

Strategic Concealment of Environment Representations in Competitive Games

Yue Guan^{ID}, *Member, IEEE*, Dipankar Maity^{ID}, *Senior Member, IEEE*,
and Panagiotis Tsiotras^{ID}, *Fellow, IEEE*

Abstract—This letter studies the strategic concealment of environment representations used by players in competitive games. We consider a defense scenario in which one player (the Defender) seeks to infer and exploit the representation used by the other player (the Attacker). The interaction between the two players is modeled as a Bayesian game: the Defender infers the Attacker's representation from its trajectory and places barriers to counteract, while the Attacker obfuscates its representation type to mislead the Defender. We solve for the Perfect Bayesian Equilibrium via a bilinear program that integrates Bayesian inference, strategic planning, and belief manipulation. Simulations show that purposeful concealment naturally emerges: the Attacker randomizes its trajectory to manipulate the Defender's belief, inducing suboptimal barrier selections and thereby gaining a strategic advantage.

Index Terms—Game theory, agent-based systems.

I. INTRODUCTION

MODERN autonomous systems must operate in complex and high-dimensional environments, where maintaining a fully detailed representation of the world is infeasible. Sensing limitations such as finite resolution and range naturally lead to coarse or partial representations [1], [2], while computational limits require agents to employ environment representations that deliberately reduce resolution and remove irrelevant information for tractable decision-making [3]. Although such reduced representations are indispensable, they inevitably induce suboptimality in the resulting strategies.

In single-agent settings, such suboptimality is often acceptable as a trade-off for real-time solutions, and prior work has studied the design of such reduced representations [3], [4]. In multi-agent environments, however, employing reduced representations is more delicate, since an agent's reduced representation, if exposed, may lead to structural weaknesses exploitable by adversaries [5], [6]. This motivates a novel and underexplored problem: how should agents behave

strategically to conceal their representations, and how can they infer and exploit their opponents' representations?

Motivated by this problem, we investigate strategic representation concealment within a Bayesian game framework. We formulate a defense game, where an Attacker reasons under a private representation only known to itself and tries to reach a target. The Defender, without the knowledge of the Attacker's representation, needs to select a barrier to obstruct the Attacker. Consequently, the Attacker needs to conceal its representation and intentionally manipulate the Defender's belief to reduce the risk of exploitation. To properly counteract, the Defender needs to infer the Attacker's representation from the Attacker's trajectory.

Our Bayesian formulation offers two key advantages over existing intent inference and concealment frameworks [7], [8]. First, it captures the notion of *purposeful concealment*—the Attacker conceals its representation only when doing so improves its performance, rather than merely maximizing the uncertainty or the bias in the Defender's belief system. Second, the Bayesian approach avoids prescribing a pre-defined Defender inference algorithm, which is often impractical in adversarial settings [9], [10].

Previous work has developed efficient algorithms for solving Bayesian games using linear programming [11], information decomposition [12], and sequential rationalization [13]. In those settings, an agent's type typically corresponds to latent factors such as its goal or unobserved state. In contrast, our setting links the Attacker's type directly to its environment representation, so that the Attacker's type not only shapes rewards but also restricts the set of admissible strategies. Specifically, we impose that all states within a 'superstate' in a given representation share the same action distribution, which is motivated by principles from minimum-attention control [14] and hierarchical learning [4], where decisions are made at the coarse-resolution level instead of at each state.

a) Contributions: The main contributions are twofold: (i) A Bayesian game formulation that models an agent's environment representation as private information, and (ii) a non-iterative bilinear-programming method for computing the Perfect Bayesian Equilibrium (PBE). We show that purposeful representation concealment naturally arises at equilibrium. Specifically, the concealing player strategically manipulates its opponent's belief by shaping a distribution over game trajectories, thereby rendering its environment representation ambiguous to the opponent. A numerical example validates the equilibrium solution and demonstrates that such strategic concealment yields notable performance gains over path-optimal strategies that directly reveal the representation used.

Received 11 September 2025; revised 4 November 2025; accepted 15 November 2025. Date of publication 25 November 2025; date of current version 3 December 2025. This work was supported by the ARL Grant ARL DCIST CRA under Grant W911NF-17-2-0181. Recommended by Senior Editor C. Briat. (*Corresponding author: Yue Guan.*)

Yue Guan and Panagiotis Tsiotras are with the School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30345 USA (e-mail: yguan44@gatech.edu; tsiotras@gatech.edu).

Dipankar Maity is with the Department of Electrical and Computer Engineering, University of North Carolina at Charlotte, Charlotte, NC 28223 USA (e-mail: dmaity@charlotte.edu).

Digital Object Identifier 10.1109/LCSYS.2025.3637014

2475-1456 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

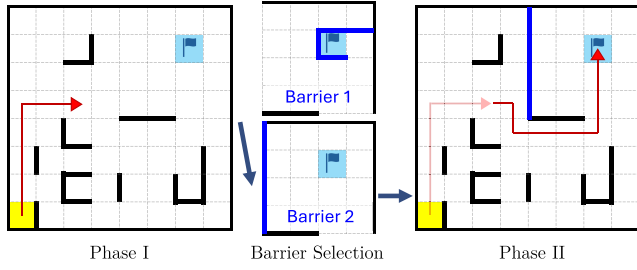


Fig. 1. An example of the proposed two-phase defense game.

b) Notation: We write $\mathcal{P}(E)$ for the set of probability distributions over a finite set E , and $[T] = \{0, \dots, T\}$. We use $\mathbf{v} = [v_i]_{i \in \mathcal{I}}$ for vectors whose i -th element is v_i , and $\mathbf{a} = (a_{i,j})_{i \in \mathcal{I}, j \in \mathcal{J}}$ for the collection of elements $a_{i,j}$.

