# Represent and Infer Goals in Markov Games

**Yizhe Huang**
School of Intelligence Science and Technology
Peking University
szhyz@pku.edu.cn

## Abstract

Goals play a key factor in the agent's decision-making, and goals can be regarded as a good representation of an agent's behavior over a period of time. This essay introduces an approach to incorporate goals into the framework of Markov games. Additionally, it provides a methodology for executing goal inference within and across episodes within this representational context.

## 1 Introduction

Goals play a pivotal role in steering behavior. It is a concept ingrained in individuals from infancy [6]. The majority of human actions can be explained as goal-directed [3], necessitating their incorporation when modeling the behavioral decision-making of an agent.

In the realm of computational models, the choice of goal level yields distinct effects. Broadly speaking, a goal with great abstraction capabilities encapsulates an agent's behavioral patterns. At this juncture, the goal serves as a classifier, distinguishing, for instance, whether the agent is cooperative or competitive [9, 14]. A more specific goal, like navigating to a specific location, functions akin to a "macro-action." Here, the goal guides decision-making over a defined period, enabling planning within the goal space. To facilitate extensive sequence planning, goals act as a high-level layer in hierarchical planning [13]. No matter which level of goal is adopted, the overarching aim is to abstractly capture agent behavior over time. In this sense, learning goal seems to be learning a good representation of behavior.

Beyond using goals to help ourselves make decisions and plans, we can also infer other people's goals. This ability assists in establishing accurate and appropriate relationships with peers, informing action choices accordingly. This dynamic naturally gives rise to a multi-agent problem, broadening the scope of goal representation and inference in the context of Markov games.

## 2 Markov games with goals

In the exploration of agent decision-making, a widely employed modeling framework is the Markov Decision Process (MDP) [8]. As we transition to scenarios involving multiple agents, this framework seamlessly evolves into the realm of Markov Games [10]. Introducing the element of goals further extends this framework, giving rise to the concept of Markov Game with goals. This specialized extension is distinctly represented by the tuple $< N, S, \mathbf{A}, T, \mathbf{R}, \gamma, \mathbf{G} >$, where

- Agent $i \in N = \{1, 2, \cdots, n\}$.
- State space of agent $i$: $S_i = \{s_i\}$. $\mathbf{S} = S_1 \times S_2 \times \cdots \times S_n$.
- Action space of agent $i$: $A_i = \{a_i\}$. $\mathbf{A} = A_1 \times A_2 \times \cdots \times A_n$.
- Transition function $T : S \times \mathbf{A} \times S \to [0, 1]$. After agents take the joint action $\boldsymbol{a}_{1:n}$ the state of the environment will transit from $s$ to $s'$ with probability $T(s'|s, \boldsymbol{a}_{1:n})$.
- Reward function $R_i : S \times \mathbf{A} \to \mathbb{R}$, which denotes the immediate reward received by agent $i$ after joint action $\boldsymbol{a}_{1:n}$ is taken on state $s \in S$.

- Discount factor for future rewards: $\gamma$.
- Goal space of agent $i$: $G_i = \{g_i\}$. $\mathbf{G} = G_1 \times G_2 \times \cdots \times G_n$.

$\pi_i : S \times A_i \rightarrow [0, 1]$ denotes agent $i$'s policy, specifying the probability $\pi_i(a_i|s)$ that agent $i$ chooses action $a_i$ at state $s$. For any two agents $i$ and $j$, $j$'s true goal is inaccessible to $i$. However, $i$ can infer $j$'s goal based on its action sequence. Specifically, $i$ maintains a belief over $j$'s goals, $b_{ij} : G_j \rightarrow [0, 1]$, which is a probability distribution over $G_j$.

In the proposed representation, the goal appears as an autonomous entity, distinct from the original Markov Game. A critical exploration involves examining the potential connections between goals, states, and actions. One intriguing approach is to conceptualize the goal as a composite of trajectories, denoted as $g = \{\tau\}$, where $\tau =< s^0, a^0, s^1, a^1, \cdots, s^{terminal} >$ - with the superscript indicating the current timestep. For instance, reaching a specific state $s$, a common goal, can be expressed as $\{\tau|\tau.s^{terminal} = s\}$.

