A Dynamic Game Framework for Rational and Persistent Robot Deception With an Application to Deceptive Pursuit-Evasion

Linan Huang, *Student Member, IEEE*, and Quanyan Zhu, *Member, IEEE*

***Abstract*— This article studies rational and persistent deception among intelligent robots to enhance security and operational efficiency. We present an N-player K-stage game with an asymmetric information structure where each robot's private information is modeled as a random variable or its type. The deception is persistent as each robot's private type remains unknown to other robots for all stages. The deception is rational as robots aim to achieve their deception goals at minimum cost. Each robot forms a dynamic belief of others' types based on intrinsic or extrinsic information. Perfect Bayesian Nash equilibrium (PBNE) is a natural solution concept for dynamic games of incomplete information. Due to its requirements of sequential rationality and belief consistency, PBNE provides a reliable prediction of players' actions, beliefs, and expected cumulative costs over the entire K stages. The contribution of this work is fourfold. First, we identify the PBNE computation as a nonlinear stochastic control problem and characterize the structures of players' actions and costs under PBNE. We further derive a set of extended Riccati equations with cognitive coupling under the linear-quadratic (LQ) setting and extrinsic belief dynamics. Second, we develop a receding-horizon algorithm with low temporal and spatial complexity to compute PBNE under intrinsic belief dynamics. Third, we investigate a deceptive pursuit-evasion game as a case study and use numerical experiments to corroborate the results. Finally, we propose metrics, such as deceivability, reachability, and the price of deception (PoD), to evaluate the strategy design and the system performance under deception.**

***Note to Practitioners*—Recent advances in automation and adaptive control in multi-agent systems enable robots to use deception to accomplish their objectives. Deception involves intentional information hiding to compromise the security and operational efficiency of the robotic systems. This work proposes a dynamic game framework to quantify the impact of deception, understand the robots' behaviors and intentions, and design cost-efficient strategies under the deception that persists over stages. Existing research studies on robot deception have relied on experiments while this work aims to lay a theoretical foundation of deception with quantitative metrics, such as deceivability and the PoD. The proposed model has wide applications, including cooperative robots, pursuit and evasion, and human–robot teaming. The pursuit-evasion games are used as case studies to show how the deceiver can amplify the deception by belief manipulation and how the deceived robots can reduce the negative impact of deception by enhanced maneuverability and Bayesian learning. The future work would focus on designing cooperative deception among swarm robotics and robotic systems that are robust to or further benefit from the deception.**

***Index Terms*— Discrete-time Riccati equations, linear-quadratic (LQ) games, perfect Bayesian equilibrium, pursuit-evasion, robot deception.**

NOMENCLATURE

Variable & Meaning

$\mathcal{N} := \{1, 2, \ldots, N\}$	Set of N players in the dynamic game.
$\mathcal{K} := \{0, 1, 2, \ldots, K\}$	Set of K discrete stages in the dynamic game.
$\Theta_i := \{\theta_i^1, \theta_i^2, \ldots, \theta_i^{N_i}\}$	Set of N_i possible types for player $i \in \mathcal{N}$.
$\theta_i \in \Theta_i$	Type of player $i \in \mathcal{N}$.
$\theta := [\theta_1, \ldots, \theta_N]$	N players' joint type.
$\Theta_{-i} := \prod_{j \in \mathcal{N} \setminus \{i\}} \Theta_j$	Set of types of all players except for player i.
$\theta_{-i} := [\theta_j]_{j \in \mathcal{N} \setminus \{i\}} \in \Theta_{-i}$	Types of all players except for player i.
$\Delta(\Theta_{-i})$	Set of probability distributions over set Θ_{-i}.
$\Xi_i(\cdot)$	Probability distribution of player i's type.
$\Xi = [\Xi_i]_{i \in \mathcal{N}}$	Probability distribution of the joint type θ.
$\Xi_w(\cdot)$	Probability distribution of noise $w^k, \forall k \in \mathcal{K}$.
$x^k \in \mathbb{R}^{n \times 1}$	System state of dimension n at stage k.
$x_i^k \in \mathbb{R}^{n_i \times 1}$	Player i's state of dimension n_i at stage k.
$[\hat{x}_i^k(\theta_i)]_{k \in \mathcal{K}}$	Reference trajectory for player i of type θ_i.
$\beta_i^k \in \Lambda_i \subseteq [0,1]^{\lvert\Theta_{-i}\rvert \times \lvert\Theta_i\rvert}$	Player i's belief state at stage k.
$\beta^k = [\beta_i^k]_{i \in \mathcal{N}} \in \Lambda$	N players' joint belief state at stage k.

Manuscript received January 13, 2021; revised April 5, 2021; accepted July 8, 2021. This article was recommended for publication by Associate Editor L. Ferrarini and Editor C. Seatzu upon evaluation of the reviewers' comments. This work was supported in part by the National Science of Foundation (NSF) under Award ECCS-1847056, Award CNS-1544782, Award CNS-2027884, and Award SES-1541164; and in part by the Army Research Office (ARO) under Grant W911NF-19-1-0041. *(Corresponding author: Linan Huang.)*

The authors are with the Department of Electrical and Computer Engineering, New York University, New York, NY 11201 USA (e-mail: lh2328@nyu.edu; qz494@nyu.edu).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TASE.2021.3097286.

Digital Object Identifier 10.1109/TASE.2021.3097286

$h^k := [x^0, \ldots, x^k] \in \mathscr{H}^k$	State history.
f^k	State transition function at stage k.
Γ_i^k	Player i's belief transition function at stage k.
g_i^k	Player i's cost function at stage k.
$V_i^k(\beta^k, x^k, \theta_i)$	Player i's PBNE cost.
$\bar{V}_i^k(x^k, \theta)$	Player i's PBNE cost when all players' types are common knowledge.
$u_i^k \in \mathbb{R}^{m_i \times 1}$	Player i's action of dimension m_i at stage k.
$u^k := [u_1^k, \ldots, u_N^k]$	N players' joint action at stage k.
$u_i^{k_0:K} := [u_i^{k_0}, \ldots, u_i^K]$.	Player i's action sequence from k_0 to K.
$u^{k_0:K} := [u_i^{k_0:K}, u_{-i}^{k_0:K}]$.	Player i's and all other players' control sequences from stage k_0 to K.
$l_i^k(\theta_{-i}\|h^k, \theta_i)$	Player i's belief at stage k, i.e., the probability of other players' types being θ_{-i} based on player i's available information of h^k, θ_i.

I. Introduction

DECEPTION is a ubiquitous phenomenon in biology [1], military [2], politics and media [3], and cyberspace [4]. In particular, deception plays an increasingly significant role in cyber–physical systems, including autonomous vehicles and robots driven by artificial intelligence (AI). Recent advances in these AI-enabled technologies have not only allowed robots to adapt to the dynamic environment via real-time observations, but also made them deceivable. A deceiver can intentionally hide or reveal selected information to alter the beliefs and behaviors of the target robots for a higher reward. Since deception has many forms and delivery methods, understanding deception in a unified and quantitative framework is an indispensable step toward assessing the outcomes, measuring the impact, and designing strategies. This work aims to design robots that can interact with others efficiently under deceptive environments.

We identify the following challenges and features of robot deception. First, by definition, deception involves at least two participants interacting with each other. An intelligent robot should further consider other participants' rationality, predict their potential deceptive behaviors, and adjust its actions accordingly to alleviate the negative effect of deception. Second, due to the robots' dynamic nature, one-shot deception can exert a subsequent influence. The participating robots need to form long-term objectives to deceive or counter-deceive other robots. The multistage interactions also make it possible for the deceiver to apply deception at different stages. Third, each robot contains heterogeneous private information, which results in an asymmetric cognition structure; i.e., robots can form different beliefs over the same piece of unknown information. Thus, besides the couplings of state dynamics and costs, the multi-agent system further has cognitive coupling; i.e., each robot's behaviors are not only affected by its own belief but also the beliefs of the others.

To capture these features, we model the deceptive interaction between N strategic robots as a dynamic game of incomplete information. During the finite K stages of interaction, N robots accomplish non-cooperative tasks such as pursuit-evasion in the battlefield [5] or cooperative tasks such as collective towing [6]. Robots introduce deception in the above interacting scenarios due to antagonism, selfishness, and privacy concerns. Following Harsanyi's approach [7], we capture each robot's private information by a random variable. The realization of the random variable, which is called the robot's type, is known only to itself, while the support of the random variable, which contains all its possible types, is known to all robots. Take the pursuit-evasion scenario as an example, due to the constraints of weather, terrain, and weapon, both the evading and the pursuing robots know the feasible beachheads for the evader to land on. However, the evader chooses only one beachhead as his true target and the evader's choice, i.e., his type is unknown to the pursuer. The pursuer in the battlefield knows the existence of the deception and learns to counter the deception by forming and updating her belief based on real-time observations. Since these tasks are usually time-constrained, robots cannot wait and freeze until they have learned the true type. Instead, they have to take concurrent actions while the deceiver's type remains uncertain.

We consider two classes of belief dynamics based on whether robots exploit the intrinsic information such as the prediction of other robots' actions, or the extrinsic information to update their beliefs. Each robot aims to minimize its expected cumulative cost over K stages. Since the expectation involves its K-stage belief sequence of other players' private types, its actions should be sequentially rational under its belief sequence and the belief sequence should be consistent with the belief dynamics as well. These two requirements lead to the solution concept of perfect Bayesian Nash equilibrium (PBNE) where a player's unilateral deviation from the equilibrium increases his long-run cost. By appending the belief state (i.e., all players' beliefs under all possible types) to the system state, the PBNE computation is equivalent to a multi-agent nonlinear stochastic control problem and the method of dynamic programming applies. Without loss of generality, we characterize the structure of the action and the cost under PBNE as a feedback function of the belief state and the system state at the current stage. To provide an offline evaluation metric of the equilibrium cost under incomplete information, we use the expected equilibrium cost under complete information as a benchmark and define the price of deception (PoD).

Due to their tractability and generality, we focus on incomplete-information linear-quadratic (LQ) games with extrinsic belief dynamics to obtain the PBNE action that is unique and affine to the system state. We obtain a set of extended Riccati equations, which explicitly characterizes the coupling in the state dynamics, costs, and cognition of all robots. Under proper decoupling structures, the extended Riccati equations degenerate to the classical Riccati equations

for the problems of optimal control or complete-information LQ games. Under the incomplete-information LQ games with intrinsic belief dynamics, the equilibrium action is in general not affine feedback of the system state. Thus, we adopt a receding-horizon approach to provide a reasonable approximation of PBNE; i.e., instead of offline planning of all K-stage actions before the game starts, players recompute their actions based on the real-time observations and their updated beliefs at each new stage during the interaction.

