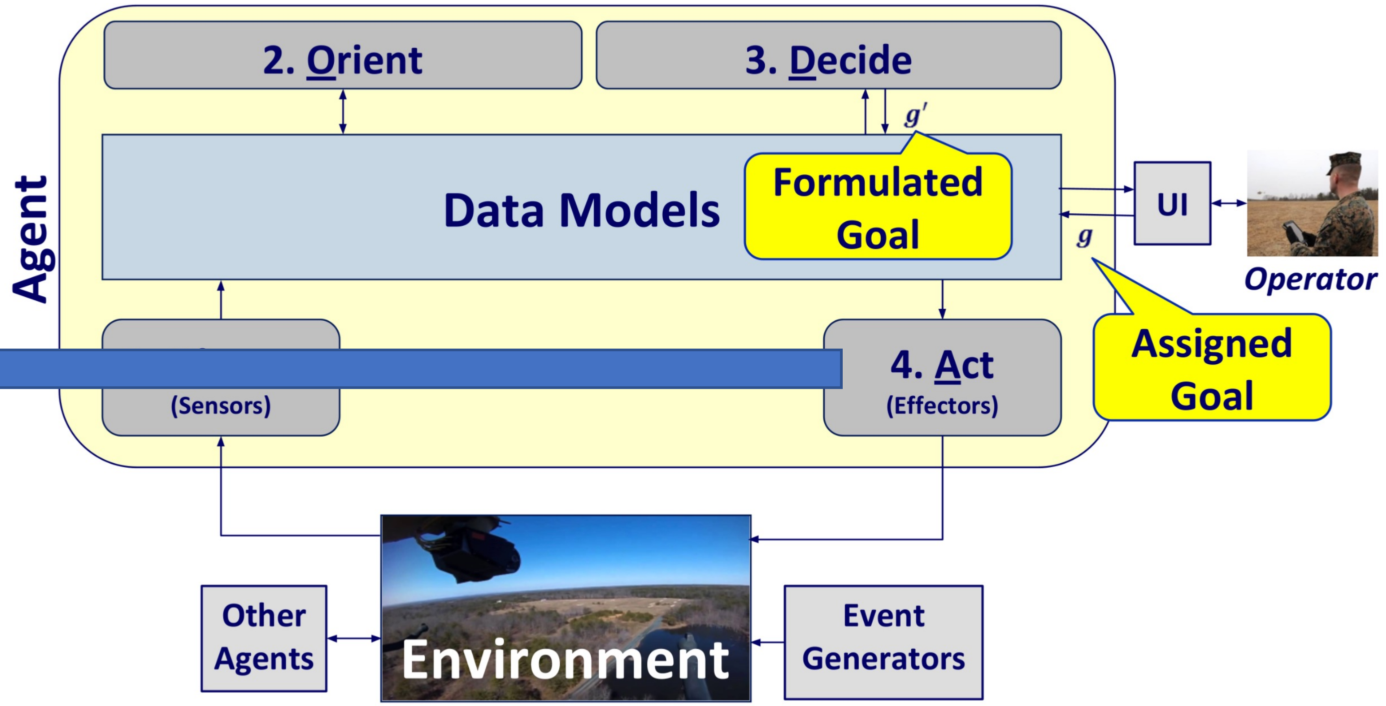


- Decide selecting which goals to manage and among those managed which goals to pursue.
- In GoalAct: there is only one goal that is pursued at any point of time.
- The list of all possible goals remains fixed for the lifetime of the system.
- In principle, any changes in the list will require to re-learn by running the system on all episodes experienced so far

H-DQN's Like Architectures



Goal Reasoning



- Once a goal has been decided, the control of the agent is given to the Actor, which takes actions in order to achieve the goal.
- The effects of each action causes the environment to transition from one state to the next.
- Current implementations: The Actor continues to take actions until either the goal is achieved or a fixed number of steps is taken.

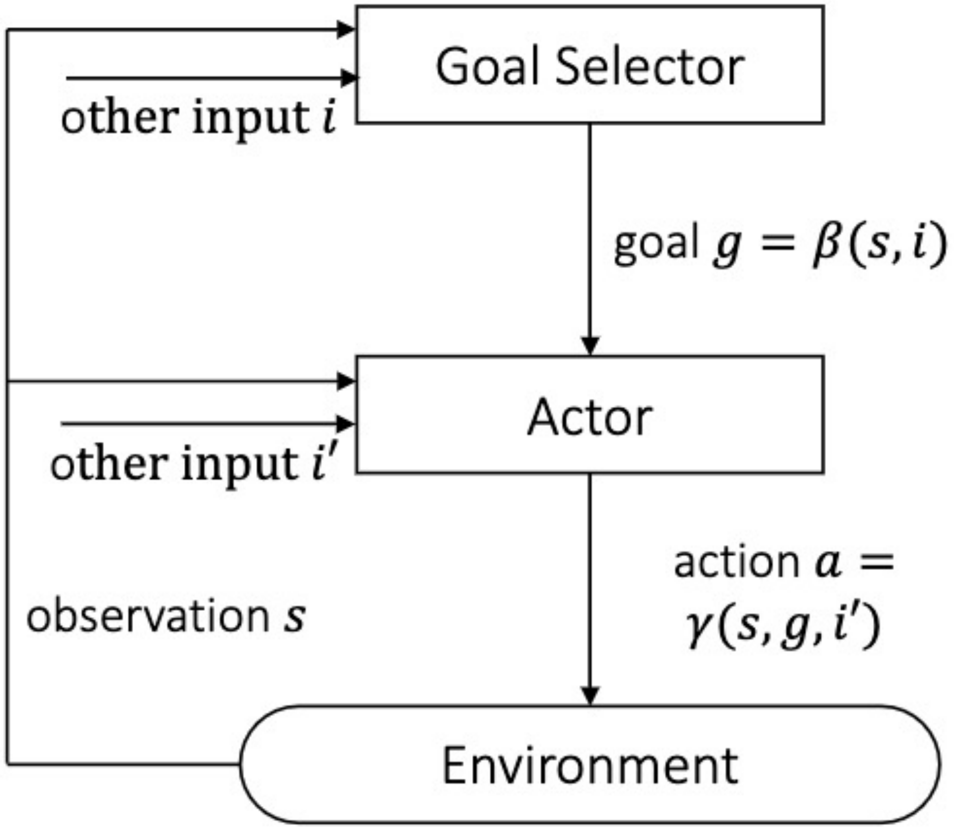
Taxonomy of Goals

Conditions	Condition 1	Condition 2
Declarative vs Procedural	✓	✓
Concrete vs Abstract	✓	
Static-time vs Durative	✓	🚫
Knowledge vs regular goals		✓
Interruptible vs non-i goals		✓

van Riemsdijk, M. B.; Dastani, M.; and Winikoff, M. 2008. Goals In Agent Systems: A Unifying Framework. In Proceedings of the Seventh International Conference on Autonomous Agents and Multi-Agent Systems, 713– 720. New York

What kinds of outputs is the GoalAct architecture capable of generating?