While the current representation provides a comprehensive framework, a notable challenge arises when the trajectories corresponding to the same goal may have many constraints, potentially leading to dimension curse. Nevertheless, given that goals inherently offer a powerful abstraction of agent behavior, their constraints typically remain manageable, easing the computational burden. This indicates that the detailed trajectory representation of goals may, in this context, introduce some redundancy.

## 3 Infer other's goals in Markov games

In the setting of multi-agent reinforcement learning (MARL), we will go through multiple episodes, each resembling a Markov game. The task of inferring goals introduces two distinct types of inference: goal inference within the episode and goal inference between episodes.

The process of goal inference within an episode can be calculated via Bayesian updates [2, 1], aligning seamlessly with the principles of teleological reasoning [4]. Specifically, in episode $K$, agent $i$'s belief about agent $j$'s goals at time $t$, $b_{ij}^{K,t}(g_j)$, is updated according to:

$$\begin{aligned} b_{ij}^{K,t+1}(g_j) &= Pr(g_j\ |s^{K,0:t+1}, a_j^{K,0:t}) \\ &= \frac{Pr(g_j|s^{K,0:t}, a_j^{K,0:t-1})Pr(a_j^{K,t}|s^{K,0:t}, a_j^{K,0:t-1}, g_j)Pr(s^{K,t+1}|s^{K,0:t}, a_j^{K,0:t}, g_j)}{Pr(s^{K,t+1}, a_j^{K,t}|s^{K,0:t}, a_j^{K,0:t-1})} \\ &= \frac{b_{ij}^{K,t}(g_j)Pr_i(a_j^{K,t}|s^{K,0:t}, g_j)}{\int_{g\in G_j} b_{ij}^{K,t}(g)Pr_i(a_j^{K,t}|s^{K,0:t}, g)}, \end{aligned} \tag{1}$$

As we discuss in Sec. 2, goals typically exhibit simplicity and possess high abstraction capabilities. Consequently, they often manifest as discrete entities with a limited and manageable number, rendering Eq. (1) tractable.

Another issue is the goal inference between episodes. The agent may choose different goals due to different initial states. We hope to give a more accurate prior $b_{ij}^{K,0}(g_j)$ based on the experience of past episodes. One of the simplest methods is Monte Carlo estimation:

$$b_{ij}^{K,0}(g_j) = (1 - \frac{1}{K})b_{ij}^{K-1,0}(g_j) + \frac{1}{K}\mathbf{1}(g_j^{K-1} = g_j) \tag{2}$$

Given the dynamic nature of an opponent's goal preferences, employing updates with fixed weights emerges as a potentially advantageous strategy. This approach facilitates a more effective tracking of the opponent's evolving goal preferences, providing stability and consistency in our efforts to understand and respond to their latest objectives:

$$b_{ij}^{K,0}(g_j) = (1 - \alpha)b_{ij}^{K-1,0}(g_j) + \alpha\mathbf{1}(g_j^{K-1} = g_j) \tag{3}$$

A drawback of Equation 3 lies in the necessity of knowing the goal accomplished by agent $j$ in the preceding episode during the updating process. This requirement mandates that our model possesses the capability to parse the opponent's trajectory towards the goal, which is an assumption with potential challenges. However, if we have confidence in the accuracy of our goal inference for agent $j$ in the previous episode, we can substitute $b_{ij}^{K-1,T_{max}}(g_j)$ for $\mathbf{1}(g_j^{K-1} = g_j)$:

$$b_{ij}^{K,0}(g_j) = (1-\alpha)b_{ij}^{K-1,0}(g_j) + \alpha b_{ij}^{K-1,T_{max}}(g_j) \tag{4}$$

However, none of the above methods directly take into account the initial state of the new episode. In fact, when we model agent $j$ accurately, we may opt to directly ascertain its goal for the current episode based on the initial state $Pr(g_j|s^{K,0})$—a process that can be facilitated through supervised learning. Simultaneously, the likelihood term in Eq. (1) can also be refined through supervised learning, contingent on the availability of ground-truth goal labels during training. The challenge of discovering goals without the aid of ground-truth labels remains a noteworthy issue that often garners significant attention in research [12, 5, 15].