Finally, we investigate a target protection problem where an evader aims to deceptively reach one of the possible targets and simultaneously evade the pursuer. The game has doubled-sided asymmetric information. The evader's private or hidden information is his true target while the pursuer's private information is her capability to maneuver or the maneuverability. We propose multi-dimensional metrics, including the stage of truth revelation and the endpoint distance, to assess the deception impact. We define the concept of deceivability to characterize the fundamental limits of deception and investigate how it is affected by the distinguishability of private information. We compare the proposed control policy with two heuristic policies to demonstrate its efficacy to counter deception at a much lower cost. We show that Bayesian learning can significantly reduce the impact of initial belief manipulation and result in a win–win situation for some cases. The increase of the pursuer's maneuverability improves her control performance under deception yet has a marginal effect. We also find that applying deception to counter deception is not always effective; e.g., it can be beneficial for a less maneuverable pursuer to disguise as a more maneuverable pursuer but not vice versa. The numerical results corroborate that PoD can exceed 1; i.e., deception among players may not only benefit the deceiver but also the deceivee.

A. Related Works

The secure and efficient operation of robots, autonomous vehicles, and industrial control systems is vital for recent advances in technologies. Many works [8]–[10] have investigated how to protect these systems from various attacks on sensor measurements [11], communication channels [12], and control signals [13], [14]. Deception is a key feature of sophisticated attacks with a focus on intentionally hiding private information [15], [16], introducing randomness [17], and manipulating other players' beliefs [18], [19]. Deception in robotic systems can be conducted through visual displays [20], facial expressions and body gestures [21], and trajectories [15], [22]. Existing works on robot deception are largely based on experimental approaches [15], [23], [24]. There is a need for a formal and quantitative framework to assess the deception impact, understand the fundamental limit and tradeoff of deception, and determine real-time strategies. Compared to the theoretical works of deceptive path planning and goal recognition [25], [26], which focus on identifying the true target behind deception, our work further determines optimal and cost-effective control policies to counteract deception and physically protect the true target; e.g., the pursuer adopts the action sequence of the minimum cost to reach and protect the true beachhead selected by the evader. Compared to control-theoretic deception frameworks based on Markov decision processes [17], [18] and stochastic games [27], we adopt a state–space representation to better characterize the physical dynamics of robots and autonomous vehicles.

Game models such as hypergames [28], dynamic Bayesian games [16], partially observable stochastic games [19], [29], and games that involve signaling mechanisms [30], [31] have been adopted as natural analytic paradigms to understand deception between intelligent players. The computation of equilibrium solutions for dynamic games of incomplete information, especially ones with non-classical information structure [32], is often a challenging task. Previous works have adopted conjugate prior assumptions to simplify Bayesian update and decouple the forward type estimation and backward action optimization under a finite state space and a continuous type space [33], [34]. To solve the coupling between players' belief dynamics and the multi-agent optimal control problem in the context of robotic systems where states are continuous and constrained by physical dynamics with noises, we adopt a receding-horizon approach to compute PBNE, which yields computationally tractable online strategies for the players. Similar receding-horizon approaches have been used in other contexts, including cyber-physical systems [35], military air operation [36], and autonomous racing [37].

B. Notations and Organization of the Article

Calligraphic letter $\mathscr{A}$ defines a set and $|\mathscr{A}|$ represents its cardinality. Define $\mathscr{B} \setminus \mathscr{A}$ as the set of elements in $\mathscr{B}$ but not in $\mathscr{A}$. The Euclidean norm of a vector x is represented by $||x||_2$. Let $\mathbb{E}_{a\sim A}[f(a)]$ denote the expectation of $f(a)$ over random variable a whose probability distribution is A. Let $'$ represent matrix transpose and $\mathrm{Diag}[a_1, \ldots, a_N]$ represent a block diagonal matrix with possibly non-square matrices $a_i, i \in \mathscr{N}$, on its diagonal. Define $\{a_i\}_{i\in\mathscr{N}} := \{a_1, \ldots, a_N\}$ as a set of N elements, $[a_i]_{i\in\mathscr{N}} := [a_1, \ldots, a_N]$ as N block matrices of the same number of rows arranged in one row vector, and $[a_1; \cdots; a_N] = [a_1, \ldots, a_N]'$ as N block matrices of the same number of columns arranged in one column vector. Let $\mathbf{I}_r, \mathbf{0}_{m,n}$ be the $r \times r$ identity matrix and the $m \times n$ zero matrix, respectively. The superscript $k \in \mathscr{K}$ is the stage index and the subscript $i \in \mathscr{N}$ is the player index. We omit a function's arguments when there is no ambiguity, e.g., $S_i^k := S_i^k(\beta_i^k, \theta_i)$. A piece of information for a group of players is called common knowledge if all players know it, all players know that all players know it, and so on ad infinitum. We summarize main notations in Nomenclature.

The rest of the article is organized as follows. Section II introduces the dynamic game of incomplete information and the solution concept of PBNE. To obtain explicit and practical solutions, we consider a class of a LQ problems in Section III and obtain a set of extended Riccati equations. We present a case study of deceptive pursuit-evasion in Section IV and Section V concludes this article.

II. Dynamic Game With Private Types

We model deception as a K-stage game consisting of N robots as players and each robot has asymmetric information. Let $\mathscr{N} := \{1, \ldots, N\}$ be the set of N players and

$\mathscr{K} := \{0, 1, 2, \ldots, K\}$ be the set of K discrete stages. Private information of player $i \in \mathscr{N}$, i.e., his type θ_i, is modeled as the realization of a discrete random variable with a finite support $\Theta_i := \{\theta_i^1, \theta_i^2, \ldots, \theta_i^{N_i}\}$ and a prior probability distribution $\Xi_i(\cdot)$. Hence, N_i is the number of possible types for player i and $\Xi_i(\theta_i)$ is the probability that player i's type is θ_i. Define shorthand notation $\Xi := [\Xi_i]_{i\in\mathscr{N}}$ and let $\Theta_{-i} := \prod_{j\in\mathscr{N}\setminus\{i\}} \Theta_j$ be the set of types of all players except for player $i \in \mathscr{N}$. Each player i knows the value of his own type θ_i, but does not know the values of other players' types $\theta_{-i} := [\theta_j]_{j\in\mathscr{N}\setminus\{i\}} \in \Theta_{-i}$, throughout K stages of the game. The system state dynamics under N players' joint action $u^k := [u_1^k, \ldots, u_N^k]$, joint type $\theta := [\theta_1, \ldots, \theta_N]$, and an additive external noise $w^k \in \mathbb{R}^{n\times 1}$ are shown in the following equation:

$$x^{k+1} = f^k\big(x^k, u_1^k, \ldots, u_N^k, \theta_1, \ldots, \theta_N\big) + w^k, \quad k \in \mathscr{K} \setminus \{K\}. \tag{1}$$

The dynamics in (1) can have different interpretations based on applications. In the pursuit-evasion scenario as in [5], $x_i^k \in \mathbb{R}^{n_i\times 1}$ represents robot i's local states such as its location and speed. The system state $x^k \in \mathbb{R}^{n\times 1}$ can be explicitly represented by N robots' joint state $[x_1^k, \ldots, x_N^k]$ with $n = \sum_{i=1}^N n_i$. In the application where N robots cooperatively transport a payload, e.g., [6], [38], system state $x^k \in \mathbb{R}^{n\times 1}$ represents the payload's location and posture, which does not explicitly relate to robots' local states. The noise sequence $[w^k]_{k\in\mathscr{K}}$ assumed to be independent with probability density function $\Xi_w(\cdot)$, i.e., $\mathbb{E}_{w^k, w^h\sim\Xi_w}[w^k(w^h)'] = 0, \forall k \in \mathscr{K}, h \in \mathscr{K} \setminus \{k\}$. The noise is not necessarily Gaussian distributed but is assumed to have a zero mean, i.e., $\mathbb{E}_{w^k\sim\Xi_w}[w^k] = 0, \forall k \in \mathscr{K}$. We assume that system dynamics (1) are multi-agent controllable as defined in Definition 1 so that players can design their deceptive actions to reach the entire state space in finite stages.

Definition 1 (Multi-Agent Controllability): System dynamics (1) are called multi-agent controllable if for any target state $x^k \in \mathbb{R}^{n\times 1}$ at stage $k \in \mathscr{K} \setminus \{0\}$, initial state $x^0 \in \mathbb{R}^{n\times 1}$, and joint type $\theta \in \Theta$, there exists a sequence of finite joint actions $u^{0:k}$ that drive the system state from x^0 to x^k in expectation.

A. Forward Belief Dynamics

At each stage $k \in \mathscr{K}$, the information available to player i compromises all players' state history $h^k := [x^0, \ldots, x^k] \in \mathscr{H}^k$ as well as his own type value θ_i. Define $\Delta(\Theta_{-i})$ as the set of probability distributions over set Θ_{-i}. Each player i at stage k forms a belief $l_i^k : \mathscr{H}^k \times \Theta_i \mapsto \Delta\Theta_{-i}$ based on his available information. Thus, $l_i^k(\cdot|h^k, \theta_i)$ is a probability measure of other players' types, i.e., $\sum_{\theta_{-i}\in\Theta_{-i}} l_i^k(\theta_{-i}|h^k, \theta_i) = 1, \forall h^k \in \mathscr{H}^k, \theta_i \in \Theta_i$. Define a vector

$$\beta_i^k := \Big[l_i^k\big(\theta_{-i}|h^k, \theta_i^1\big), l_i^k\big(\theta_{-i}|h^k, \theta_i^2\big), \ldots, l_i^k\Big(\theta_{-i}|h^k, \theta_i^{N_i}\Big)\Big]_{\theta_{-i}\in\Theta_{-i}}$$

as player i's belief state at stage $k \in \mathscr{K}$. We assume that the set of belief states is independent of stages, i.e., $\beta_i^k \in \Lambda_i \subseteq [0, 1]^{|\Theta_{-i}|\times|\Theta_i|}$. Then, we can represent player i's belief dynamics as

$$\beta_i^{k+1} := \Gamma_i^k\big(\beta_i^k, u^k, w^k, \theta_i\big) \quad \forall k \in \{0, \ldots, K-1\}. \tag{2}$$

Note that the belief transition function Γ_i^k can be different for each i and k, i.e., players' belief updates can be heterogeneous and time-varying. Define $\beta^k := [\beta_i^k]_{i\in\mathscr{N}} \in \Lambda := \prod_{i\in\mathscr{N}} \Lambda_i$. In this work, we assume that the initial beliefs of all players of all types β^0 and the belief update rules $\Gamma_i^k, \forall i \in \mathscr{N}, \forall k \in \{0, \ldots, K-1\}$, are common knowledge. In Sections II-A1 and II-A2, we provide two specific forms of Γ_i^k that rely on intrinsic and extrinsic information, respectively.