GoalAct Architecture



Outputs: system generates a sequence:

$s_0\, s_1\ \ldots\ s_n$

Cox et al.' goal formulation rule:

$$\beta(s, g) \rightarrow g'$$

Cox, Michael, Dustin Dannenhauer, and Sravya Kondrakunta. "Goal operations for cognitive systems." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 31. No. 1. 2017.

$$\beta(s, g) \rightarrow g'$$ (Cox et al.)

$$\beta(s, \_) \rightarrow g$$ (H-DQN)

$$\beta(s, g) \rightarrow g' \qquad \text{(Cox et al.)}$$

$$\beta(s, \_) \rightarrow g \qquad \text{(H-DQN)}$$

$$\beta(s, \tilde{s}) \rightarrow g \qquad \text{(FUNS)}$$

“list of previously visited states”

Vezhnevets, A. S.; Osindero, S.; Schaul, T.; Heess, N.; Jaderberg, M.; Silver, D.; and Kavukcuoglu, K. 2017. Feudal networks for hierarchical reinforcement learning. In Proceedings of the 34th International Conference on Machine Learning-Volume 70, 3540–3549

$$\boxed{\beta(s, i) \rightarrow g}$$ (Direct generalization)

$$\beta(s, g) \rightarrow g'$$ (Cox et al.)

$$\beta(s, \_) \rightarrow g$$ (H-DQN)

$$\beta(s, \tilde{s}) \rightarrow g$$ (FUNS)

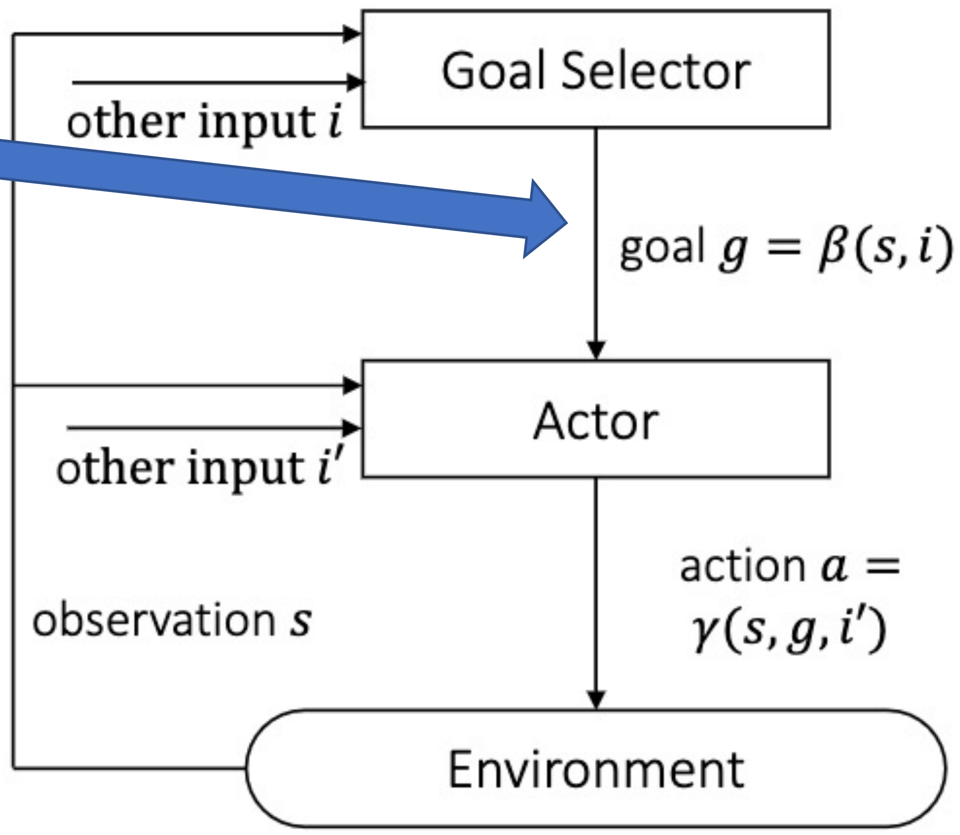
“list of previously visited states”

Vezhnevets, A. S.; Osindero, S.; Schaul, T.; Heess, N.; Jaderberg, M.; Silver, D.; and Kavukcuoglu, K. 2017. Feudal networks for hierarchical reinforcement learning. In Proceedings of the 34th International Conference on Machine Learning-Volume 70, 3540–3549

$$\beta(s, i) \to g$$

GoalAct Architecture

$$\beta(s, i) \rightarrow g$$



$$\beta(s, i) \to g$$

$$\beta(s, i) \rightarrow g$$

$$i\ s\ \langle \text{GOAL} \rangle \rightarrow i\ s\ g\ \langle \text{ACT} \rangle$$

$$\beta(s, i) \to g$$

$$i\ s\ \langle \text{GOAL} \rangle \to i\ s\ g\ \langle \text{ACT} \rangle$$

$$\beta(s, i) \rightarrow g$$

$$i \; s \; \langle \mathrm{GOAL} \rangle \rightarrow i \; s \; g \; \langle \mathrm{ACT} \rangle$$