## 4 Discussion

The focus here delves into a singular computational approach aimed at modeling the objectives within the landscape of a Markov game. There's a growing emphasis among decision intelligence scholars on integrating goals into computations. Research is actively exploring goal-conditioned policy modeling [11], offering substantial utility in both decision-making and opponent modeling.

In various scenarios, the completion of a task inherently implies the existence of a goal, whether explicitly considered in decision-making or not. This signifies that irrespective of whether an agent's decision-making module encompasses explicit goals or more broadly, a mental state, one can conceptualize the agent as possessing a mental state and consequently draw inferences. But what do you infer when inferring agents that have no mental state? It may be a reasonable representation and abstraction of the opponent's behavior, just like humans interpret intentions from the movements of inanimate geometric figures [7].

## References

[1] Anonymous. Planning with theory of mind for few-shot adaptation in sequential social dilemmas. In *Submitted to The Twelfth International Conference on Learning Representations*, 2023. URL `https://openreview.net/forum?id=Y8OaqdX5Xt`. under review. 2

[2] Chris Baker, Rebecca Saxe, and Joshua Tenenbaum. Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the annual meeting of the cognitive science society*, volume 33, 2011. 2

[3] Jennifer Sootsman Buresh and Amanda L. Woodward. Infants track action goals within and across agents. *Cognition*, 104(2):287–314, August 2007. doi: 10.1016/j.cognition.2006.07.001. URL `https://doi.org/10.1016/j.cognition.2006.07.001`. 1

[4] Gergely Csibra and György Gergely. 'obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta psychologica*, 124(1):60–78, 2007. 2

[5] Carlos Florensa, David Held, Xinyang Geng, and Pieter Abbeel. Automatic goal generation for reinforcement learning agents. In *International conference on machine learning*, pages 1515–1528. PMLR, 2018. 3

[6] György Gergely, Zoltán Nádasdy, Gergely Csibra, and Szilvia Bíró. Taking the intentional stance at 12 months of age. *Cognition*, 56(2):165–193, 1995. 1

[7] Fritz Heider and Marianne Simmel. An experimental study of apparent behavior. *The American journal of psychology*, 57(2):243–259, 1944. 3

[8] Ronald A Howard. Dynamic programming and markov processes. 1960. 1

[9] Max Kleiman-Weiner, Mark K Ho, Joseph L Austerweil, Michael L Littman, and Joshua B Tenenbaum. Coordinate to cooperate or compete: abstract goals and joint intentions in social interaction. In *CogSci*, 2016. 1

[10] Michael L Littman. Markov games as a framework for multi-agent reinforcement learning. In *Machine learning proceedings 1994*, pages 157–163. Elsevier, 1994. 1

[11] Minghuan Liu, Menghui Zhu, and Weinan Zhang. Goal-conditioned reinforcement learning: Problems and solutions. *arXiv preprint arXiv:2201.08299*, 2022. 3

[12] Amy McGovern and Andrew G Barto. Automatic discovery of subgoals in reinforcement learning using diverse density. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pages 361–368, 2001. 3

[13] Miquel Ramırez and Hector Geffner. Goal recognition over pomdps: Inferring the intention of a pomdp agent. In *IJCAI*, pages 2009–2014. IJCAI/AAAI, 2011. 1

[14] Sarah A Wu, Rose E Wang, James A Evans, Joshua B Tenenbaum, David C Parkes, and Max Kleiman-Weiner. Too many cooks: Bayesian inference for coordinating multi-agent collaboration. *Topics in Cognitive Science*, 13(2):414–432, 2021. 1

[15] Fuxiang Zhang, Chengxing Jia, Yi-Chen Li, Lei Yuan, Yang Yu, and Zongzhang Zhang. Discovering generalizable multi-agent coordination skills from multi-task offline data. In *The Eleventh International Conference on Learning Representations*, 2022. 3