1) Bayesian Belief Dynamics: The most common belief update rule Γ_i^k in (2) for player i at stage $k+1$ uses Bayesian inference. Given the knowledge of the sequential state observations x^k, x^{k+1} and all players' actions u^k, each player i of type $\theta_i \in \Theta_i$ at stage $k+1$ can update his belief as follows: $\forall \theta_{-i} \in \Theta_{-i}$,

$$l_i^{k+1}\big(\theta_{-i}|h^{k+1}, \theta_i\big) = \frac{l_i^k\big(\theta_{-i}|h^k, \theta_i\big)\Pr\big(x^{k+1}|\theta_{-i}, x^k, \theta_i\big)}{\sum_{\bar\theta_{-i}\in\Theta_{-i}} l_i^k\big(\bar\theta_{-i}|h^k, \theta_i\big)\Pr\big(x^{k+1}|\bar\theta_{-i}, x^k, \theta_i\big)}. \tag{3}$$

In (3), we use the Markov property, i.e., $\Pr(x^{k+1}|\theta_{-i}, h^k, \theta_i) = \Pr(x^{k+1}|\theta_{-i}, x^k, \theta_i) = \Xi_w(x^{k+1} - f^k(x^k, u^k, \theta))$. The denominator is positive as $w^k \in \mathbb{R}^{n\times 1}$.

Remark 1 (Actions Reveal Type Information): Even if the state dynamics f^k in (1) are independent of $\theta_j, \forall j \in \mathscr{N} \setminus \{i\}$, player $i \in \mathscr{N}$ can still learn player j' type via (3) as player j's action u_j^k is a function[1] of his type θ_j.

2) Markov-Chain Belief Dynamics: In Section II-A1, we assume that players can exploit the intrinsic information of state dynamics f^k, state observations x^k, x^{k+1}, and the prediction of all players' actions u^k. Since the above intrinsic information may not be available in practice, we consider the belief dynamics with extrinsic information in this section. In particular, we assume that each player i's belief dynamics $\beta_i^{k+1} := \Gamma_i^k(\beta_i^k, w^k, \theta_i), \forall k \in \{0, \ldots, K-1\}$ are a discrete-time Markov chain where the extrinsic information at stage k is characterized by the transition function $\Gamma_i^k(\cdot, w^k, \theta_i)$. Note that the transition function only characterizes how players update their beliefs at each stage yet does not guarantee that a player can learn the true types of others. The following example illustrates a class of players whose belief dynamics exhibit the confirmation bias [39] where players tend to ignore intrinsic evidence such as u^k and preserve their belief update rules Γ_i^k at each stage k.

Example 1: Consider a two-person game $N = 2$ where the first player has two types $N_1 = 2, \Theta_1 = \{\theta_1^1, \theta_1^2\}$ and the second player only has one type $N_1 = 1, \Theta_2 = \{\theta_2^1\}$. The second player's belief state $\beta_2^k = [l_2^k(\theta_1^1|\theta_2^1), l_2^k(\theta_1^2|\theta_2^1)]$ toward the first player's type belongs to a finite set $\Lambda_2 = \{[0.2, 0.8], [0.5, 0.5], [0.8, 0.2]\}$. The transition function Γ_2^k is independent of k: if the current belief state is $[0.5, 0.5]$, then the belief at the next stage is $[0.2, 0.8]$, $[0.5, 0.5]$, or $[0.8, 0.2]$ with probability $0.4, 0.2, 0.4$, respectively. If the current belief state is $[0.8, 0.2]$ (resp. $[0.2, 0.8]$), then the belief at the next stage is $[0.8, 0.2]$ (resp. $[0.2, 0.8]$) or $[0.5, 0.5]$ with probability 0.9 and 0.1, respectively. The above transition function Γ_2^k

[1]Each player's action is a function of his type as his cost is related to his type and the action aims to minimize his cost.

means that the second player tends to interpret the extrinsic information of the first player's type based on his current belief. If the second player already believes that the first player is of type θ_1^1 with a high probability of 0.8 at stage k, i.e., $\beta_2^k = [0.8, 0.2]$, then the second player is more inclined to enhance his current belief, i.e., his belief state at the next stage, i.e., β_2^{k+1}, will remain to be $[0.8, 0.2]$ with a high probability of 0.9. The above transition function represents the phenomena of attitude polarization and confirmation bias where players preserve their existing beliefs and the disagreement becomes more extreme at each stage even when players are exposed to the same evidence.

B. Nonzero-Sum Cost Function and Equilibrium Concept

At nonterminal stage $k \in \mathscr{K} \setminus \{K\}$, player i's cost function is $g_i^k : \mathbb{R}^{n\times 1} \times \prod_{j=1}^N \mathbb{R}^{m_j\times 1} \times \Theta_i \mapsto \mathbb{R}$. The final stage cost is $g_i^K : \mathbb{R}^{n\times 1} \times \Theta_i \mapsto \mathbb{R}$. Define $u_i^{k_0:K-1} := [u_i^{k_0}, \ldots, u_i^{K-1}]$ as player i's action sequence from stage k_0 to $K-1$ and $u^{k_0:K-1} := [u_i^{k_0:K-1}, u_{-i}^{k_0:K-1}]$ as player i's and all other players' action sequences from stage k_0 to $K-1$. Player i's expected cumulative cost from arbitrary initial stage $k_0 \in \mathscr{K}$ to the terminal stage K is defined as

$$\begin{aligned} J_i^{k_0}&\left(l_i^{k_0:K-1}, u^{k_0:K-1}, x^{k_0}, \theta_i\right) \\ &= \mathbb{E}_{w^{K-1}\sim\Xi_w}\left[g_i^K\left(x^K, \theta_i\right)\right] \\ &\quad + \sum_{k=k_0}^{K-1} \mathbb{E}_{w^{k-1}\sim\Xi_w}\left[\mathbb{E}_{\theta_{-i}\sim l_i^k}\left[g_i^k\left(x^k, u^k, \theta_i\right)\right]\right]. \end{aligned} \tag{4}$$

The expectations are taken first over the external noise sequence w^k and then over other players' internal type uncertainty. We cannot exchange the order of these two expectations as l_i^k is a function of w^{k-1}. Each player i at stage $k_0 \in \mathscr{K}$ aims to minimize $J_i^{k_0}$ by choosing only his action sequence $u_i^{k_0:K-1}$ but not other players' action sequence $u_{-i}^{k_0:K-1}$. The following definition of sequential rationality in Definition 2 guarantees that each player i has no motivation to deviate from the sequentially rational action at any stage $k \in \{k_0, \ldots, K-1\}$ during the interaction if all other players adopt the sequentially rational actions.

Definition 2 (Sequential Rationality): An action sequence $u^{*,k_0:K-1} := \{u_i^{*,k_0:K-1}, u_{-i}^{*,k_0:K-1}\}$ is called sequentially rational for player i under the belief sequence $l_i^{k_0:K-1}$, state x^{k_0}, and type θ_i, if for any state x^k at stage $k \in \{k_0, \ldots, K-1\}$, player i does not benefit from taking any other action sequence $u_i^{k:K-1}$, i.e., $J_i^k(l_i^{k:K-1}, u_i^{*,k:K-1}, u_{-i}^{*,k:K-1}, x^k, \theta_i) \le J_i^k(l_i^{k:K-1}, u_i^{k:K-1}, u_{-i}^{*,k:K-1}, x^k, \theta_i), \forall u_i^{k:K-1}$.

Since players' actions may affect their future beliefs as captured by the belief dynamics Γ_i^k in (2), we further require the equilibrium action $u^{*,k_0:K-1}$ in Definition 2 to be consistent with the belief dynamics, which leads to the following definition of PBNE.

Definition 3: (Perfect Bayesian Nash Equilibrium): Consider the N-player dynamic game of private types and asymmetric information defined by the state dynamics (1) and the expected cumulative cost (4). The action sequence $u^{*,0:K-1} := \{u_i^{*,0:K-1}, u_{-i}^{*,0:K-1}\}$ of N players over K stages compromises the PBNE if, regardless of each player i's type $\theta_i \in \Theta_i$, the following statements hold.

1) *Sequential Rationality:* $u^{*,0:K-1}$ is sequential rational for each player $i \in \mathscr{N}$ under his belief sequence $l_i^{*,0:K-1}$.
2) *Belief Consistency:* Each player i's belief sequence $l_i^{*,0:K-1}$ is consistent with (2) under $u^{*,0:K-1}$.

Proposition 1: It is sufficient to represent player i's equilibrium cost $J_i^k(l_i^{*,k:K-1}, u^{*,k:K-1}, x^k, \theta_i)$ under the PBNE action $u^{*,k:K-1}$ at stage $k \in \mathscr{K}$ as a function of β^k, x^k and θ_i, which is defined as $V_i^k(\beta^k, x^k, \theta_i)$. Under the boundary condition $V_i^K(\beta^K, x^K, \theta_i) := g_i^K(x^K, \theta_i)$, the following holds for all $k \in \{0, \ldots, K-1\}$ and all $x^k \in \mathbb{R}^{n\times 1}$, $\beta^k \in \Lambda$, that is:

$$\begin{aligned} &V_i^k\left(\beta^k, x^k, \theta_i\right) \\ &= \min_{u_i^k} \sum_{\theta_{-i}} l_i^k\left(\theta_{-i}|h^k, \theta_i\right)\left\{g_i^k\left(x^k, u^k, \theta_i\right)\right. \\ &\quad \left. + \mathbb{E}_{w^k\sim\Xi_w}\left[V_i^{k+1}\left(\beta^{k+1}, x^{k+1}, \theta_i\right)\right]\right\} \quad \forall\theta_i \in \Theta_i \ \forall i \in \mathscr{N} \end{aligned} \tag{5}$$

where β^{k+1} and x^{k+1} satisfy (2) and (1), respectively.