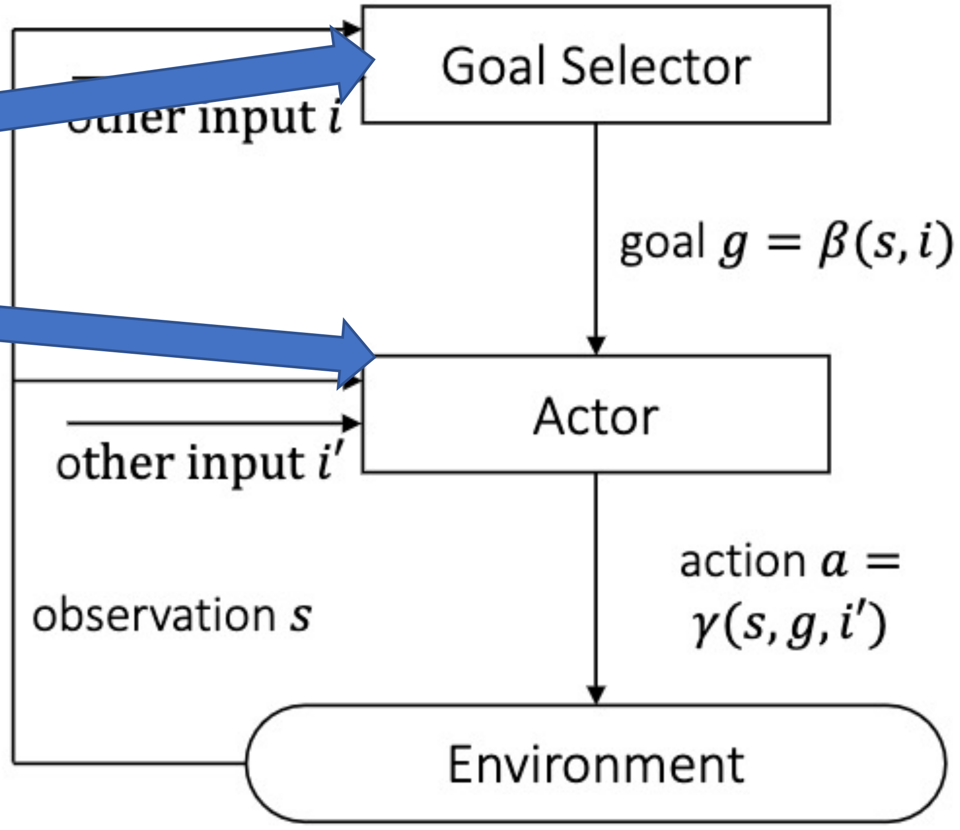
$$\beta(s, i) \rightarrow g$$

$$i \; s \; \langle \mathrm{GOAL} \rangle \rightarrow i \; s \; g \; \langle \mathrm{ACT} \rangle$$

GoalAct Architecture

$$\beta(s, i) \rightarrow g$$

$$i\ s\ \langle \mathrm{GOAL} \rangle \rightarrow i\ s\ g\ \langle \mathrm{ACT} \rangle$$



$$\beta(s, i) \rightarrow g$$

$$i \ s \ \langle \text{GOAL} \rangle \rightarrow i \ s \ g \ \langle \text{ACT} \rangle$$

(H-DQN)

$$\beta(s, i) \rightarrow g$$

$$\beta(s, \_) \rightarrow g$$

$$i\ s\ \langle\text{GOAL}\rangle \rightarrow i\ s\ g\ \langle\text{ACT}\rangle$$

$$s\langle\text{GOAL}\rangle \rightarrow sg\langle\text{ACT}\rangle$$

(H-DQN)

$$\beta(s, i) \rightarrow g$$

$$i\ s\ \langle\text{GOAL}\rangle \rightarrow i\ s\ g\ \langle\text{ACT}\rangle$$

$$\beta(s, \_) \rightarrow g.$$

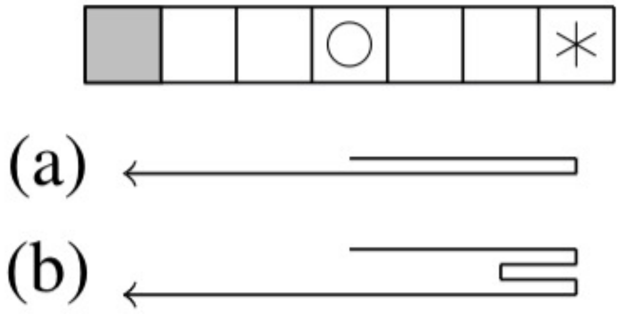
$$s\langle\text{GOAL}\rangle \rightarrow sg\langle\text{ACT}\rangle$$

(FUNS)

$$\beta(s, \tilde{s}) \rightarrow g$$

$$\tilde{s}s\langle\text{GOAL}\rangle \rightarrow \tilde{s}sg\langle\text{ACT}\rangle.$$

Formal System Around these Rules
(exemplified for H-DQN)



Formal System Around these Rules – **Basic GoalAct**

(exemplified for H-DQN)

(H-DQN)

$$\beta(s, \_) \rightarrow g$$

$$s\langle\text{GOAL}\rangle \rightarrow sg\langle\text{ACT}\rangle$$

$$S \rightarrow s_3\langle\text{GOAL}\rangle$$

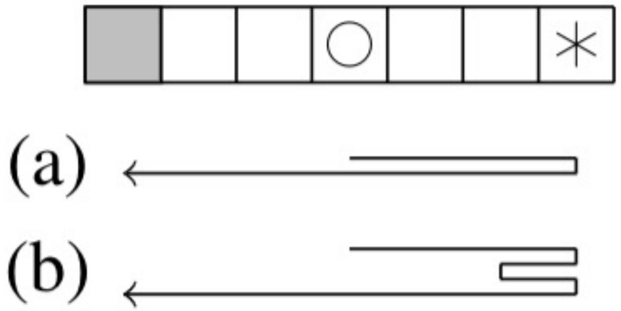
$$s_3\langle\text{GOAL}\rangle \rightarrow s_3 g_6\langle\text{ACT}\rangle$$

$$s_6\langle\text{GOAL}\rangle \rightarrow s_6 g_0\langle\text{ACT}\rangle$$

$$s_i g_j\langle\text{ACT}\rangle \rightarrow s_i s_j\langle\text{GOAL}\rangle \quad (1 \leq i, j \leq 6 \wedge i \neq j)$$

$$s_i g_0\langle\text{ACT}\rangle \rightarrow s_i s_0 \quad (1 \leq i \leq 6)$$

$$s_i g_i\langle\text{ACT}\rangle \rightarrow s_i g_i\langle\text{ACT}\rangle \quad (1 \leq i \leq 6)$$



Formal System Around these Rules – **Basic GoalAct**

(exemplified for H-DQN)

$$\boxed{S \rightarrow s_3\langle\text{GOAL}\rangle}$$

One such rule for each starting state

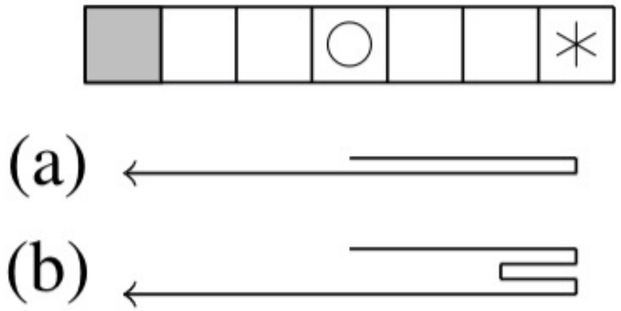
$$s_3\langle\text{GOAL}\rangle \rightarrow s_3 g_6\langle\text{ACT}\rangle$$

$$s_6\langle\text{GOAL}\rangle \rightarrow s_6 g_0\langle\text{ACT}\rangle$$

$$s_i g_j\langle\text{ACT}\rangle \rightarrow s_i s_j\langle\text{GOAL}\rangle \quad (1 \leq i, j \leq 6 \wedge i \neq j)$$

$$s_i g_0\langle\text{ACT}\rangle \rightarrow s_i s_0 \quad (1 \leq i \leq 6)$$

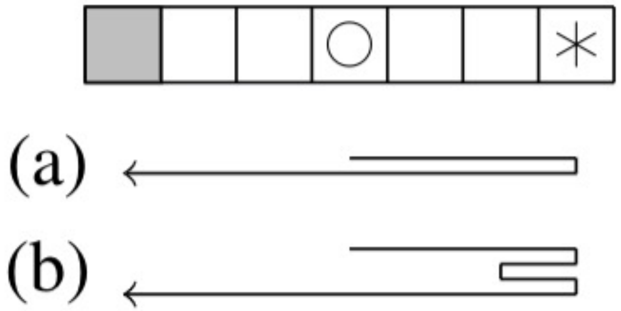
$$s_i g_i\langle\text{ACT}\rangle \rightarrow s_i g_i\langle\text{ACT}\rangle \quad (1 \leq i \leq 6)$$