II. PROBLEM FORMULATION

We consider a two-player defense game in a grid environment, as shown in Fig. 1. At the start, the Attacker is privately assigned an environment representation, which is used by the Attacker to reach its goal while *minimizing traverse time*. The Defender infers the representation from the Attacker's behavior and selects a barrier at a *pre-defined* time T to obstruct the Attacker's progress. We model this strategic interaction as a two-phase Bayesian game defined as follows.

a) State and action spaces: At time t , the game state is the Attacker's position $s_t \in \mathcal{S}$, where $\mathcal{S}$ is the set of grid cells, and the Attacker is tasked with reaching a goal within the goal set $G \subseteq \mathcal{S}$. At each step, the Attacker selects an action $a_t \in \mathcal{A} = \{\text{Up}, \text{Down}, \text{Left}, \text{Right}\}$ to move to s_{t+1} . The Attacker stays at s_t if its move is blocked by a wall or barrier. The Defender selects a barrier configuration ω at time T from a finite set of allowed configurations Ω . To avoid trivial outcomes, no configuration may fully enclose the goal set.

b) Game phases: The game unfolds in two sequential phases, separated by a *pre-defined* time step T .

- *Phase I* ($t = 0, \dots, T$): The Attacker moves in the grid environment, while the Defender observes the Attacker's state-action trajectory.
- *Barrier selection* ($t = T$): Given the observed Attacker trajectory in Phase I, the Defender deploys a barrier, thereby altering the traversability of the grid environment.
- *Phase II* ($t = T + 1, \dots, T_f$): With the barrier introduced to the environment, the Attacker proceeds until reaching a goal state $g \in G$, at which point the game terminates.

The Phase I history is given by $h_t = (s_0, a_0, \dots, s_{t-1}, a_{t-1}, s_t)$. In Phase II, we have $h_t = (s_0, a_0, \dots, s_T, \omega, s_{T+1}, a_{T+1}, \dots, s_t)$, where $s_T = s_{T+1}$ since only the Defender acts at T . The set of admissible histories at time t is denoted as $\mathcal{H}_t$. Fig. 1 presents an example of the game.

c) Environment representations: An environment representation Γ is a partition of the state space $\mathcal{S}$ into superstates: $\Gamma = \{\gamma^1, \dots, \gamma^{|\Gamma|}\}$, with $\gamma^i \subseteq \mathcal{S}$ and $\gamma^i \cap \gamma^j = \emptyset$ for $i \neq j$. While the framework is general, we focus on quad-tree-based representations [3] for their natural compatibility with grid worlds (see Fig. 2 for examples).

At the beginning of the game, the Attacker is privately assigned a type $\theta \in \Theta$, drawn according to the prior distribution $\mathbf{b}_0 \in \mathcal{P}(\Theta)$. Each type θ specifies a representation Γ^θ in the collection of representations $(\Gamma^1, \dots, \Gamma^{|\Theta|})$. A type- θ Attacker is constrained to use the same action distribution

for all states within each superstate $\gamma \in \Gamma^\theta$. This **same action distribution (SAD) constraint** reduces the flexibility of Attacker strategies and creates opportunities for the Defender to infer and exploit the representation.

Fig. 2 illustrates such exploitation: The type-1 representation in (a)–(c) features coarse resolution near the goal and fine resolution near the start, whereas type-2 in (d)–(e) exhibits the opposite pattern. For type-1 representations, barrier 1 yields poor Attacker performance and is therefore favorable to the Defender: the coarse region near the goal forces a mixed Attacker strategy, and a suboptimal realization is shown in (b). Barrier 2, however, enables a deterministic straight-line path to the goal. For type 2, the roles reverse: barrier 2 requires randomization, while barrier 1 permits straight-line entry.

d) Information structure: The prior $\mathbf{b}_0$, goal set G , representations $(\Gamma^\theta)_{\theta=1}^{|\Theta|}$, and barrier configurations Ω are common knowledge. Both players observe the history h_t , but only the Attacker knows its assigned representation Γ^θ ; the Attacker observes the barrier choice $\omega \in \Omega$ at the start of Phase II.

e) Strategy sets: The Attacker's strategy is a mapping $\pi_t^\theta : \mathcal{H}_t \rightarrow \mathcal{P}(\mathcal{A})$ subject to the **SAD constraint**: for any two histories ending in the same superstate $\gamma \in \Gamma^\theta$ (in the same phase), the action distributions must be identical, i.e.,

$$\pi_t^\theta(h_t) = \pi_t^\theta(h'_t) \quad \text{if } s_t, s'_t \in \gamma,$$

where s_t and s_t' are the last states of histories h_t and h'_t . Since this constraint applies separately in each phase for different history lengths, the Attacker's strategy can be represented as stationary strategies: $\pi_\emptyset^\theta, \pi_\omega^\theta : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{A})$, where $\pi_\emptyset^\theta$ denotes the type- θ Attacker's strategy before the Defender's barrier selection, and π_ω^θ denotes the strategy after barrier ω . The SAD-constraint requires $\pi_\emptyset^\theta$ and π_ω^θ to satisfy

$$\pi_\bullet^\theta(a|s) = \pi_\bullet^\theta(a|s'), \quad \forall s, s' \in \gamma, \forall a \in \mathcal{A}.$$

We denote the type- θ full strategy across both phases as $\pi^\theta \in \Pi^\theta$, and the complete strategy profile across types as $\pi \in \Pi$.

The Defender acts only once at T , selecting barrier $\omega \in \Omega$ via $\sigma : \mathcal{H}_T \rightarrow \mathcal{P}(\Omega)$ without knowing θ , where $\sigma(\omega|h_T)$ is the probability of choosing barrier ω given the Phase I history h_T . The Defender's strategy set is denoted by Σ .