Proof: According to the definition of PBNE, at the second last stage $k = K-1$, each player i's equilibrium action $u_i^{*,k} = \arg\min_{u_i^k} \mathbb{E}_{\theta_{-i}\sim l_i^k}[g_i^k(x^k, u^k, \theta_i)] + \mathbb{E}_{w^k\sim\Xi_w}[g_i^K(x^K, \theta_i)]$ is in general a function of $\theta_i, x^k, l_i^{*,k}, u_{-i}^{*,k}$. Due to the coupling between $u_i^{*,k}$ and $u_{-i}^{*,k}$, we need to solve a set of system equations for all $i \in \mathscr{N}$ and $\theta_i \in \Theta_i$. Then, $u_i^{*,k}$ will be a function of β^k, x^k, θ_i and we obtain (5) at stage $k = K-1$. We can repeat the above procedure from $k = K-2$ to $k = 0$ to obtain the recursive form in (5). □

Proposition 1 characterizes the structure of the equilibrium action $u_i^{*,k}$ and the equilibrium cost $V_i^k(\beta^k, x^k, \theta_i)$ for each player i of type θ_i under the solution concept of PBNE; i.e., both terms are feedback functions of the belief state β^k, the physical state x^k, and the player' type θ_i. Although J_i^k is a function of beliefs $l_i^{k:K-1}$ over all the remaining stages, $V_i^k(\beta^k, x^k, \theta_i)$ only depends on the belief state at the current stage k. If all players' types are common knowledge, PBNE still applies and we can define a new function $\bar{V}_i^k(x^k, \theta)$ to represent the resulting equilibrium cost $V_i^k(\beta^k, x^k, \theta_i)$ for all $k \in \mathscr{K}$ without loss of generality.

C. Offline Evaluation of Equilibrium Cost

If each player i's initial belief confirms to the prior distribution of other players' types, i.e., $l_i^0(\theta_j|x^0, \theta_i) = \Xi_j(\theta_j)$, $\forall\theta_i \in \Theta_i$, $j \in \mathscr{N}, \theta_j \in \Theta_j, \forall x^0$, then each player i at system state x^0 with belief state β^0 can use his expected equilibrium cost $\mathbb{E}_{\theta_i\sim\Xi_i}[V_i^0(\beta^0, x^0, \theta_i)]$ over his type uncertainty Ξ_i as an offline performance measure of the equilibrium action $u^{*,0:K}$. As a comparison, player i's expected equilibrium cost $\mathbb{E}_{\theta\sim\Xi}[\bar{V}_i^0(x^0, \theta)]$ under the complete information game serves as a benchmark. Note that player i does not need to know the realization of the joint type θ to compute $\mathbb{E}_{\theta\sim\Xi}[\bar{V}_i^0(x^0, \theta)]$. Due to the coupling in dynamics, costs, and cognition among N players, obtaining more information and knowing the type of another player $j \in \mathscr{N}\setminus\{i\}$ may not always improve player i's performance; i.e., there is no guarantee

that $\mathbb{E}_{\theta_i\sim\Xi_i}[V_i^0(\beta^0,x^0,\theta_i)] \geq \mathbb{E}_{\theta\sim\Xi}[\bar{V}_i^0(x^0,\theta)]$. Besides the above performance evaluation for an individual player $i\in\mathscr{N}$ under deception, we may also aim to evaluate the overall performance of multiple players or all N players. We define the PoD in Definition 4 with a set of coefficients $\eta_i\in[0,1]$, $\forall i\in\mathscr{N}$, $\sum_{i\in\mathscr{N}}\eta_i=1$. Since the equilibrium cost can be negative, we let $\eta_0(\Xi):=-\min(0,\{\mathbb{E}_{\theta_i\sim\Xi_i}[V_i^0(\beta^0,x^0,\theta_i)]\}_{i\in\mathscr{N}},\{\mathbb{E}_{\theta\sim\Xi}[\bar{V}_i^0(x^0,\theta)]\}_{i\in\mathscr{N}})$ be the normalizing constant to guarantee that $p^\eta(\Xi)$ is non-negative for all chosen coefficients $\eta_i, i\in\mathscr{N}$.

Definition 4 (Price of Deception): For a given set of coefficients $\eta:=\{\eta_i\}_{i\in\mathscr{N}\cup\{0\}}$, the PoD of the N-player K-stage game defined by (1), (4), and (2) under the prior probability distribution $\Xi=[\Xi_i]_{i\in\mathscr{N}}$ is

$$p^\eta(\Xi):=\frac{\sum_{i\in\mathscr{N}}\eta_i\mathbb{E}_{\theta\sim\Xi}\big[\bar{V}_i^0\big(x^0,\theta\big)\big]+\eta_0(\Xi)}{\sum_{i\in\mathscr{N}}\eta_i\mathbb{E}_{\theta_i\sim\Xi_i}\big[V_i^0\big(\beta^0,x^0,\theta_i\big)\big]+\eta_0(\Xi)}\in[0,\infty).$$

The PoD is a crucial evaluation and design metric. We can endow PoD with different meanings by properly choosing the weighting coefficients $\eta_i, i\in\mathscr{N}$. For example, if besides N players, there is a central planner who aims to minimize the total cost of all N players under their deceptive interaction. Then, we can pick $\eta_i=1/N, i\in\mathscr{N}$, to represent the overall system performance. Although the central planner cannot control players' state dynamics, costs, and belief dynamics directly, he can still affect their deceptive interaction if he can design the prior probability distribution Ξ of the joint type θ. If the central planner instead only aims to reduce the cost of one player $j\in\mathscr{N}$, then we can pick $\eta_j=1$ and $\eta_h=0,\forall h\in\mathscr{N}\setminus\{j\}$. With a given weighting parameters η, a larger value of $p_\eta(\Xi)$ indicates a better accomplishment of the above goals. Note that individual deception may improve the system performance, i.e., $p^\eta(\Xi)>1$.

III. LQ Specification

LQ game is an important class of dynamic games. They can also be applied iteratively to approximate nonlinear stochastic systems with general cost functions and obtain equilibrium actions [40]. In Sections III and IV, we consider linear state dynamics:

$$f^k\big(x^k,u^k,\theta\big):=A^k(\theta)x^k+\sum_{i=1}^N B_i^k(\theta_i)u_i^k \tag{6}$$

with stage-varying matrices $A^k(\theta)\in\mathbb{R}^{n\times n}$, $B_i^k(\theta_i)\in\mathbb{R}^{n\times m_i}$.

Remark 2: System (6) is multi-agent controllable if and only if matrices $H_i^k(\theta):=[B_i^{k-1}(\theta_i),\ldots,\prod_{h=2}^{k-1}A^h(\theta)B_i^1(\theta_i),\prod_{h=1}^{k-1}A^h(\theta)B_i^0(\theta_i)]$, $\forall i\in\mathscr{N},\forall\theta\in\Theta,\forall k\in\mathscr{K}$, are of full rank as noise w^k has zero mean and we can obtain $\mathbb{E}[x^k]=\prod_{h=0}^{k-1}A^h(\theta)x^0+\sum_{r=1}^N H_r^k(\theta)[u_r^{k-1};\cdots;u_r^0]$ by induction.

Each player i's cost is quadratic in both x^k and u^k; that is

$$\begin{aligned}g_i^k\big(x^k,u^k,\theta_i\big)=&\big(x^k-\hat{x}_i^k(\theta_i)\big)'D_i^k(\theta_i)\big(x^k-\hat{x}_i^k(\theta_i)\big)\\&+\hat{f}_i^k\big(\hat{x}_i^k(\theta_i)\big)+\sum_{j=1}^N\big(u_j^k\big)'F_{ij}^k(\theta_i)u_j^k\quad\forall k\in\mathscr{K}\end{aligned}\tag{7}$$

where $[\hat{x}_i^k(\theta_i)]_{k\in\mathscr{K}}$ is a known type-dependent reference trajectory for player $i\in\mathscr{N}$ and $\hat{f}_i^k$ is a known function of $\hat{x}_i^k(\theta_i)$. The cost matrices $D_i^k(\theta_i)\in\mathbb{R}^{n\times n}$, $F_{ij}^k(\theta_i)\in\mathbb{R}^{m_i\times m_i}$, $\forall i,j\in\mathscr{N},k\in\mathscr{K}$, are symmetric. At the final stage, $F_{ij}^K(\theta_i)\equiv\mathbf{0}_{m_i,m_i},\forall i,j\in\mathscr{N},\forall\theta_i\in\Theta_i$. We introduce the following three sets of notations for the belief matrix, the extended Riccati equations, and the matrix-form equilibrium action, respectively.

A. Belief Matrix

With a little abuse of notation, we can define the marginal probability $l_i^k(\theta_j|h^k,\theta_i):=\sum_{\theta_r\in\Theta_r,r\in\mathscr{N}\setminus\{i,j\}}l_i^k(\theta_{-i}|h^k,\theta_i)$, $\forall j\in\mathscr{N}\setminus\{i\}$, as the player i's belief toward the player j's type at stage k. Define the belief matrix for all $i\in\mathscr{N}$, $j\in\mathscr{N}\setminus\{i\},k\in\{0,\ldots,K-1\}$, as

$$\mathbf{L}_{ij}^k:=\begin{bmatrix}\mathbf{L}_i^k\big(\theta_j^1|h^k,\theta_i^1\big),&\cdots&\mathbf{L}_i^k\big(\theta_j^{N_j}|h^k,\theta_i^1\big)\\\mathbf{L}_i^k\big(\theta_j^1|h^k,\theta_i^2\big),&\cdots&\mathbf{L}_i^k\big(\theta_j^{N_j}|h^k,\theta_i^2\big)\\\vdots&\ddots&\vdots\\\mathbf{L}_i^k\big(\theta_j^1|h^k,\theta_i^{N_i}\big),&\cdots&\mathbf{L}_i^k\big(\theta_j^{N_j}|h^k,\theta_i^{N_i}\big)\end{bmatrix}\tag{8}$$

where each block element $\mathbf{L}_i^k(\theta_j^r|h^k,\theta_i^h)=\mathrm{Diag}[l_i^k(\theta_j^r|h^k,\theta_i^h),\ldots,l_i^k(\theta_j^r|h^k,\theta_i^h)]\in\mathbb{R}^{n\times n},\forall r\in\{1,\ldots,N_j\}$, $\forall h\in\{1,\ldots,N_i\}$. Since all its elements are positive and all rows sum to one, the belief matrix $\mathbf{L}_{ij}^k$ is a right stochastic matrix.