Formal System Around these Rules – **Basic GoalAct**
(exemplified for H-DQN)

$$
\begin{aligned}
S &\rightarrow s_3\langle\text{GOAL}\rangle \\
s_3\langle\text{GOAL}\rangle &\rightarrow s_3 g_6\langle\text{ACT}\rangle \\
s_6\langle\text{GOAL}\rangle &\rightarrow s_6 g_0\langle\text{ACT}\rangle \\
&\boxed{s_i g_j\langle\text{ACT}\rangle \rightarrow s_i s_j\langle\text{GOAL}\rangle \quad (1 \leq i, j \leq 6 \wedge i \neq j)} \\
s_i g_0\langle\text{ACT}\rangle &\rightarrow s_i s_0 \quad (1 \leq i \leq 6) \\
s_i g_i\langle\text{ACT}\rangle &\rightarrow s_i g_i\langle\text{ACT}\rangle \quad (1 \leq i \leq 6)
\end{aligned}
$$

When state reached satisfies the goal



Formal System Around these Rules – **Basic GoalAct**

(exemplified for H-DQN)

$$S \rightarrow s_3\langle\text{GOAL}\rangle$$
$$s_3\langle\text{GOAL}\rangle \rightarrow s_3 g_6\langle\text{ACT}\rangle$$
$$s_6\langle\text{GOAL}\rangle \rightarrow s_6 g_0\langle\text{ACT}\rangle$$
$$s_i g_j\langle\text{ACT}\rangle \rightarrow s_i s_j\langle\text{GOAL}\rangle \quad (1 \leq i, j \leq 6 \wedge i \neq j)$$
$$\boxed{s_i g_0\langle\text{ACT}\rangle \rightarrow s_i s_0 \quad (1 \leq i \leq 6)}$$
$$s_i g_i\langle\text{ACT}\rangle \rightarrow s_i g_i\langle\text{ACT}\rangle \quad (1 \leq i \leq 6)$$

When state reached
is a terminal state

Formal System Around these Rules – **Basic GoalAct**

(exemplified for H-DQN)

$$S \rightarrow s_3\langle\text{GOAL}\rangle$$

$$s_3\langle\text{GOAL}\rangle \rightarrow s_3 g_6\langle\text{ACT}\rangle$$

$$s_6\langle\text{GOAL}\rangle \rightarrow s_6 g_0\langle\text{ACT}\rangle$$

$$s_i g_j\langle\text{ACT}\rangle \rightarrow s_i s_j\langle\text{GOAL}\rangle \quad (1 \leq i, j \leq 6 \wedge i \neq j)$$

$$s_i g_0\langle\text{ACT}\rangle \rightarrow s_i s_0 \quad (1 \leq i \leq 6)$$

$$\boxed{s_i g_i\langle\text{ACT}\rangle \rightarrow s_i g_i\langle\text{ACT}\rangle \quad (1 \leq i \leq 6)}$$

Actor falls in an infinite loop

Formal System Around these Rules – **Basic GoalAct**

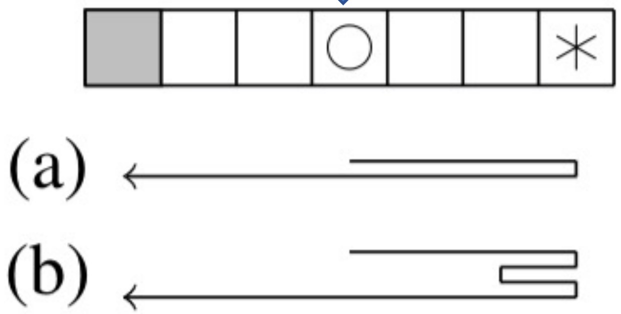
(exemplified for H-DQN)

$$S \xrightarrow{1} s_3\langle\text{GOAL}\rangle \xrightarrow{2} s_3 g_6\langle\text{ACT}\rangle \xrightarrow{4} s_3 s_6\langle\text{GOAL}\rangle \xrightarrow{3} s_3 s_6 g_0\langle\text{ACT}\rangle \xrightarrow{5} s_3 s_6 s_0$$

(H-DQN)

$$\beta(s, \_) \rightarrow g.$$

$$s\langle\text{GOAL}\rangle \rightarrow sg\langle\text{ACT}\rangle$$



Formal System Around these Rules – **Basic GoalAct**

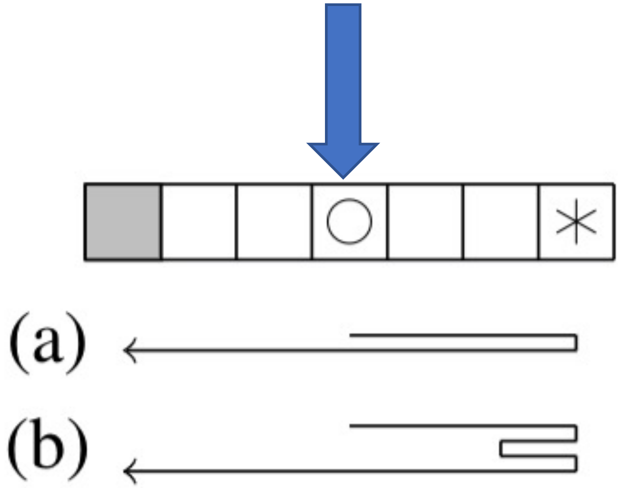
(exemplified for H-DQN)

(H-DQN)

$$\beta(s, \_) \to g.$$

$$s\langle\mathrm{GOAL}\rangle \to sg\langle\mathrm{ACT}\rangle$$

$$\begin{aligned}
S &\xrightarrow{1} s_3\langle\mathrm{GOAL}\rangle \\
&\xrightarrow{2} s_3 g_6\langle\mathrm{ACT}\rangle \\
&\xrightarrow{4} s_3 s_6\langle\mathrm{GOAL}\rangle \\
&\xrightarrow{3} s_3 s_6 g_0\langle\mathrm{ACT}\rangle \\
&\xrightarrow{5} s_3 s_6 s_0
\end{aligned}$$