Remark 1: While SAD constraints yield stationary Attacker strategies within each phase, the Defender's strategy remains history-dependent since the Defender relies on the observed trajectory to infer and exploit the Attacker's representation.

f) Optimization and equilibrium: We consider a running reward $r(s, a)$, and the expected cumulative reward for a type- θ Attacker starting at state s_0 is given by

$$J^\theta(s_0; \pi^\theta, \sigma) = \mathbb{E}_{\pi^\theta, \sigma} \left[\sum_{t=0}^{T_f} \beta^t r(s_t, a_t) \mid s_0 \right], \quad (1a)$$

where $\beta \in (0, 1)$ is the discount factor.

Since the Defender does not observe θ , it evaluates its expected performance with respect to the prior:

$$J(s_0; \pi, \sigma, \mathbf{b}_0) = \sum_{\theta \in \Theta} \mathbf{b}_0(\theta) J^\theta(s_0; \pi^\theta, \sigma). \quad (1b)$$

Before a representation is assigned by Nature, the Attacker maximizes its objective J^θ for each θ (i.e., minimizes traverse time to the goal), while the Defender minimizes the expected objective J , leading to the following equilibrium formulation.

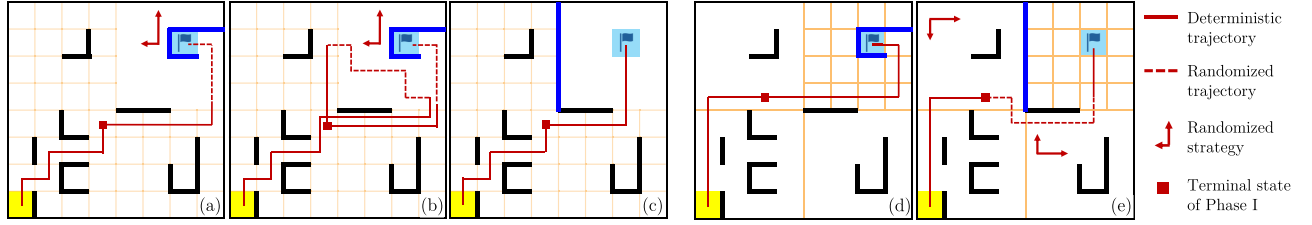


Fig. 2. Examples of two representations with different barrier placements. Arrows indicate mixed strategies in coarse regions. Solid-line trajectory segments are achieved using deterministic strategies; dashed-line segments require randomization. Subplots (a) and (b) show different realizations.

Definition 1: An Attacker strategy $\pi^* = (\pi^{\theta*})_{\theta}$ and a Defender strategy σ^* constitutes an equilibrium if, for all $\pi \in \Pi$,

$$J^{\theta}(s_0; \pi^{\theta*}, \sigma^*(\pi^*)) \geq J^{\theta}(s_0; \pi^{\theta}, \sigma^*(\pi)), \quad \forall \theta \in \Theta, \quad (2a)$$

$$\sigma^*(\pi^*) \in \operatorname{argmin}_{\sigma \in \Sigma} J(s_0; \pi^*, \sigma). \quad (2b)$$

III. PHASE II: SAD-CONSTRAINED MDP

Since the game unfolds in two phases, we adopt a backward induction approach: we first solve Phase II, which reduces to a single-agent Markov Decision Process (MDP) given the barrier placed by the Defender, with the additional SAD constraints; the Phase II solution, which represents the game's future outcome, is then used as terminal conditions in Phase I. We will show that both phases can be formulated as bilinear programs, with the bilinearity arising from the SAD constraint.

For each Attacker type $\theta \in \Theta$, barrier $\omega \in \Omega$, and Phase II initial state $s_T \in \mathcal{S}$ (Attacker's position at the end of Phase I), the Phase II problem is defined as a constrained MDP:

$$\text{cMDP}_{\omega, s_T}^{\theta} = \langle s_T, \mathcal{S}, \mathcal{A}, r, f^{\omega}, \beta \rangle,$$

where $\mathcal{S}$, $\mathcal{A}$, r and β denote the original state space, action space, Attacker reward function, and discount factor, respectively. The transition function f^{ω} encodes the grid world traversability with the added barrier ω . Specifically, the transition function is a mapping $f^{\omega} : \mathcal{S} \times \mathcal{S} \times \mathcal{A} \rightarrow [0, 1]$.

One can solve the SAD-constrained MDP using the state-action occupation measure [15], which leads to the following bilinear program, with the policy π^{θ} and the discounted cumulative state-occupation d as decision variables.

$$\begin{aligned} \max_{d, \pi^{\theta}} \quad & \sum_{s \in \mathcal{S}} d(s) \sum_{a \in \mathcal{A}} \pi^{\theta}(a|s) r(s, a) \\ \text{s.t.} \quad & d(s) = \mu(s) + \beta \sum_{s' \in \mathcal{S}} \sum_{a' \in \mathcal{A}} f^{\omega}(s|s', a') \pi^{\theta}(a'|s') d(s'), \\ & \sum_a \pi^{\theta}(a|s) = 1, \pi^{\theta}(a|s) \geq 0, d(s) \geq 0, \quad \forall s \in \mathcal{S}, \\ & \pi^{\theta}(a|s) = \pi^{\theta}(a|s'), \quad \forall a \in \mathcal{A}, \quad \forall s, s' \in \gamma, \forall \gamma \in \Gamma^{\theta}, \end{aligned} \quad (3)$$

where μ is the initial state distribution. The first constraint is the flow constraint and the last one is the SAD constraint. Unlike unconstrained MDPs, the solution to the SAD-constrained MDP depends on the initial distribution due to the policy coupling within each superstate. To obtain values for every Phase I terminal state s_T (initial state for Phase II), we solve (3) with $\mu = \mathbf{e}_{s_T}$, where $\mathbf{e}_s$ is the one-hot vector. We repeat this for every representation Γ^{θ} and every barrier configuration ω . The optimal values $V_{\omega, s_T}^{\theta*}$ serve as the terminal payoffs for Phase I, whereas the optimal policies $\pi_{\omega, s_T}^{\theta*}$ govern the Attacker's behavior throughout Phase II.