B. Extended Riccati Equations

Let a sequence of symmetric matrices $S_i^k(\beta^k,\theta_i)\in\mathbb{R}^{n\times n}$, vectors $N_i^k(\beta^k,\theta_i)\in\mathbb{R}^{n\times 1}$, and scalars $q_i^k(\beta^k,\theta_i)\in\mathbb{R}$ satisfy the following extended Riccati equations for all $\beta^k\in\Lambda$, $i\in\mathscr{N},\theta_i\in\Theta_i,k\in\{0,\ldots,K-1\}$:

$$\begin{aligned}S_i^k=&D_i^k+\mathbb{E}_{\theta_{-i}\sim l_i^k}\Bigg[\bigg(A^k+\sum_{j=1}^N B_j^k\Psi_j^{1,k}\bigg)'\mathbb{E}_{w^k\sim\Xi_w}\big[S_i^{k+1}\big]\\&\cdot\bigg(A^k+\sum_{j=1}^N B_j^k\Psi_j^{1,k}\bigg)+\sum_{j=1}^N\Big(\Psi_j^{1,k}\Big)'F_{ij}^k\Psi_j^{1,k}\Bigg]\end{aligned}\tag{9}$$

$$\begin{aligned}N_i^k=&-2D_i^k\hat{x}_i^k+\mathbb{E}_{\theta_{-i}\sim l_i^k}\Bigg[\bigg(\sum_{j=1}^N B_j^k\Psi_j^{1,k}+A^k\bigg)'\big(\mathbb{E}_{w^k\sim\Xi_w}\big[N_i^{k+1}\big]\\&+2\mathbb{E}_{w^k\sim\Xi_w}\big[S_i^{k+1}\big]\sum_{j=1}^N B_j^k\Psi_j^{2,k}\bigg)+2\sum_{j=1}^N\Big(\Psi_j^{1,k}\Big)'F_{ij}^k\Psi_j^{2,k}\Bigg]\end{aligned}\tag{10}$$

$$\begin{aligned}q_i^k=&\big(\hat{x}_i^k\big)'D_i^k\hat{x}_i^k+\hat{f}_i^k\big(\hat{x}_i^k\big)+\mathbb{E}_{w^k\sim\Xi_w}\Big[\big(w^k\big)'S_i^{k+1}w^k+q_i^{k+1}\Big]\\&+\mathbb{E}_{\theta_{-i}\sim l_i^k}\Bigg[\bigg(\sum_{j=1}^N B_j^k\Psi_j^{2,k}\bigg)'\mathbb{E}_{w^k\sim\Xi_w}\big[S_i^{k+1}\big]\sum_{j=1}^N B_j^k\Psi_j^{2,k}\\&+\bigg(\sum_{j=1}^N B_j^k\Psi_j^{2,k}\bigg)'\mathbb{E}_{w^k\sim\Xi_w}\big[N_i^{k+1}\big]+\sum_{j=1}^N\Big(\Psi_j^{2,k}\Big)'F_{ij}^k\Psi_j^{2,k}\Bigg]\end{aligned}\tag{11}$$

where functions $\Psi_i^{1,k}, \Psi_i^{2,k}, \forall i \in \mathcal{N}$, are defined below. The boundary conditions of the extended Riccati equations are

$$S_i^K = D_i^K;\ N_i^K = -2D_i^K \hat{x}_i^K;\ q_i^K = \left(\hat{x}_i^K\right)' D_i^K \hat{x}_i^K + \hat{f}_i^K\left(\hat{x}_i^K\right). \tag{12}$$

C. Equilibrium Action in Matrix Form

We need to represent the equilibrium action of all players under all types in matrix form as each player's action is coupled with other players' actions under PBNE. Since each player i has different equilibrium actions under different types, with a little abuse of notation, we write each player i's action as a function of his type θ_i and define two action vectors $\mathbf{u}_i^k := [u_i^k(\theta_i^1), \ldots, u_i^k(\theta_i^{N_i})]' \in \mathbb{R}^{m_i N_i \times 1}$ and $\mathbf{u}^k := [\mathbf{u}_1^k, \mathbf{u}_2^k \ldots, \mathbf{u}_N^k]' \in \mathbb{R}^{\sum_{r=1}^N m_r N_r \times 1}$. For all $i \in \mathcal{N}, l_i^k, \theta_i \in \Theta_i, k \in \{0, \ldots, K-1\}$, define a series of (m_i)-by-(m_i) square matrices

$$R_i^k\left(\beta^k, \theta_i\right) := F_{ii}^k(\theta_i) + \left(B_i^k(\theta_i)\right)' S_i^{k+1}\left(\beta^k, \theta_i\right) B_i^k(\theta_i).$$

Let $\mathbf{B}_i^k := \mathrm{Diag}[B_i^k(\theta_i^1) \ldots, B_i^k(\theta_i^{N_i})]$ be $(N_i n)$-by-$(N_i m_i)$ block matrices and $\mathbf{S}_i^k(\beta^k) := \mathrm{Diag}[S_i^k(\beta^k, \theta_i^1), \ldots, S_i^k(\beta^k, \theta_i^{N_i})]$ be $(N_i n)$-by-$(N_i n)$ block matrices. Finally, define parameter matrices $\mathbf{W}^{1,k}(\beta^k) = [W_1^{1,k}(\beta^k); \ldots; W_N^{1,k}(\beta^k)] \in \mathbb{R}^{\sum_{r=1}^N m_r N_r \times n}$, $\mathbf{W}^{2,k}(\beta^k) = [W_1^{2,k}(\beta^k); \cdots; W_N^{2,k}(\beta^k)] \in \mathbb{R}^{\sum_{r=1}^N m_r N_r \times 1}$, and $\mathbf{W}^{0,k}(\beta^k) := [W_{ij}^{0,k}(\beta^k) \in \mathbb{R}^{m_i N_i \times m_j N_j}]_{i,j \in \mathcal{N}}$ for any $\beta^k \in \Lambda$. Their elements are given as follows, i.e., $\forall i \in \mathcal{N}, \forall k \in \{0, \ldots, K-1\}$:

$$\begin{aligned}
W_i^{1,k}\left(\beta^k\right) &= \Big[\left(B_i^k\left(\theta_i^1\right)\right)' S_i^{k+1}\left(\beta^k, \theta_i^1\right) \mathbb{E}_{\theta_{-i} \sim l_i^k}\left[A^k\left(\theta_i^1, \theta_{-i}\right)\right]; \cdots; \\
&\qquad \left(B_i^k\left(\theta_i^{N_i}\right)\right)' S_i^{k+1}\left(\beta^k, \theta_i^{N_i}\right) \mathbb{E}_{\theta_{-i} \sim l_i^k}\left[A^k\left(\theta_i^{N_i}, \theta_{-i}\right)\right]\Big] \\
W_i^{2,k}\left(\beta^k\right) &= \frac{1}{2}\Big[\left(B_i^k\left(\theta_i^1\right)\right)' N_i^{k+1}\left(\beta^k, \theta_i^1\right); \\
&\qquad \cdots; \left(B_i^k\left(\theta_i^{N_i}\right)\right)' N_i^{k+1}\left(\beta^k, \theta_i^{N_i}\right)\Big] \\
W_{ii}^{0,k}\left(\beta^k\right) &= \mathrm{Diag}\Big[R_i^k\left(\beta^k, \theta_i^1\right), \cdots, R_i^k\left(\beta^k, \theta_i^{N_i}\right)\Big] \\
W_{ij}^{0,k}\left(\beta^k\right) &= \left(\mathbf{B}_i^k\right)' \mathbf{S}_i^{k+1}\left(\beta^k\right) \mathbf{L}_{ij}^k \mathbf{B}_j^k \quad \forall j \in \mathcal{N} \setminus \{i\}.
\end{aligned}$$

Let matrix $\mathbf{M}_i^k(\beta^k, \theta_i^l) \in \mathbb{R}^{m_i \times \sum_{r=1}^N m_r N_r}, l \in \{1, 2, \ldots, N_i\}, i \in \mathcal{N}, k \in \{0, \ldots, K-1\}$, be the truncated row block, i.e., from row $\sum_{r=1}^{i-1} m_r N_r + m_i(l-1)$ to $\sum_{r=1}^{i-1} m_r N_r + m_i l$, of matrix $(-\mathbf{W}^{0,k}(\beta^k))^{-1}$. Define shorthand notations $\Psi_i^{1,k}(\beta^k, \theta_i) := \mathbf{M}_i^k(\beta^k, \theta_i)\mathbf{W}^{1,k}(\beta^k)$ and $\Psi_i^{2,k}(\beta^k, \theta_i) := \mathbf{M}_i^k(\beta^k, \theta_i)\mathbf{W}^{2,k}(\beta^k)$.

D. Extrinsic Belief Dynamics and Extended Riccati Equations

In this section, we focus on the extrinsic belief dynamics where Γ_i^k is independent of players' actions u^k for all $i \in \mathcal{N}, k \in \{0, \ldots, K-1\}$. The proof of Theorem 1 generalizes the one of classical LQ games (e.g., [41, Ch. 5.5 and 6.2]) where we further incorporate players' asymmetric belief dynamics into their objective functions to minimize their expected costs under deception. We apply dynamic programming from stage $K-1$ backward to stage 0 to obtain a closed-form solution of PBNE.

Theorem 1: An N-player K-stage LQ game of incomplete information defined by (6) and (7), and extrinsic belief dynamics $\beta_i^{k+1} = \Gamma_i^k(\beta_i^k, w^k, \theta_i), \forall i \in \mathcal{N}, \forall k \in \{0, \ldots, K-1\}$, admits a unique state-feedback PBNE

$$u_i^{*,k}\left(\beta^k, x^k, \theta_i\right) = \Psi_i^{1,k}\left(\beta^k, \theta_i\right) x^k + \Psi_i^{2,k}\left(\beta^k, \theta_i\right) \tag{13}$$

if and only if $R_i^k(\beta^k, \theta_i)$ is positive definite and $\mathbf{W}^{0,k}(\beta^k)$ is non-singular for all $\beta^k \in \Lambda, i \in \mathcal{N}, \theta_i \in \Theta_i, k \in \{0, \ldots, K-1\}$. The equilibrium cost V_i^k is quadratic in x^k, that is

$$\begin{aligned}
V_i^k\left(\beta^k, x^k, \theta_i\right) = {} & q_i^k\left(\beta^k, \theta_i\right) + \left(x^k\right)' N_i^k\left(\beta^k, \theta_i\right) \\
& + \left(x^k\right)' S_i^k\left(\beta^k, \theta_i\right) x^k \quad \forall i \in \mathcal{N},\ k \in \mathcal{K}.
\end{aligned} \tag{14}$$

Proof: We use backward induction to prove the result. At the final stage K, the value function $V_i^K(\beta^K, x^K, \theta_i) = (x^K - \hat{x}_i^K(\theta_i))' D_i^K(\theta_i)(x^K - \hat{x}_i^K(\theta_i)) + \hat{f}_i^K(\hat{x}_i^K(\theta_i))$ is quadratic in x^K and we obtain the boundary conditions for S_i^K, N_i^K, q_i^K in (12) by matching the right-hand side (RHS) of (14). At any stage $k \in \{0, \ldots, K-1\}$, if (14) is true at stage $k+1$, we can expand $\mathbb{E}_{w^k \sim \Xi_w}[V_i^{k+1}(\beta^{k+1}, x^{k+1}, \theta_i)]$ by plugging in the state dynamics $x^{k+1} = A^k(\theta)x^k + \sum_{i=1}^N B_i^k(\theta_i)u_i^k + w^k$ and the belief dynamics $\beta_i^{k+1} = \Gamma_i^k(\beta_i^k, w^k, \theta_i)$. Then, the RHS of (5) is quadratic in u_i^k for each player i. If the coefficient matrix R_i^k of the quadratic form $(u_i^k)' R_i^k u_i^k$ is positive definite, then the first-order necessary conditions for minimization are also sufficient and we obtain the following unique set of equations for the equilibrium action $u^{*,k}$ by differentiating the RHS of (5) and setting it to zero, i.e., $\forall \theta_i \in \Theta_i$:

$$\begin{aligned}
& -R_i^k u_i^{*,k}(\theta_i) \\
& = \left(B_i^k\right)' S_i^{k+1} \mathbb{E}_{\theta_{-i} \sim l_i^k}\left[A^k\right] x^k + \frac{1}{2}\left(B_i^k\right)' N_i^{k+1} \\
& \quad + \left(B_i^k\right)' S_i^{k+1} \sum_{j \neq i} \mathbb{E}_{\theta_j \sim l_i^k}\left[B_j^k(\theta_j) u_j^{*,k}(\theta_j)\right] \quad \forall i \in \mathcal{N}.
\end{aligned} \tag{15}$$