Formal System Around these Rules – **Basic GoalAct**

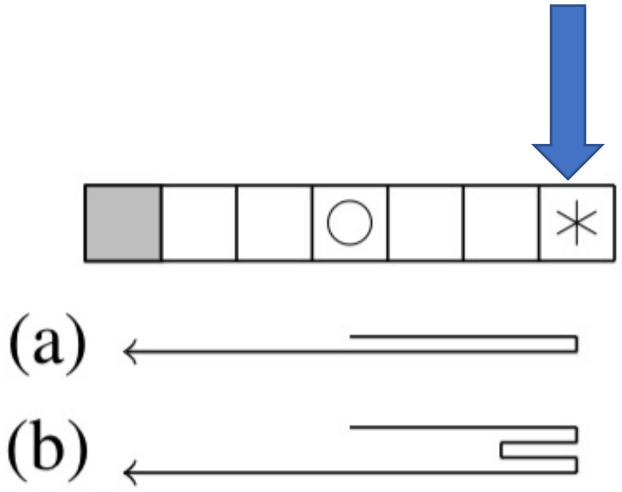
(exemplified for H-DQN)

(H-DQN)

$$\beta(s, \_) \rightarrow g$$

$$s\langle\text{GOAL}\rangle \rightarrow sg\langle\text{ACT}\rangle$$

$$\begin{aligned} S &\xrightarrow{1} s_3\langle\text{GOAL}\rangle \\ &\xrightarrow{2} s_3 g_6\langle\text{ACT}\rangle \\ &\xrightarrow{4} s_3 s_6\langle\text{GOAL}\rangle \\ &\xrightarrow{3} s_3 s_6 g_0\langle\text{ACT}\rangle \\ &\xrightarrow{5} s_3 s_6 s_0 \end{aligned}$$



Formal System Around these Rules – **Basic GoalAct**

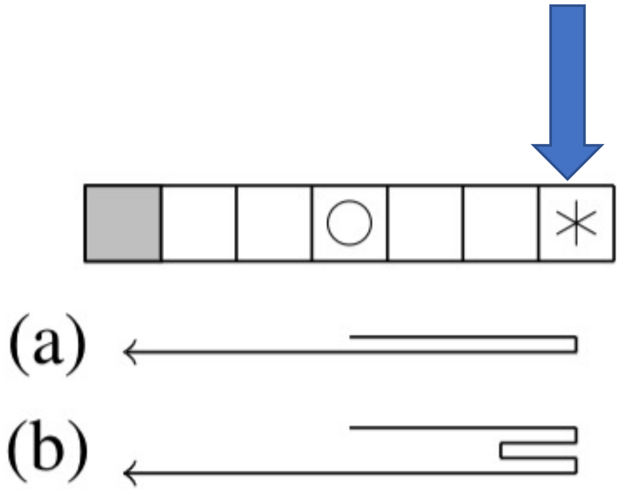
(exemplified for H-DQN)

(H-DQN)

$$\beta(s, \_) \to g.$$

$$s\langle\text{GOAL}\rangle \to sg\langle\text{ACT}\rangle$$

$$\begin{aligned}
S &\xrightarrow{1} s_3\langle\text{GOAL}\rangle \\
&\xrightarrow{2} s_3 g_6\langle\text{ACT}\rangle \\
&\xrightarrow{4} s_3 s_6\langle\text{GOAL}\rangle \\
&\xrightarrow{3} s_3 s_6 g_0\langle\text{ACT}\rangle \\
&\xrightarrow{5} s_3 s_6 s_0
\end{aligned}$$



Formal System Around these Rules – **Basic GoalAct**

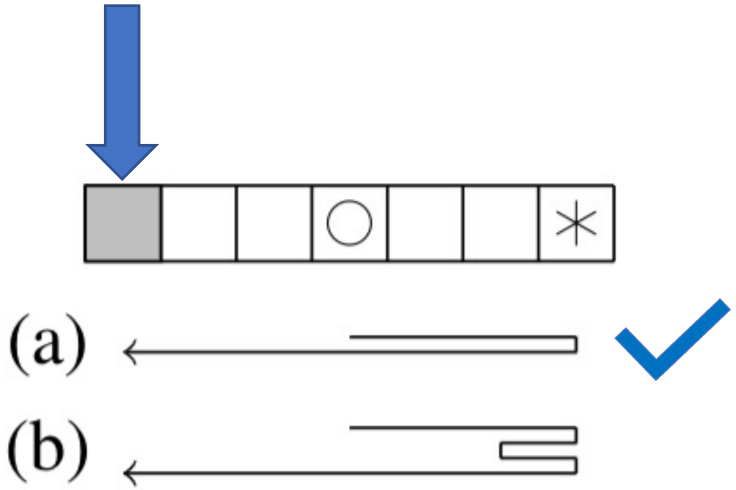
(exemplified for H-DQN)

$$
\begin{aligned}
S &\xrightarrow{1} s_3\langle \mathrm{GOAL}\rangle \\
&\xrightarrow{2} s_3 g_6\langle \mathrm{ACT}\rangle \\
&\xrightarrow{4} s_3 s_6\langle \mathrm{GOAL}\rangle \\
&\xrightarrow{3} s_3 s_6 g_0\langle \mathrm{ACT}\rangle \\
&\xrightarrow{5} s_3 s_6 s_0
\end{aligned}
$$

(H-DQN)

$$\beta(s, \_) \rightarrow g$$

$$s\langle \mathrm{GOAL}\rangle \rightarrow sg\langle \mathrm{ACT}\rangle$$



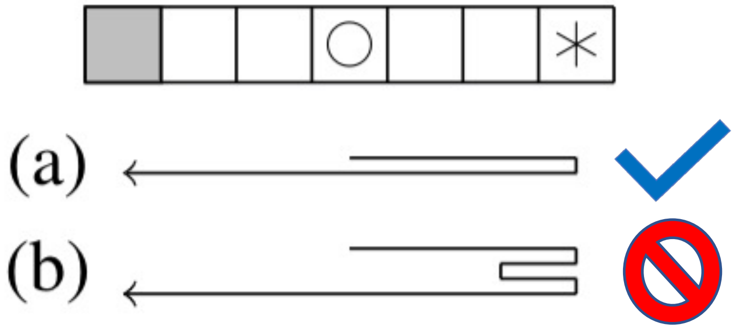
Formal System Around these Rules – **Basic GoalAct**

(exemplified for H-DQN)

(H-DQN)

$$\beta(s, \_) \rightarrow g$$

$$s\langle \mathrm{GOAL} \rangle \rightarrow sg\langle \mathrm{ACT} \rangle$$



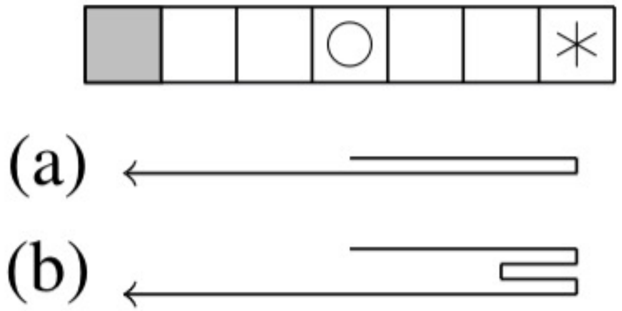
Formal System Around these Rules – **Mnemonic GoalAct**
(exemplified for FUNS)