IV. PHASE I: SAD-CONSTRAINED BAYESIAN GAME

The key distinction between Phase I and Phase II is that the Defender selects a barrier configuration $\omega \in \Omega$ at the Phase I terminal time T , thereby influencing the outcome of Phase I. Accordingly, Phase I is a *game* between the two players, rather than a single-agent MDP. Moreover, since the Defender does not observe the representation assigned to the Attacker, Phase I is an asymmetric-information game [9].

In this setting, the Attacker faces a trade-off between path optimality and concealing its environment representation. The Attacker may forgo fine-resolution maneuvers around obstacles and instead follow a suboptimal path that mimics behavior under a coarser representation, misleading the Defender into choosing a less effective barrier.

We formalize this interaction as a *Bayesian game*, where the Defender maintains a belief over the Attacker's representation types and selects a barrier accordingly. In turn, the Attacker optimizes its strategy against the Defender's belief, balancing path optimality with representation concealment. We formalize the SAD-constrained Bayesian game as

$$\text{BG} = \langle \mathcal{S}, \mathcal{A}, f, r, (V_{\omega, s}^{\theta*})_{s \in \mathcal{S}, \omega \in \Omega}, \Theta, (\Gamma^{\theta})_{\theta \in \Theta}, \beta, T \rangle,$$

where f is the transition function before the Defender places its barrier, and the terminal values $(V_{\omega, s}^{\theta*})_{s \in \mathcal{S}, \omega \in \Omega}$ come from the Phase II solution. The set Θ consists of all possible Attacker types, and $(\Gamma^{\theta})_{\theta}$ is the set of representations. Finally, T is the pre-defined terminal time for Phase I.

A. Value Functions

For convenience, let $\pi_{h_t}^{\theta} \in \mathcal{P}(\mathcal{A})$ denote the action distribution of a type- θ Attacker given history h_t , and $\pi_{h_t} = (\pi_{h_t}^{\theta})_{\theta}$ the joint strategy profile. Similarly, $\sigma_{h_T} \in \mathcal{P}(\Omega)$ is the Defender's action distribution given Phase I history h_T . We denote the belief that the Attacker is type θ by $b_t^{\theta} = \mathbf{b}_t(\theta)$.

The type- θ Attacker value at Phase I is evaluated by the following recursive formula¹:

$$J_T^{\theta}(h_T, \sigma_{h_T}) = \sum_{\omega \in \Omega} V_{\omega, s_T}^{\theta*} \sigma_{h_T}(\omega), \quad (4a)$$

$$\begin{aligned} J_t^{\theta}(h_t, \pi_{h_t}^{\theta}) = & \sum_{a_t \in \mathcal{A}} \pi_{h_t}^{\theta}(a_t) \left(r(s_t, a_t) + \right. \\ & \left. \beta \sum_{s_{t+1}} f(s_{t+1}|s_t, a_t) J_{t+1}^{\theta}(h_{t+1}, \pi_{h_{t+1}}^{\theta}) \right), \end{aligned} \quad (4b)$$

where $h_{t+1} = (h_t, a_t, s_{t+1})$ and $V_{\omega, s}^{\theta*}$ is the optimal value from the solution of the Phase II SAD-constrained MDP.

¹Since J_T^{θ} depends on σ_{h_T} , the recursion at $t = T - 1$ differs slightly from earlier steps. For brevity we omit this distinction in the sequel.

Next, we derive the recursive formula for the Defender value. For the terminal time step, we have by definition,

$$J_T(h_T, \mathbf{b}_T, \sigma_{h_T}) = \sum_{\theta \in \Theta} b_T^\theta J_T^\theta(h_T, \sigma_{h_T}) = \sum_{\theta \in \Theta} b_T^\theta V_{\omega, s_T}^{\theta*} \sigma_{h_T}(\omega).$$

For time step $t \in [T-1]$, define the reward vector $\mathbf{r}(s) = [r(s, a)]_{a \in \mathcal{A}}$, then we have that

$$\begin{aligned} J_t(h_t, \mathbf{b}_t, \pi_{h_t}) &= \sum_{\theta \in \Theta} b_t^\theta J_t^\theta(h_t, \sigma_{h_t}) = \sum_{\theta \in \Theta} b_t^\theta \left(\mathbf{r}(s_t)^\top \pi_{h_t}^\theta \right. \\ &\quad \left. + \beta \sum_{a_t, s_{t+1}} \pi_{h_t}^\theta(a_t) f(s_{t+1}|s_t, a_t) J_{t+1}^\theta(h_{t+1}, \pi_{h_{t+1}}^\theta) \right) \\ &= \left(\sum_{\theta} b_t^\theta \mathbf{r}(s_t)^\top \pi_{h_t}^\theta \right) + \beta \sum_{s_{t+1}} \sum_{a_t} f(s_{t+1}|s_t, a_t) \\ &\quad \times \underbrace{\bar{\chi}(\mathbf{b}_t, \pi_{h_t}, a_t)}_{=\sum_{\theta} b_t^\theta \pi_{h_t}^\theta(a_t)} \left(\sum_{\theta} \underbrace{\frac{b_t^\theta \pi_{h_t}^\theta(a_t)}{\bar{\chi}(\mathbf{b}_t, \pi_{h_t}, a_t)}}_{b_{t+1}^\theta} J_{t+1}^\theta(h_{t+1}, \pi_{h_{t+1}}^\theta) \right) \\ &= \left(\sum_{\theta} b_t^\theta \mathbf{r}(s_t)^\top \pi_{h_t}^\theta \right) + \beta \sum_{s_{t+1}} \sum_{a_t} f(s_{t+1}|s_t, a_t) \\ &\quad \times \bar{\chi}(\mathbf{b}_t, \pi_{h_t}, a_t) J_{t+1}(h_{t+1}, \mathbf{b}_{t+1}, \pi_{h_{t+1}}), \end{aligned} \quad (5)$$