Due to the coupling in players' actions and beliefs, we rewrite (15) in matrix form, i.e., $-\mathbf{W}^{0,k}(\beta^k)\mathbf{u}^{*,k} = \mathbf{W}^{1,k}(\beta^k)x^k + \mathbf{W}^{2,k}(\beta^k)$, to solve the set of equations. Given the existence of $(-\mathbf{W}^{0,k}(\beta^k))^{-1}$, each player i's equilibrium action is an affine function in x^k, i.e., $u_i^{*,k}(\beta^k, x^k, \theta_i) = \Psi_i^{1,k}(\beta^k, \theta_i)x^k + \Psi_i^{2,k}(\beta^k, \theta_i)$. Note that the coefficients $\Psi_i^{1,k}, \Psi_i^{2,k}$ for player i are functions of β^k, i.e., the beliefs of all players under all types at stage k. Finally, after substituting the equilibrium action $u_i^{*,k}(\beta^k, x^k, \theta_i) = \Psi_i^{1,k}(\beta^k, \theta_i)x^k + \Psi_i^{2,k}(\beta^k, \theta_i)$ into the RHS of (5) and representing V_i^k in the left-hand side (LHS) in its quadratic form of x^k, we can match the coefficients of quadratic, linear, and constant terms in the LHS and RHS to obtain the extended Riccati equations (9)–(11). □

Remark 3 (Positive Definiteness): If $D_i^k(\theta_i)$ and $F_{ij}^k(\theta_i)$, $\forall j \in \mathcal{N}$, are positive definite for all $k \in \mathcal{K}$, then $R_i^k(\beta^k, \theta_i)$ is positive definite for all $k \in \mathcal{K}, \beta^k \in \Lambda$, because the linear combination of positive definite matrices in (9) preserves positive definiteness. Note that the above condition is only a necessary condition; i.e., D_i^k and F_{ij}^k do not need to be positive definite to make R_i^k positive definite as shown in Section IV.

Remark 4 (Cognitive Coupling): Compared with the classical LQ games (e.g., [41, Ch. 6]), the deception of players' types results in a unique feature of cognitive coupling represented by the belief matrix in (8); i.e., each player's action hinges on not only his own belief but also all other players' beliefs as these beliefs can affect their actions and further the outcome of the interaction. Thus, player i can change other players' actions by manipulating their beliefs of his type θ_i, i.e., $l_j^k, \forall j \in \mathscr{N} \setminus \{i\}$, or making them believe that his belief l_i^k on their types θ_{-i} has changed.

We introduce matrix block partitions as follows. For each type $\theta_i \in \Theta_i$, we divide $A^k(\theta), D_i^k(\theta_i), S_i^k(\theta_i)$ into N-by-N blocks where the (i,i) block is $A_i^k(\theta), \bar{D}_i^k(\theta_i), \bar{S}_i^k(\theta_i) \in \mathbb{R}^{n_i \times n_i}$, respectively. The ith row block of $N_i^k(\theta_i), \hat{x}_i^k(\theta_i)$ is $\bar{N}_i^k(\theta_i), \bar{x}_i^k(\theta_i) \in \mathbb{R}^{n_i \times 1}$, respectively. The i-th row block of $B_i^k(\theta_i)$ is $\bar{B}_i^k(\theta_i) \in \mathbb{R}^{n_i \times m_i}$. When the system state x^k can be represented by players' joint states $[x_i^k]_{i \in \mathscr{N}}$, Corollary 1 shows that the LQ game of asymmetric information degenerates to an LQ control problem if players have decoupled cost and state dynamics defined as follows.

Definition 5 (Decoupled Dynamics and Cost): Player $i \in \mathscr{N}$ has decoupled dynamics if for all $k \in \mathscr{K}$, $A_i^k(\theta) = \bar{A}_i^k(\theta_i), \forall \theta \in \Theta$, while all other elements in the ith row block and the ith column block of $A^k(\theta)$ are 0. Besides, all elements of $B_i^k(\theta_i)$ except for the row block $\bar{B}_i^k(\theta_i)$ are required to be 0. Player $i \in \mathscr{N}$ has a decoupled cost if for all stage $k \in \mathscr{K}$, $F_{ij}^k(\theta_i) = \mathbf{0}_{m_i, m_i}, \forall \theta_i \in \Theta_i, j \in \mathscr{N} \setminus \{i\}$, and all elements of $D_i^k(\theta_i)$ equal 0 except for $\bar{D}_i^k(\theta_i)$.

Corollary 1 (Degeneration to LQ Control): If $x^k = [x_i^k]_{i \in \mathscr{N}}$ for all stage $k \in \mathscr{K}$ and player i has both decoupled cost and state dynamics, then his action under PBNE is independent of other players' actions, types, and beliefs, i.e., $u_i^{*,k} = -(R_i^k)^{-1}(\bar{B}_i^k)' \bar{S}_i^{k+1} A_i^k x_i^k - \frac{1}{2}(R_i^k)^{-1}(\bar{B}_i^k)' \bar{N}_i^{k+1}$, where $R_i^k = F_{ii}^k + (\bar{B}_i^k)' \bar{S}_i^{k+1} \bar{B}_i^k$, $(G_i^k)' = \mathbf{I}_n - \bar{S}_i^{k+1} \bar{B}_i^k (R_i^k)^{-1} (\bar{B}_i^k)'$, $\bar{S}_i^k = (A_i^k)'(G_i^k)' \bar{S}_i^{k+1} A_i^k + \bar{D}_i^k$, and $\bar{N}_i^k = (A_i^k)'(G_i^k)' \bar{N}_i^{k+1} - 2\bar{D}_i^k \bar{x}_i^k$.

Proof: We show by induction that $S_i^k, N_i^k, \forall k \in \mathscr{K}$, satisfy the sparsity condition that only the (i,i) block of S_i^k and the i-th row block of N_i^k are nonzero. At stage K, $S_i^K = D_i^K$ and $N_i^K = -2D_i^K \hat{x}_i^K$ satisfy the above condition. At stage $k \in \{0, \ldots, K-1\}$, if S_i^{k+1}, N_i^{k+1} satisfy the sparsity condition, $\mathbf{W}^{0,k}(\beta^k)$ becomes a diagonal block matrix where $W_{ij}^{0,k}(\beta^k) = \mathbf{0}_{m_i N_i, m_j N_j}$ and $\mathbf{M}_i^k(\beta^k, \theta_i) = -(R_i^k(\beta^k, \theta_i))^{-1}$ for all $\beta^k \in \Lambda$. Then, S_i^k, N_i^k satisfy the condition based on (9) and (10). □

E. Intrinsic Belief Dynamics and Receding-Horizon Control

If there exists a player $i \in \mathscr{N}$ whose belief dynamics Γ_i^k depend on intrinsic information at some stage $k \in \{0, \ldots, K-1\}$ as shown in (2), then the equilibrium action $u_i^{*,k}$ is in general a nonlinear function of x^k and the equilibrium cost V_i^k is not quadratic in x^k even under the LQ setting of (6) and (7). Besides the static cognitive coupling among N players in Remark 4, the intrinsic information of u^k in the belief update introduces another dynamic cognitive coupling between the forward belief dynamics via (2) and the backward equilibrium computation via (5), which makes it challenging to compute PBNE. To reduce the computational complexity and further obtain implementable actions, we adopt a receding-horizon approach that computes the sequentially rational action sequence of all the future stages $u^{*,k:K-1}$ at current stage $k \in \{0, \ldots, K-1\}$ assuming $\beta^{\bar{k}} = \beta^k$, $\forall \bar{k} \in \{k, \ldots, K-1\}$, yet only implements the current-stage action $u^{*,k}$. Then, at the new stage $k+1$, each player observes the new system state x^{k+1} and updates the belief to β^{k+1} and recomputes the entire action sequence $u^{*,k+1:K-1}$ under assumption of $\beta^{\bar{k}} = \beta^{k+1}, \forall \bar{k} \in \{k+1, \ldots, K-1\}$, yet still only implements the new current-stage action $u^{*,k+1}$. Players repeat the above procedure until they reach the final stage of the interaction.

Compared with PBNE, which produces an offline planning for all future stages under all possible scenarios before the game has taken place, the receding-horizon approach enables an online replanning of their actions repeatedly at the beginning of each new stage as the interaction continues. Although we assume that players' beliefs at the future stages are the same as the current beliefs during the phase of equilibrium computation, players can correct and update their beliefs and actions based on the online observation of x^k during each replanning phase. Thus, the receding-horizon approach provides a reasonable approximation of the PBNE action and is more adaptive to unexpected environmental changes of the state dynamics f^k and cost structure $g_i^k, \forall i \in \mathscr{N}$.

Under the LQ specification in (6) and (7) and Bayesian belief dynamics in (3), we summarize the computation phase and online implementation phase in Algorithm 1 and 2, respectively. To investigate the scalability of our algorithms, we analyze the temporal and spatial complexity concerning N, K, and N_i. To simplify the notation and enhance readability, we focus on the symmetric setting where $N_i = N_0 \in \mathbb{Z}^+, \forall i \in \mathscr{N}$. For each player $i \in \mathscr{N}$ of type $\theta_i \in \Theta_i$ at the beginning of the interaction, i.e., $k = 0$, he needs to store the game parameters $A^0, B_r^0(\theta_r), D_r^0(\theta_r), F_{rh}^0(\theta_r), \forall \theta_r \in \Theta_r$, and the belief matrix $\mathbf{L}_{rh}^0$ for all $r, h \in \mathscr{N}$, which are common knowledge. The spatial complexity to store the game parameters and the belief matrix is $O(N^2\ N_0)$ and $O(N^2\ N_0^2)$, respectively. Note that in general, player i has coupled cognition as shown in Remark 4 and has to keep track of not only his belief $\mathbf{L}_{i,j}^k, \forall j \in \mathscr{N}$, but also other players' beliefs $\mathbf{L}_{r,h}^k, \forall r \in \mathscr{N} \setminus \{i\}, h \in \mathscr{N}$, to decide his equilibrium action under deception at each stage k. During the K-stage interaction, each player $i \in \mathscr{N}$ of type $\theta_i \in \Theta_i$ observes the system state x^k and computes his equilibrium action $u_i^{*,k}(\beta^k, x^k, \theta_i)$ at stage k based on Algorithm 1. After all players implement their equilibrium actions at stage k, the system state evolves to x^{k+1}. Based on the new state observation x^{k+1}, each player i updates the belief matrix in (8) via (3). Since player i can delete the game parameters and the belief matrices of previous stages, the spatial complexity remains the same as the real-time stage index k increases. Thus, our algorithm can handle the interaction of long duration. All players repeat the above procedure stated in lines 14–17 of Algorithm 2 until reaching the terminal stage $k = K$.