(FUNS)

$$\beta(s, \tilde{s}) \rightarrow g$$

$$\tilde{s}s\langle\text{GOAL}\rangle \rightarrow \tilde{s}sg\langle\text{ACT}\rangle.$$

$$S \rightarrow s_3\langle\text{GOAL}\rangle$$
$$s_3\langle\text{GOAL}\rangle \rightarrow s_3g_6\langle\text{ACT}\rangle$$
$$s_3s_6\langle\text{GOAL}\rangle \rightarrow s_3s_6g_5\langle\text{ACT}\rangle$$
$$s_3s_6s_5\langle\text{GOAL}\rangle \rightarrow s_3s_6s_5g_6\langle\text{ACT}\rangle$$
$$s_3s_6s_5s_6\langle\text{GOAL}\rangle \rightarrow s_3s_6s_5s_6g_0\langle\text{ACT}\rangle$$
$$s_ig_j\langle\text{ACT}\rangle \rightarrow s_is_j\langle\text{GOAL}\rangle \quad (1 \leq i, j \leq 6 \wedge i \neq j)$$
$$s_ig_0\langle\text{ACT}\rangle \rightarrow s_is_0 \quad (1 \leq i \leq 6)$$
$$s_ig_i\langle\text{ACT}\rangle \rightarrow s_ig_i\langle\text{ACT}\rangle \quad (1 \leq i \leq 6)$$



Formal System Around these Rules – **Mnemonic GoalAct**

(exemplified for FUNS)

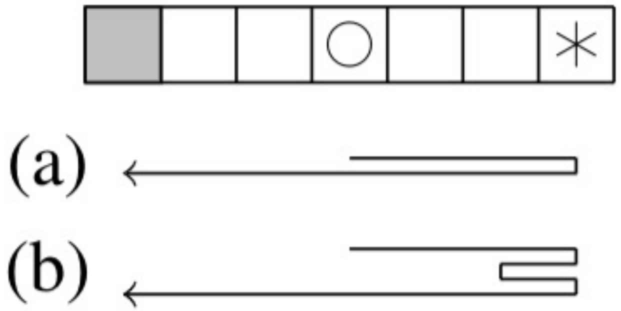
(FUNS)

$$S \rightarrow s_3\langle \text{GOAL} \rangle$$

$$\begin{aligned}
s_3\langle \text{GOAL} \rangle &\rightarrow s_3 g_6 \langle \text{ACT} \rangle \\
s_3 s_6\langle \text{GOAL} \rangle &\rightarrow s_3 s_6 g_5 \langle \text{ACT} \rangle \\
s_3 s_6 s_5\langle \text{GOAL} \rangle &\rightarrow s_3 s_6 s_5 g_6 \langle \text{ACT} \rangle \\
s_3 s_6 s_5 s_6\langle \text{GOAL} \rangle &\rightarrow s_3 s_6 s_5 s_6 g_0 \langle \text{ACT} \rangle
\end{aligned}$$

$$\begin{aligned}
s_i g_j \langle \text{ACT} \rangle &\rightarrow s_i s_j \langle \text{GOAL} \rangle \quad (1 \leq i, j \leq 6 \wedge i \neq j) \\
s_i g_0 \langle \text{ACT} \rangle &\rightarrow s_i s_0 \quad (1 \leq i \leq 6) \\
s_i g_i \langle \text{ACT} \rangle &\rightarrow s_i g_i \langle \text{ACT} \rangle \quad (1 \leq i \leq 6)
\end{aligned}$$

Same as before



Formal System Around these Rules – **Mnemonic GoalAct**

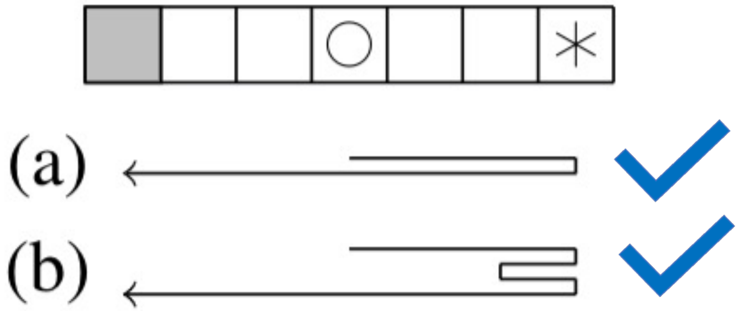
(exemplified for FUNS)

(FUNS)

$$\beta(s, \tilde{s}) \rightarrow g$$

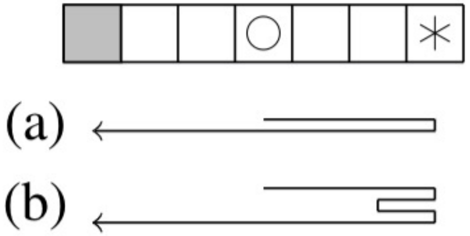
$$\tilde{s}s\langle \mathrm{GOAL} \rangle \rightarrow \tilde{s}sg\langle \mathrm{ACT} \rangle.$$

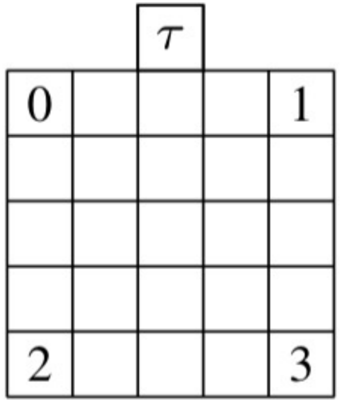
$$\begin{aligned}
S &\rightarrow s_3\langle \mathrm{GOAL} \rangle \\
s_3\langle \mathrm{GOAL} \rangle &\rightarrow s_3 g_6\langle \mathrm{ACT} \rangle \\
s_3 s_6\langle \mathrm{GOAL} \rangle &\rightarrow s_3 s_6 g_5\langle \mathrm{ACT} \rangle \\
s_3 s_6 s_5\langle \mathrm{GOAL} \rangle &\rightarrow s_3 s_6 s_5 g_6\langle \mathrm{ACT} \rangle \\
s_3 s_6 s_5 s_6\langle \mathrm{GOAL} \rangle &\rightarrow s_3 s_6 s_5 s_6 g_0\langle \mathrm{ACT} \rangle \\
s_i g_j\langle \mathrm{ACT} \rangle &\rightarrow s_i s_j\langle \mathrm{GOAL} \rangle \quad (1 \leq i, j \leq 6 \wedge i \neq j) \\
s_i g_0\langle \mathrm{ACT} \rangle &\rightarrow s_i s_0 \quad (1 \leq i \leq 6) \\
s_i g_i\langle \mathrm{ACT} \rangle &\rightarrow s_i g_i\langle \mathrm{ACT} \rangle \quad (1 \leq i \leq 6)
\end{aligned}$$