where we use the definition $J_{t+1} = \sum_{\theta} b_{t+1}^\theta J_{t+1}^\theta$, and arrive at the recursion for the Defender's value J_t in (5). In particular, $h_{t+1} = (h_t, a_t, s_{t+1})$ is the new history induced after applying action a_t and transitioning to s_{t+1} , and $\bar{\chi}(\mathbf{b}_t, \pi_{h_t}, a_t) = \sum_{\theta} b_t^\theta \pi_{h_t}^\theta(a_t)$ is the total probability that the Attacker selects action a_t given the current belief $\mathbf{b}_t$ and its strategy π_{h_t} . The next belief $\mathbf{b}_{t+1}$ after observing action a_t is updated as

$$\mathbf{b}_{t+1}(\theta) = b_{t+1}^\theta = \frac{b_t^\theta \pi_{h_t}^\theta(a_t)}{\bar{\chi}(\mathbf{b}_t, \pi_{h_t}, a_t)}. \quad (6)$$

The above is the Bayesian belief update-rule for the Defender, since the numerator $b_t^\theta \pi_{h_t}^\theta(a_t)$ admits the probability that action a_t is selected by a type- θ Attacker.

B. Perfect Bayesian Equilibrium (PBE)

With asymmetric information, the Defender must update its belief about the Attacker's type from observed trajectories, while the Attacker, though knowing its own type, must account for the Defender's evolving belief when planning its actions. This interplay is captured by the Perfect Bayesian Equilibrium (PBE), which requires agents to act rationally given the game state and their beliefs, while ensuring consistency between strategies and Bayesian belief updates. We consider a sequential decision structure, where the Attacker's strategy first induces beliefs via (6); the Defender subsequently optimizes its barrier placement strategy under the resulting beliefs.

Definition 2: A strategy pair (π^*, σ^*) forms a (weak) Perfect Bayesian Equilibrium if the following holds for all nonzero-probability histories $h_t \in \mathcal{H}_t$ under (π^*, σ^*) .

- $J_t(h_t, \mathbf{b}_t, \pi_{h_t}^*) \geq J_t(h_t, \mathbf{b}_t, \pi_{h_t})$ for all $\pi_{h_t} \in (\mathcal{P}(\mathcal{A}))^{|\Theta|}$ and $t \in [T-1]$;
- $\sigma_{h_T}^* \in \arg\min_{\sigma_{h_T} \in \mathcal{P}(\Omega)} J_T(h_T, \mathbf{b}_T, \sigma_{h_T})$;
- The beliefs $\mathbf{b}_t$ are propagated with Bayes' rule along the history h_t conditioned on the Attacker strategy π^* .

The above definition is a "weak" PBE, because it only considers histories with nonzero probability under the equilibrium strategies. For treatments of zero-probability (off-equilibrium) histories using perturbation limits, see [9].

C. Optimal Value Functions

Definition 2 yields a min-max dynamic program for the optimal value function:

$$J_T^*(h_T, \mathbf{b}_T) = \min_{\sigma_{h_T}} \sum_{\theta} b_T^\theta V_{s_T}^{\theta*} \sigma_{h_T}, \quad (7a)$$

$$J_t^*(h_t, \mathbf{b}_t) = \max_{\pi_{h_t}} \sum_{\theta} b_t^\theta \mathbf{r}(s_t)^\top \pi_{h_t}^\theta \quad (7b)$$

$$+ \beta \sum_{s_{t+1}, a_t} f(s_{t+1}|s_t, a_t) \bar{\chi}(\mathbf{b}_t, \pi_{h_t}, a_t) J_{t+1}^*(h_{t+1}, \mathbf{b}_{t+1}),$$

where $V_{s_T}^{\theta*} = [V_{\omega, s_T}^{\theta*}]_{\omega \in \Omega}$ is the vector of the terminal values.

We next state a scaling property of the optimal value function [11], which is essential to deriving the bilinear program solution of the min-max problem.

Lemma 1: For any constant $\alpha > 0$ and time step $t \in [T]$, the game value $J_t^*(h_t, \mathbf{b}_t)$ satisfies

$$J_t^*(h_t, \alpha \mathbf{b}_t) = \alpha J_t^*(h_t, \mathbf{b}_t).$$

Proof: The result follows directly from the definition $J_t(h_t, \mathbf{b}_t, \pi_{h_t}) = \sum_{\theta \in \Theta} \mathbf{b}_t(\theta) J_t^\theta(h_t, \pi_{h_t}^\theta)$ for $t \in [T-1]$, and $J_T(h_T, \mathbf{b}_T, \sigma_{h_T}) = \sum_{\theta \in \Theta} b_T^\theta V_{s_T}^{\theta*} \sigma_{h_T}$. ■

D. Bilinear-Program Approach

We reformulate the min-max problem in (7) as a bilinear program. A key challenge is that the belief $\mathbf{b}_t$ lies in a continuous space, and thus a direct DP approach would require discretization, leading to errors and high computational complexity. Our bilinear program leverages the fact that the set of histories is finite and constructs beliefs along each history. We first present the optimization problem for the Attacker's optimal strategy without the SAD constraint, focusing on belief propagation and the T -step optimization; the SAD constraints are incorporated afterward.

a) Terminal time T : At the Phase I terminal time T , the Defender's minimization problem can be transformed to an equivalent maximization through duality [9]:

$$\begin{aligned} 3J_T^*(h_T, \mathbf{b}_T) &= \max_{z_{h_T}, \ell_{h_T}} \ell_{h_T} \\ \text{s.t.} \quad &\sum_{\theta \in \Theta} z_{h_T}^\theta V_{s_T}^{\theta*} \geq \ell_{h_T} \mathbf{1}, \\ &z_{h_T}^\theta = b_T^\theta, \quad \forall \theta \in \Theta, \end{aligned} \quad (8)$$

where for consistency with later derivations, we make the change of variable $z_{h_T}^\theta = b_T^\theta$ and denote $z_{h_T} = (z_{h_T}^\theta)_{\theta \in \Theta}$.