The computational complexity of the belief matrix update in the line 15 of Algorithm 2 is $O(N_0^N N)$. For any β^k,

Algorithm 1 PBNE Computation With $\beta^{\bar{k}} = \beta^k, \forall \bar{k} \in \{k, \ldots, K-1\}$ at Stage $k \in \{0, \ldots, K-1\}$ for Player $i \in \mathcal{N}$ of Type $\theta_i \in \Theta_i$

Load game parameters A^k, $B_r^k(\bar{\theta}_r)$, $D_r^k(\bar{\theta}_r)$, $F_{rh}^k(\bar{\theta}_r)$, $\forall \bar{\theta}_r \in \Theta_r$ and the belief matrix $\mathbf{L}_{r,h}^k$ for all $r, h \in \mathcal{N}$;
Input state observation x^k;
for $\bar{k} \leftarrow K-1$ **to** k **do**
  for $j \leftarrow 1$ **to** N **do**
    for $\theta_j \leftarrow \theta_j^1$ **to** $\theta_j^{N_j}$ **do**
      Compute $S_j^{\bar{k}}, N_j^{\bar{k}}$ via (9), (10) with $\beta^{\bar{k}} = \beta^k$;
    end
  end
end
Return his equilibrium action $u_i^{*,k}(I_i^k, x^k, \theta_i)$ via (13);

Algorithm 2 K-Stage Receding-Horizon Control for Player $i \in \mathcal{N}$ of Type $\theta_i \in \Theta_i$

Initialize $k = 0$;
Store game parameters A^k, $B_r^k(\bar{\theta}_r)$, $D_r^k(\bar{\theta}_r)$, $F_{rh}^k(\bar{\theta}_r)$, $\forall \bar{\theta}_r \in \Theta_r$ and the belief matrix $\mathbf{L}_{r,h}^k$ for all $r, h \in \mathcal{N}$;
while $k < K$ **do**
  Call Algorithm 1 to implement $u_i^{*,k}(I_i^k, x^k, \theta_i)$;
  Observe state x^{k+1} and update all elements of the belief matrix via (3) to obtain $\mathbf{L}_{r,h}^{k+1}, \forall r, h \in \mathcal{N}$;
  Delete A^k, $B_r^k(\bar{\theta}_r)$, $D_r^k(\bar{\theta}_r)$, $F_{rh}^k(\bar{\theta}_r)$, $\mathbf{L}_{r,h}^k$ and **Store** A^{k+1}, $B_r^{k+1}(\bar{\theta}_r)$, $D_r^{k+1}(\bar{\theta}_r)$, $F_{rh}^{k+1}(\bar{\theta}_r)$, $\mathbf{L}_{r,h}^{k+1}$ for all $\bar{\theta}_r \in \Theta_r$ and for all $r, h \in \mathcal{N}$;
  Update stage index $k \leftarrow k+1$;
end

the term $\mathbf{W}^{0,k}(\beta^k)$ has computational complexity $O(N_0^N N) + O(N_0^3\ N^2)$, which is determined by the belief matrix update and the matrix chain multiplication of $W_{ij}^{0,k}(\beta^k)$, respectively. Then, the computational complexity of $(\mathbf{W}^{0,k}(\beta^k))^{-1}$ and $\mathbf{W}^{1,k}(\beta^k)$ is $O(N_0^N N)+O(N_0^3\ N^3)$ and $O(N_0^N N)+O(N_0^3 N^2)$, respectively. Given β^k and θ_i, the computational complexity of $S_i^k(\beta^k, \theta_i)$ in (9) is $O(N_0^N N) + O(N_0^3 N^3) + O(N_0^3 N^2) + O(N_0 N) = O(\max(N_0^N N, N_0^3 N^3))$, which hinges on the computational complexity of $\mathbf{M}_i^k(\beta^k, \theta_i)$ (or $(\mathbf{W}^{0,k}(\beta^k))^{-1}$), $\mathbf{W}^{1,k}(\beta^k)$, and the matrix chain multiplication in (9). Similarly, $N_i^k(\beta^k, \theta_i)$ and $W^{2,k}(\beta^k)$ both have computational complexity of $O(N_0^N N) + O(N_0 N)$. Therefore, player i's temporal complexity at each stage $k \in \{0, 1, \ldots, K-1\}$ is

$$O\big((K-k) \cdot N_0 N \cdot \max\big(N_0^N N, N_0^3 N^3\big)\big).$$

The temporal complexity has the maximum value of $O(K \cdot \max\{N_0^{N+1} N^2, N_0^4 N^4\})$ at the initial stage $k = 0$ where each player has to predict the entire K future stages to act optimally under the deception. Since the temporal complexity decreases as the real-time stage index k increases, a player who can compute the equilibrium action within the required time at the initial stage $k = 0$ is guaranteed to meet the real-time requirement in the following stages of interaction. If the number of types and agents are on the same scale, e.g., $N_0 = N$, then $\lim_{N\to\infty}(N_0^{N+1} N^2)/(N_0^4 N^4) \to \infty$ and the computation of belief matrix update plays a dominant role as each player keeps track of all players' beliefs to obtain the equilibrium action under deception. If $N_0 \ll N$, e.g., $N_0 = N^{1/N}$, then $\lim_{N\to\infty}(N_0^{N+1} N^2)/(N_0^4 N^4) \to 0$ and the inverse of $\mathbf{W}^{0,k}(\beta^k)$ becomes the most time-consuming operation due to the coupling in dynamics, costs, and cognition.

Effective deception can prevent or delay other players from learning the deceiver's private type. We define the criterion of successful learning of the deceiver's type in Definition 6 and ϵ-deceviability and ϵ-learnability in Definition 7.

Definition 6 (Stage of Truth Revelation): Consider two players $i, j \in \mathcal{N}$ with type θ_i and θ_j, respectively. Stage $k_{i,j}^{tr} \in \mathcal{K} \cup \{K+1\}$ is said to be player i's truth-revealing stage with accuracy $\delta \in (0, 1]$[2] if it satisfies the following two conditions.

1) *The Bounded Mismatch Condition:* Player i's belief mismatch remains less than δ after stage $k_{i,j}^{tr} \in \mathcal{K}$, that is
$$1 - l_i^k\big(\theta_j | h^k, \theta_i\big) \le \delta \quad \forall k \ge k_{i,j}^{tr}. \tag{16}$$
2) *The First-Hitting-Time Condition:* $k_{i,j}^{tr} \in \mathcal{K}$ is the first stage satisfying (16), i.e., $1 - l_i^{k_{i,j}^{tr}-1}(\theta_j | h^{k_{i,j}^{tr}-1}, \theta_i) > \delta, k_{i,j}^{tr} > 1$.

If there does not exist $k_{i,j}^{tr} \in \mathcal{K}$ that satisfies (16), we define $k_{i,j}^{tr} := K+1$. If there are only two players $N = 2$, we write $k_{i,j}^{tr}$ as k_i^{tr} without ambiguity.

Due to deceivers' deceptive actions and the external noises, the belief sequence may be fluctuant; i.e., there can exist $k < k_{i,j}^{tr}$ such that $1 - l_i^k(\theta_j | h^k, \theta_i) \le \delta$. Thus, as shown in Definition 6, a player should only claim successful learning of other players' types if his belief mismatch remains less than δ for the remaining stages.

Definition 7 (Deceviability and Learnability): Consider players $i, j \in \mathcal{N}$ with type θ_i and θ_j, thresholds $\delta \in (0, 1]$, $\epsilon \in [0, 1]$, and a given stage index $\tilde{k} \in \mathcal{K} \cup \{K+1\}$. Player i is $\tilde{k}$-stage ϵ-deceivable if the probability $\Pr(k_{i,j}^{tr} < \tilde{k})$, or equivalently $\Pr(l_i^{\tilde{k}}(\theta_j | x^{\tilde{k}}, \theta_i) > 1-\delta)$, is not greater than ϵ for all $l_i^0 \in (0, 1)$. If the above does not hold, player j's type is said to be $\tilde{k}$-stage ϵ-learnable by player i.

Since robot deception involves only a finite number of stages, it is essential that the deceived robot can learn the deceiver's type as quickly as possible so that he has sufficient stages to plan on and mitigate the deception impact from the previous stages. Therefore, the definition of learnability, i.e., non-deceviability in Definition 7, not only requires the deceived player to be capable of learning the deceiver's private information, but also learning it in a desirable rate, i.e., within $\tilde{k}$ stage. Due to the external noise, $k_{i,j}^{tr}$ is a random variable. Thus, the definition of learnability requires $\Pr(k_{i,j}^{tr} < \tilde{k}) > \epsilon$; i.e., player i has a large probability to correctly learn the type of player j before stage $\tilde{k}$.

IV. Dynamic Target Protection Under Deception

We investigate a pursuit-evasion scenario that contains two UAVs with the decoupled linear time-invariant state dynamics,

[2]Since the belief mismatch does not reduce to 0 in finite stages with initial belief $l_i^0 \in (0, 1)$, the accuracy threshold $\delta \neq 0$.

i.e., $A^k(\theta) = \mathbf{I}_4, \tilde{B}_i^k(\theta_i) = [\tilde{B}_i(\theta_i), 0; 0, \tilde{B}_i(\theta_i)] \in \mathbb{R}^{2\times 2}$, $\forall k \in \mathcal{K}$. We use "she" for UAV 1, the pursuer, and "he" for UAV 2, the evader. UAV i's state $x_i^k := [x_{i,x}^k, x_{i,y}^k]' \in \mathbb{R}^{2\times 1}$ represents i's location $(x_{i,x}^k, x_{i,y}^k)$ in the 2-D space, and action $u_i^k = [u_{i,x}^k, u_{i,y}^k] \in \mathbb{R}^{2\times 1}$ affects i's speed in x- and y-directions.