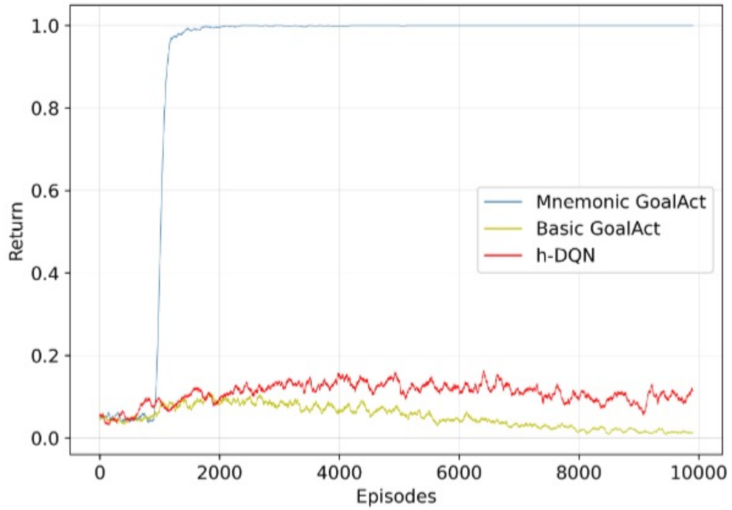
Results (1)

- We provide a characterization of the grammars generating the states sequences: k-mnemonic grammars
- Using this characterization we prove that Mnemonic GoalAct is strictly more expressive than Basic GoalAct
- This result is corroborated in empirical evaluation: corridor, Doom, Grid

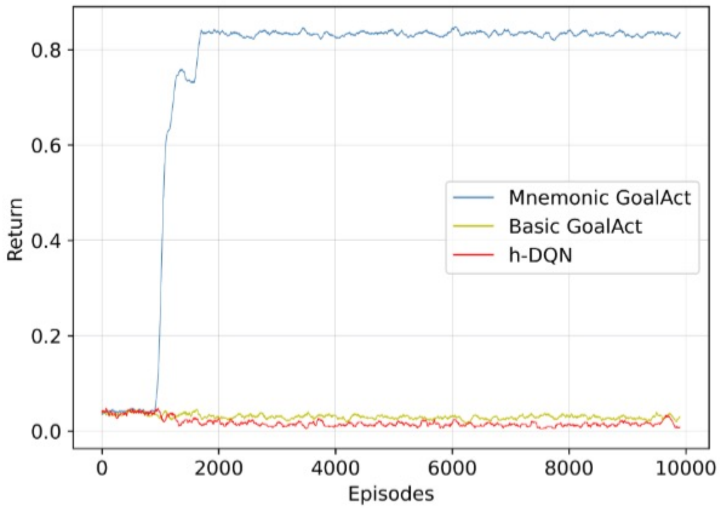




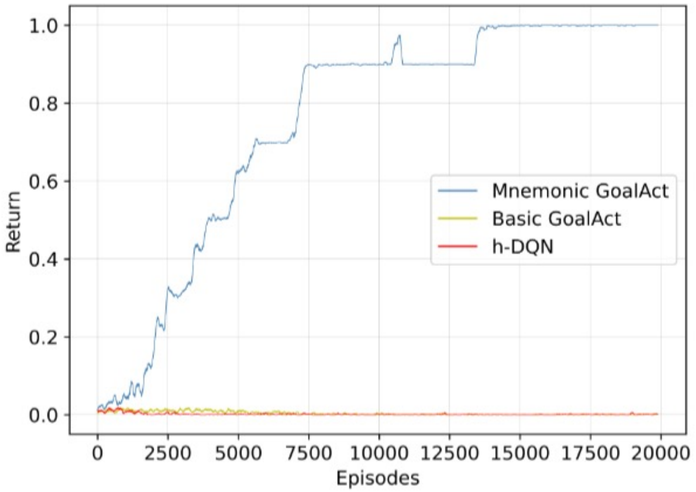
Results (2)



(a) Corridor



(b) Doom

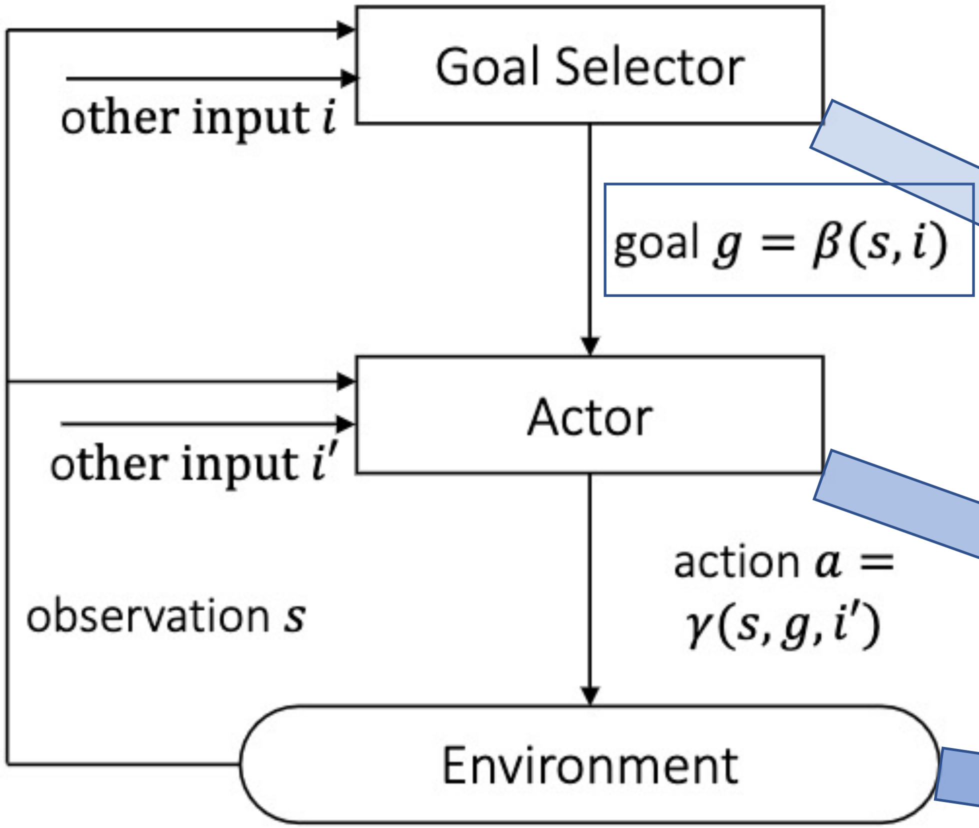


(c) Grid

Related Work - Inspirations

- DQN combines RL and DL
- DRQN combines RNN and DQN
- Options: a mechanism to jump between policies when certain states are reached
- Memory-based RL considers rewards based on states visited
- Subgoal learning: learns which goal to choose in a particular situation