b) Time step $T-1$: Now consider the maximization problem performed by the Attacker at $T-1$. From Lemma 1, we can distribute the scaling constants β , $f(s_T|s_{T-1}, a_{T-1})$ and $\chi(\mathbf{b}_T, \pi_{h_{T-1}}, a_{T-1})$ into the value function J_T^* :

$$\begin{aligned} J_{T-1}^*(h_{T-1}, \mathbf{b}_{T-1}) &= \max_{\pi_{h_{T-1}}} \left\{ \sum_{\theta \in \Theta} b_{T-1}^\theta \mathbf{r}(s_{T-1})^\top \pi_{h_{T-1}}^\theta \right. \\ &\quad \left. + \sum_{s_T, a_{T-1}} J_T^*\left(h_T, \underbrace{[\beta f_{s_T, s_{T-1}}^{a_{T-1}} b_{T-1}^\theta \pi_{h_{T-1}}^\theta(a_{T-1})]_{\theta \in \Theta}}_{\text{scaled belief}}\right) \right\}, \end{aligned} \quad (9)$$

where we use $f_{s_T, s_{T-1}}^{a_{T-1}}$ to denote the transition probability.

We first analyze the optimal terminal value J_T^* in (9) under the (scaled) belief induced by a fixed

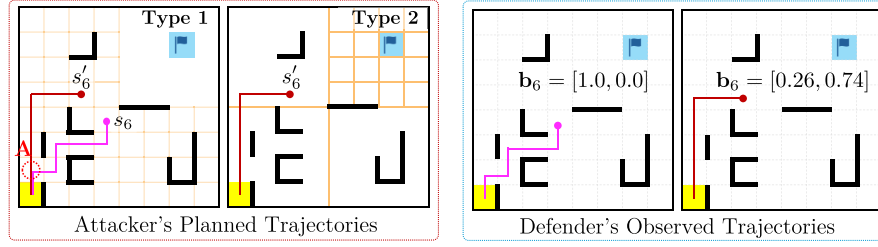


Fig. 3. **Left:** Attacker trajectories under each representation type. **Right:** Defender's observed trajectories and corresponding beliefs.

choice of $\pi_{h_{T-1}}$. Plugging in the induced (scaled) belief into (8):

$$\begin{aligned} 3J_T^* = \max_{z_{h_T}, \ell_{h_T}} \quad & \ell_{h_T} \\ \text{s.t.} \quad & \sum_{\theta \in \Theta} z_{h_T}^\theta \mathbf{V}_{s_T}^\theta \geq \ell_{h_T} \mathbf{1}, \\ & z_{h_T}^\theta = \beta f_{s_T, s_{T-1}}^{a_{T-1}} b_{T-1}^\theta \pi_{h_{T-1}}^\theta(a_{T-1}), \quad \forall \theta \in \Theta, \\ & \mathbf{1}^\top \pi_{h_{T-1}}^\theta = 1, \pi_{h_{T-1}}^\theta \geq 0, \forall \theta \in \Theta. \end{aligned} \quad (10)$$

We perform the change of variables via $z_{h_{T-1}}^\theta = b_{T-1}^\theta \pi_{h_{T-1}}^\theta$, and plug the maximization for J_T^* in (10) back into (9). Maximizing with respect to $z_{h_{T-1}}$, equivalently $\pi_{h_{T-1}}$, leads to the following maximization problem for J_{T-1}^* .

$$\begin{aligned} \max_{z_{h_T}, z_{h_{T-1}}, \ell_{h_T}, \ell_{h_{T-1}}} \quad & \ell_{h_{T-1}} + \sum_{h_T \in \mathcal{H}_T} \ell_{h_T} \\ \text{s.t.} \quad & \ell_{h_T} \mathbf{1} \leq \sum_{\theta \in \Theta} \mathbf{V}_{s_T}^\theta z_{h_T}^\theta, \quad \forall h_T \in \mathcal{H}_T, \end{aligned} \quad (11a)$$

$$z_{h_T}^\theta = \beta f_{s_T, s_{T-1}}^{a_{T-1}} z_{h_{T-1}}^\theta(a_{T-1}), \quad \forall \theta \in \Theta, h_T \in \mathcal{H}_T, \quad (11b)$$

$$\ell_{h_{T-1}} = \sum_{\theta \in \Theta} \mathbf{r}(s_{T-1})^\top z_{h_{T-1}}^\theta, \quad (11c)$$

$$\mathbf{1}^\top z_{h_{T-1}}^\theta = b_{T-1}^\theta, \quad z_{h_{T-1}}^\theta \geq 0, \quad \forall \theta \in \Theta, \quad (11d)$$

where (11a) corresponds to the optimization performed by the Defender at time T ; (11b) captures the influence of the policy $z_{h_{T-1}}$ on the belief at time T , and thus affects the optimization at the next stage of the game; finally, (11d) encodes the starting belief at $T-1$.