Conclusion

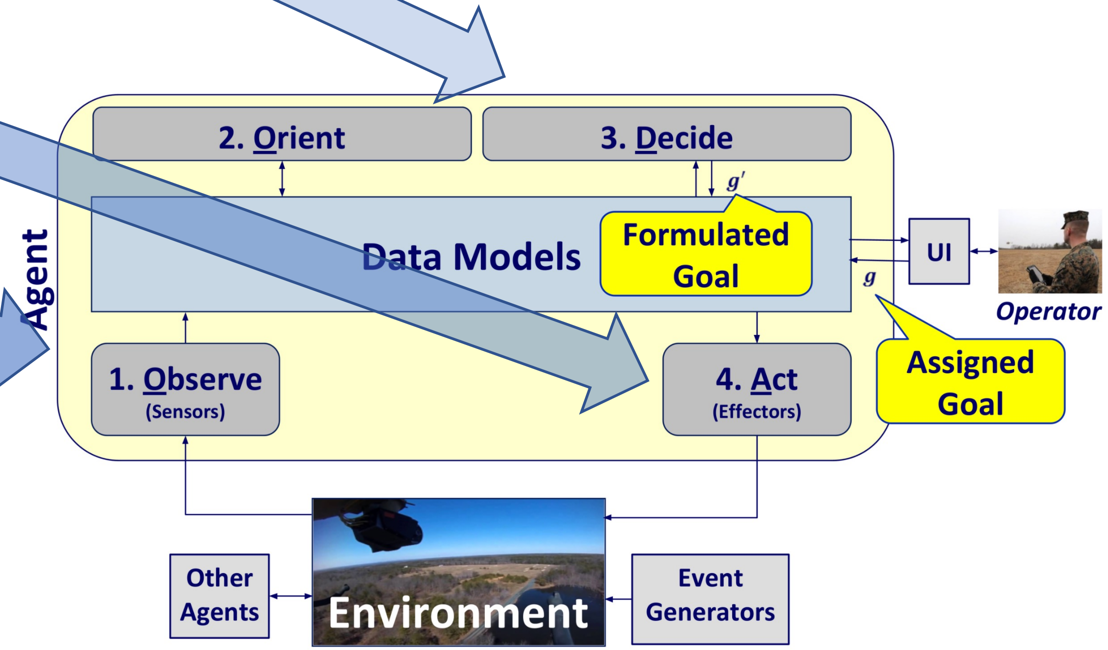


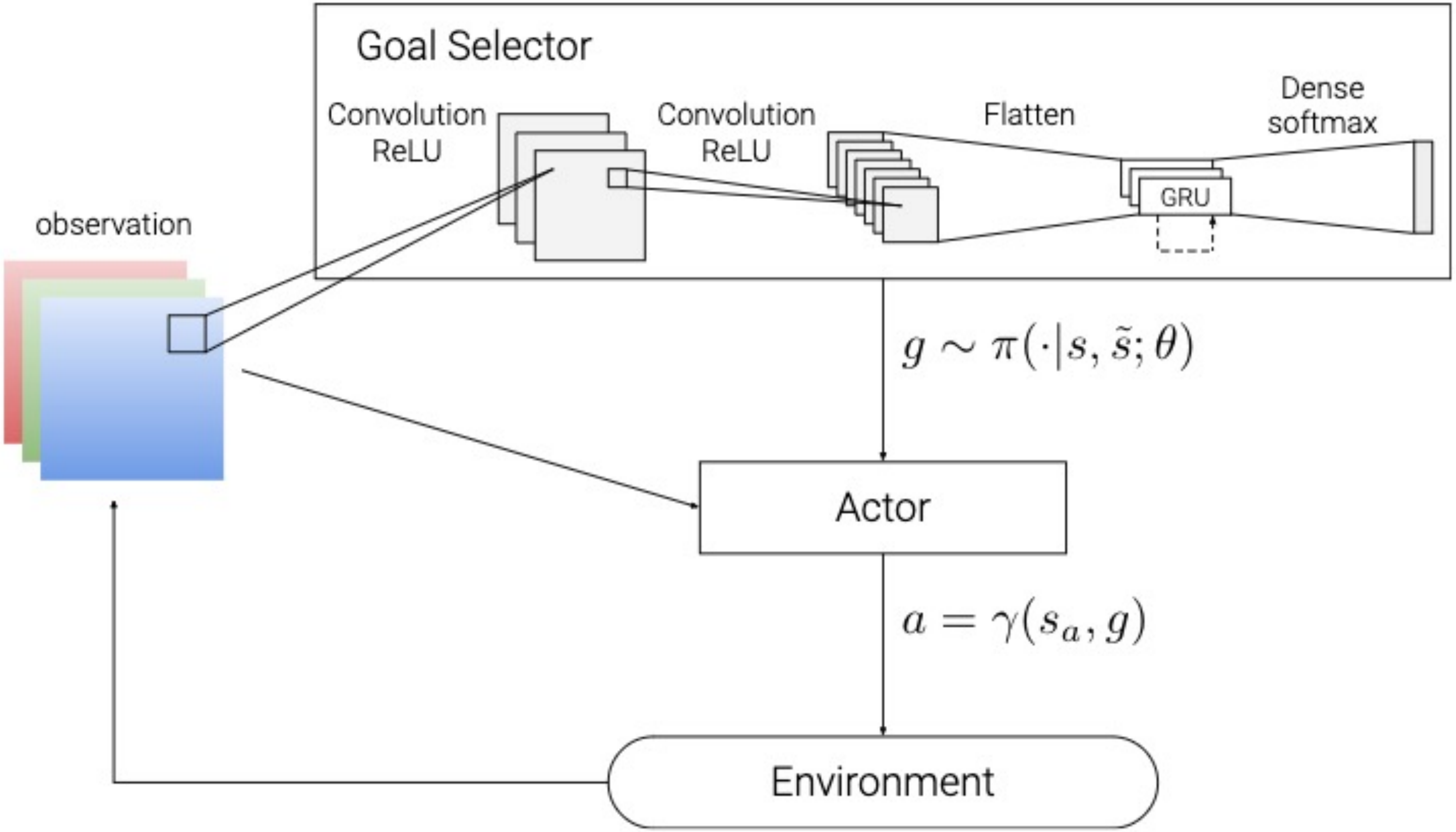
Use formal grammars to model the sequences generated by different systems:

(H-DQN) s <GOAL> → s g <ACT>

(FUNS) …s’s<GOAL> → …s’s<ACT>

Use these grammars to compare expressiveness (kinds of state sequences generated)




Goal Selector
Convolution
ReLU
Convolution
ReLU
Flatten
Dense
softmax
GRU
observation
$g \sim \pi(\cdot|s, \tilde{s}; \theta)$
Actor
$a = \gamma(s_a, g)$
Environment