Note that the ℓ -variables encode the expected stage reward for each trajectory, while the z -variables are *history-action occupation measures* analogous to state-action measures in MDP. Let $z_{h_t}^*$ denote the optimal solution to (11), the optimal Attacker strategy is then given by

$$\pi_{h_t}^*(u) = \begin{cases} \frac{z_{h_t}^*(u)}{\sum_{\bar{u}} z_{h_t}^*(\bar{u})}, & \text{if } \sum_{\bar{u}} z_{h_t}^*(\bar{u}) > 0, \\ \frac{1}{|\mathcal{A}(h_t)|}, & \text{otherwise.} \end{cases} \quad (12)$$

Note that if $\sum_{\bar{u}} z_{h_t}^*(\bar{u}) = 0$, the trajectory h_t has zero probability under the equilibrium strategies. For practical purposes, the action distribution on these off-equilibrium trajectories is inconsequential and we set them to be uniform.

c) *Generic problem:* For the entire T -stage game, one can follow the pattern in (11) and add the SAD-constraints using the relationship between the policy and the z variables in (12). The optimization for the entire Phase I game is given by

$$\begin{aligned} \max_{z_{h_t}, \ell_{h_t}} \quad & \sum_{t=0}^T \sum_{h_t \in \mathcal{H}_t} \ell_{h_t} \\ \text{s.t.} \quad & \ell_{h_0} = \sum_{\theta \in \Theta} \mathbf{r}(s_0)^\top z_{h_0}^\theta, \end{aligned} \quad (13)$$

$$\mathbf{1}^\top z_{h_0}^\theta = b_0^\theta, \quad z_{h_0}^\theta \geq 0, \quad \forall \theta \in \Theta,$$

$$\forall t = 1, \dots, T-1, \quad h_t \in \mathcal{H}_t$$

$$\ell_{h_t} = \sum_{\theta \in \Theta} \mathbf{r}(s_t)^\top z_{h_t}^\theta,$$

$$\mathbf{1}^\top z_{h_t}^\theta = \beta f_{s_t, s_{t-1}}^{a_{t-1}} z_{h_{t-1}}^\theta(a_{t-1}), \quad \forall \theta \in \Theta,$$

$$\ell_{h_T} \mathbf{1} \leq \sum_{\theta \in \Theta} \mathbf{V}_{s_T}^\theta z_{h_T}^\theta, \quad \forall h_T \in \mathcal{H}_T,$$

$$z_{h_T}^\theta = \beta f_{s_T, s_{T-1}}^{a_{T-1}} z_{h_{T-1}}^\theta(a_{T-1}), \quad \forall \theta \in \Theta, h_T \in \mathcal{H}_T, \\ \forall t, \tau \in [0, T-1]$$

$$\left(\sum_{a'} z_{h'_t}^\theta(a') \right) z_{h_t}^\theta = \left(\sum_a z_{h_t}^\theta(a) \right) z_{h'_t}^\theta, \quad \text{if } s_t, s'_t \in \gamma,$$

where the last constraint corresponds to the SAD constraint which is bilinear. The optimal Attacker strategy follows (12).

Remark 2: The z -variables are history-action occupation measures, analogous to the ones in MDPs. Thus, (13) shows that the Attacker optimizes over a distribution of game trajectories to effectively shape the Defender's belief at time T .

d) *Defender's strategy:* Given the optimal solution $z_{h_T}^*$ from (13), the Defender constructs its strategy as

$$\sigma_{h_T}^* \in \argmin_{\sigma \in \mathcal{P}(\Omega)} \sum_{\theta \in \Theta} z_{h_T}^{\theta*} \mathbf{V}_{s_T}^\theta \sigma. \quad (14)$$

V. NUMERICAL EXAMPLES

To validate the equilibrium from the bilinear program (13), we solve the example in Fig. 1 with the two representations in Fig. 2, evaluated under the performance metric in (1). We set $T = 6$, $\mathbf{b}_0 = [0.5, 0.5]$, and $r(s, a) = 1$ for $s \in G$ and -0.1 otherwise. Fig. 3 displays the Phase I Attacker trajectories.

A key observation is that the type-1 Attacker randomizes at Point A: choosing Up with probability 0.358 and Right with 0.642. By selecting Up, the type-1 Attacker mimics the behavior of a type-2 Attacker which has a coarse representation in the obstacle area, thereby introducing ambiguity into the Defender's belief. Specifically, if the Defender observes the red Attacker trajectory, its belief updates to $[0.264, 0.736]$ —i.e., a 26.4% probability that the Attacker is type-1 and 73.6% for type-2. If the Attacker is on the magenta trajectory, the Defender infers with certainty that the Attacker is type 1.

At the terminal state s'_6 of the red trajectory in Fig. 3, the SAD-constrained MDP values for each representation type and barrier are shown below. Under the “uncertain” Defender belief $\mathbf{b}_6 = [0.264, 0.736]$, the expected outcomes of both barrier configurations are -0.141 , leaving the Defender to randomize uniformly—an indifference that is a key characteristic of equilibrium behavior. The induced uncertainty in the Defender's belief along the red path benefits the Attacker. For the type-1 Attacker, the Defender may select the suboptimal

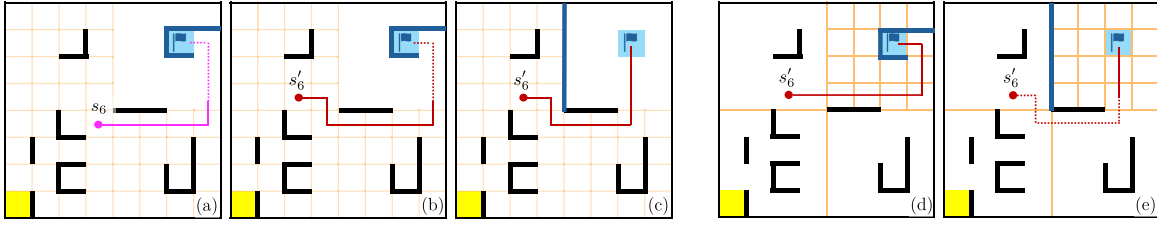


Fig. 4. Phase II trajectories: (a)–(c) type-1, (d)–(e) type-2. Dashed segments can only be achieved with randomized strategies, whose realized paths may deviate from those shown, see Fig. 2(a) and Fig. 2(b) for such realizations.

barrier 2, enabling a straight-line Attacker path to the target (Fig. 4c). Conversely, for the type-2 Attacker, the Defender may select the suboptimal barrier 1, allowing the Attacker to enter the fine-resolution area directly (Fig. 4d).

$V_{\omega, s'_6}^{\theta^*}$	$\omega = 1$	$\omega = 2$
$\theta = 1$	−0.800	0.093
$\theta = 2$	0.095	−0.225
expected outcome	−0.141	−0.141

In contrast, the *magenta* path in Fig. 3 reveals the Attacker as type 1, prompting the Defender to select barrier 1, which minimizes the Attacker’s future performance to −0.67. Meanwhile, if the type-1 Attacker follows the *red* path and barrier 1 is chosen, its performance drops to −0.8. This shows that, without the benefit of concealment, the magenta trajectory in Fig. 3 is path-optimal, whereas the red is not.

To highlight the benefit of representation concealment, consider the *full-information variant* of the game in which the Defender knows the Attacker’s representation type θ at the beginning. In this case, the Defender deterministically selects the optimal barrier for each type: barrier 1 for type-1 representation and barrier 2 for type-2. In turn, the Attacker *deterministically* responds with the corresponding path-optimal trajectory: magenta path for type 1 and red path for type 2 in Fig. 3. The Attacker’s expected performance, averaged over the prior, is given by

$$-0.978 = \underbrace{-0.530}_{\text{Phase I}} + \underbrace{0.5 \times -0.67}_{\theta=1, \text{ magenta path}, \omega=1} + \underbrace{0.5 \times -0.225}_{\theta=2, \text{ red path}, \omega=2}.$$

In contrast, in the asymmetric-information setting the Attacker employs randomized trajectories shown in Fig. 3 to introduce ambiguity in Defender’s belief as discussed. This raises the Attacker’s expected performance to −0.841. This corresponds to a 14% performance improvement when both phases are included. If the constant Phase I cumulative payoff is excluded, the performance improvement from concealment rises to 31%.

VI. CONCLUSION

This letter explored the asymmetric-information structure that arises when agents employ reduced environment representations in competitive games. We developed a Bayesian-game framework for the strategic concealment and inference of such representations in an attack–defense scenario, where the Attacker is incentivized to plan a trajectory to its goal while concealing its representation, and the Defender seeks to infer this hidden representation and place barriers to obstruct the Attacker’s progress. We integrated the Defender’s inference

and the Attacker’s planning into a unified bilinear optimization problem. Simulation results demonstrate that purposeful representation concealment naturally emerges as a means of improving the Attacker’s performance.

Future work includes generalizing the framework to two-sided private representations with mis-specified priors, as well as applying the insights of Remark 2 to broader asymmetric-information settings. Another interesting direction is to treat the barrier-placement time T as a Defender decision variable.

REFERENCES

- [1] S. LaValle, *Planning Algorithms*. Cambridge, U.K.: Cambridge Univ. Press, 2006.
- [2] S. Thrun, *Probabilistic Robotics*. Cambridge, MA, USA: MIT Press, 2005.
- [3] D. T. Larsson, D. Maity, and P. Tsotras, “Q-tree search: An information-theoretic approach toward hierarchical abstractions for agents with computational limitations,” *IEEE Trans. Robot.*, vol. 36, no. 6, pp. 1669–1685, Dec. 2020.
- [4] J. S. Baras and V. S. Borkar, “A learning algorithm for Markov decision processes with adaptive state aggregation,” in *Proc. 39th IEEE Conf. Decis. Control*, vol. 4, Dec. 2000, pp. 3351–3356.
- [5] K. Waugh, D. Schnizlein, M. H. Bowling, and D. Szafron, “Abstraction pathologies in extensive games,” in *Proc. 8th Int. Conf. Auton. Agents Multiagent Syst.*, Budapest, Hungary, May 2009, pp. 781–788.
- [6] Y. Guan, M. Afshari, Q. Zhang, and P. Tsotras, “Hierarchical decompositions of stochastic pursuit-evasion games,” in *Proc. 61st Conf. Decis. Control*, Cancún, Mexico, Dec. 2022, pp. 5062–5067.
- [7] M. Ramírez and H. Geffner, “Probabilistic plan recognition using off-the-shelf classical planners,” in *Proc. 24th AAAI Conf. Artif. Intell.*, vol. 24, Jul. 2010, pp. 1121–1126.
- [8] R. Pereira, N. Oren, and F. Meneguzzi, “Landmark-based heuristics for goal recognition,” in *Proc. 31st AAAI Conf. Artif. Intell.*, vol. 31, Feb. 2017, pp. 3622–3628.
- [9] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA, USA: MIT Press, 1991.
- [10] V. Rostobaya, Y. Guan, J. Berneburg, M. Dorothy, and D. Shishika, “Deception by motion: The eater and the mover game,” *IEEE Control Syst. Lett.*, vol. 7, pp. 3157–3162, 2023.
- [11] L. Li and J. Shamma, “LP formulation of asymmetric zero-sum stochastic games,” in *Proc. 53rd IEEE Conf. Decis. Control*, Dec. 2014, pp. 1930–1935.
- [12] Y. Ouyang, H. Tavafoghi, and D. Teneketzis, “Dynamic games with asymmetric information: Common information based perfect Bayesian equilibria and sequential decomposition,” *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 222–237, Jan. 2017.
- [13] V. Rostobaya, J. Berneburg, Y. Guan, M. Dorothy, and D. Shishika, “Deceptive path planning: A Bayesian game approach,” 2025, *arXiv:2506.13650*.
- [14] W. Brockett, “Minimum attention control,” in *Proc. 36th IEEE Conf. Decis. Control*, vol. 3, Dec. 1997, pp. 2628–2632.
- [15] D. P. Bertsekas and J. N. Tsitsiklis, *Neuro-Dynamic Programming*. Nashua, NH, USA: Athena Sci., 1996.