UAV 2 as the evader selects either the harbor in "Normandy" or "Calais" as his final location based on his type $\theta_2 \in \{\theta_2^g, \theta_2^b\}$. He aims to reach "Normandy" located at $\gamma(\theta_2^g) := (x^g, y^g)$ in $K = 40$ stages if his type is θ_2^g, otherwise "Calais" located at $\gamma(\theta_2^b) := (x^b, y^b)$ if his type is θ_2^b. UAV 1 as the pursuer can make interfering signals and aims to be close to UAV 2 at the final stage to protect the harbor targeted by the evader, i.e., $g_1^k(x^k, u^k, \theta_1) = d_{12}^k(\theta_1)((x_{2,y}^k - x_{1,y}^k)^2 + (x_{2,x}^k - x_{1,x}^k)^2) + f_{11}^k(\theta_1)((u_{1,x}^k)^2 + (u_{1,y}^k)^2) - f_{12}^k(\theta_1)((u_{2,x}^k)^2 + (u_{2,y}^k)^2), \forall k \in \mathcal{K}$, where $d_{12}^k(\theta_1) \in \mathbb{R}_{\geq 0}$ penalizes her distance from the evader at stage $k \in \mathcal{K}$, $f_{11}^k(\theta_1) \in \mathbb{R}_{\geq 0}$ prevents her from a high action cost, and $f_{12}^k(\theta_1) \in \mathbb{R}_{\geq 0}$ incites her opponent, i.e., the evader, to take costly actions. We classify UAV 1 into two types, i.e., $\Theta_1 = \{\theta_1^H, \theta_1^L\}$, based on her maneuverability represented by the value of $\tilde{B}_1(\theta_1)$. Given higher maneuverability $\tilde{B}_1(\theta_1^H) > \tilde{B}_1(\theta_1^L)$, the pursuer of type θ_1^H can obtain a higher speed under the same action u_1^k and thus cover a longer distance.

The evader's goals of deceptive target reaching and pursuit evasion are incorporated into the cost structure $g_2^k(x^k, u^k, \theta_2) = d_{2,b}^k(\theta_2)((x_{2,y}^k - y^b)^2 + (x_{2,x}^k - x^b)^2) + d_{2,g}^k(\theta_2)((x_{2,y}^k - y^g)^2 + (x_{2,x}^k - x^g)^2) - d_{21}^k(\theta_2)((x_{1,y}^k - x_{2,y}^k)^2 + (x_{1,x}^k - x_{2,x}^k)^2) + f_{22}^k(\theta_2)((u_{2,x}^k)^2 + (u_{2,y}^k)^2) - f_{21}^k(\theta_2)((u_{1,x}^k)^2 + (u_{1,y}^k)^2), \forall k \in \mathcal{K}$. Similar to the pursuer's cost parameters, $d_{21}^k(\theta_2) \in \mathbb{R}_{\geq 0}$ represents the evader's level of evasion determination to keep a distance from the pursuer along the trajectory. The action costs of the evader and the pursuer are regulated by $f_{22}^k(\theta_2) \in \mathbb{R}_{\geq 0}$ and $f_{21}^k(\theta_2) \in \mathbb{R}_{\geq 0}$, respectively. The parameters $d_{2,b}^k(\theta_2)$ and $d_{2,g}^k(\theta_2)$ represent the evader's attempt to head toward "Normandy" and "Calais," respectively, at stage $k \in \mathcal{K}$ under type $\theta_2 \in \Theta_2$. We use the ratio $d_{2,g}^k(\theta_2)/d_{2,b}^k(\theta_2)$ to represent the evader's level of trajectory deception. Since the pursuer can learn the evader's type based on the real-time observations of state x_2^k, the evader attempts to make his target ϵ_0-ambiguous at all previous stages, i.e., $|d_{2,b}^k(\theta_2)/d_{2,g}^k(\theta_2) - 1| \leq \epsilon_0, \forall \theta_2, \forall k \neq K$, and reveal his true target only at the final stage, i.e., $d_{2,b}^K(\theta_2^b) = 0$ and $d_{2,b}^K(\theta_2^g) = 0$. The evader chooses a small $\epsilon_0 \geq 0$ and achieves the maximum ambiguity when $\epsilon_0 = 0$. Two blue lines in Fig. 1(a) illustrate how the evader manages to remain ambiguous in a cost-effective manner from two different initial locations. Instead of keeping an equal distance to both potential targets, the evader heads toward the midpoint $((x^g + x^b)/2, (y^g + y^b)/2)$ at the early stages to confuse the pursuer. However, the evader starts to head toward the true target at around half of K stages rather than the last few stages so that he can reach the target with a moderate control cost $(u_2^k)' F_{22}^k(\theta_2) u_2^k$. Fig. 1(a) also shows that for a given initial location, the evader who adopts a higher level of trajectory deception heads more toward the misleading target at the early stages.

In this case study, we suppose that the evader's true target is Calais and let θ_2^b be his true type and θ_2^g be the misleading type. The following two ratios capture the evader's tradeoff of being deceptive, effective, and evasive. On one hand, the ratio $d_{2,b}^k(\theta_2^b)/d_{2,b}^K(\theta_2^b), k \neq K$, reflects the evader's tradeoff between applying deception along the trajectory and staying close to the true target at the final stage. Fig. 1(b) shows that as the evader focuses more on a deceptive trajectory represented by a larger value of $d_{2,b}^k(\theta_2^b)/d_{2,b}^K(\theta_2^b), k \neq K$, his trajectory remains ambiguous for longer stages while his final location is farther away from the true target. On the other hand, the ratio $d_{21}^k(\theta_2^b)/d_{2,b}^K(\theta_2^b), k \neq K$, reflects the evader's tradeoff between evasion and target-reaching. As the evader focuses more on keeping a distance from the pursuer along the trajectory, he takes a bigger detour and stays farther away from his true target at the final stage as shown in Fig. 1(c).

Finally, we transform UAV i's coupled cost g_i^k into the matrix form given in Section III, i.e., $\hat{x}_1^k(\theta_1) = \mathbf{0}_{4,1}$, $\hat{f}_1^k(\hat{x}_1^k(\theta_1)) = 0$, $F_{ii}^k(\theta_1) = f_{ii}^k(\theta_1) \cdot \mathbf{I}_2$, $F_{ij}^k(\theta_1) = -f_{ij}^k(\theta_1) \cdot \mathbf{I}_2, j \neq i$, $D_1^k(\theta_1) = d_{12}^k(\theta_1) \cdot [1, 0, -1, 0; 0, 1, 0, -1; -1, 0, 1, 0; 0, -1, 0, 1]$

$$D_2^k(\theta_2) = \begin{bmatrix} -d_{21}^k & 0 & d_{21}^k & 0 \\ 0 & -d_{21}^k & 0 & d_{21}^k \\ d_{21}^k & 0 & d_{2,b}^k + d_{2,g}^k - d_{21}^k & 0 \\ 0 & d_{21}^k & 0 & d_{2,b}^k + d_{2,g}^k - d_{21}^k \end{bmatrix}$$

$\hat{x}_2^k(\theta_2) = (1/d_{2,b}^k + d_{2,g}^k) \cdot [d_{2,b}^k x^b + d_{2,g}^k x^g ; d_{2,b}^k y^b + d_{2,g}^k y^g ; d_{2,b}^k x^b + d_{2,g}^k x^g ; d_{2,b}^k y^b + d_{2,g}^k y^g]$, $\hat{f}_2^k(\hat{x}_2^k(\theta_2)) = (d_{2,b}^k d_{2,g}^k((x^b - x^g)^2 + (y^b - y^g)^2)/d_{2,b}^k + d_{2,g}^k)$.

A. Deceptive Evader With Decoupled Cost Structure

We first investigate the scenario where the evader has a decoupled cost structure[3] defined in Definition 5, i.e., $d_{21}^k(\theta_2) = 0, \forall \theta_2 \in \Theta_2, \forall k \in \mathcal{K}$. According to Corollary 1, the evader's trajectory is then independent of the pursuer's action, type, and belief. Fig. 2 visualizes the pursuer's trajectories. Although the pursuer only aims to be close to the evader at the final stage, she also takes proactive actions in the previous stages to be cost-efficient. If the pursuer knows the evader's type, then she can head toward the true target directly and will not be misled by the evader's trajectory ambiguity at the early stages as illustrated by the black dashed line in Fig. 2. If the evader's type is private, then a larger initial belief mismatch $1 - l_1^0(\theta_2^b | x^0, \theta_1^H)$ makes the pursuer head more toward the misleading target at the early stages as illustrated by the three solid lines in Fig. 2. However, due to the pursuer's online learning, which is compatible, efficient, and robust as shown in Section IV-A1, she manages to approach the evader at the final stage regardless of her initial belief mismatch. Fig. 3 shows the pursuer's K-stage belief variation. The evader's ambiguous trajectory results in belief fluctuations at the early stages, yet the pursuer can quickly reduce the belief mismatch when the evader starts to head toward the true target. After the pursuer has corrected her initial belief mismatch at

[3]This article has supplementary downloadable materials available at http://ieeexplore.ieee.org, provided by the authors. This includes a video demo of two UAVs' trajectories and belief updates under the decoupled structure.

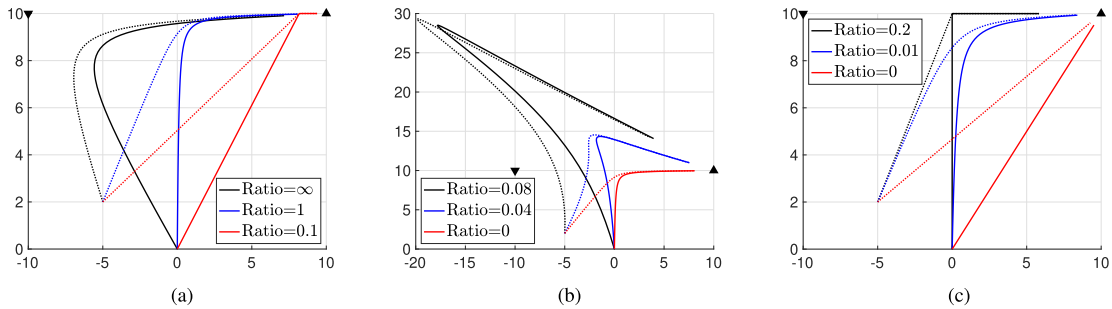

Fig. 1. Evader's trajectories from $x_2^0 = [0, 0]$ and $x_2^0 = [-5, 2]$ in solid and the dashed lines, respectively. The black downward and upward triangles represent the location of Calais $(x^b, y^b) = (-10, 10)$ and Normandy $(x^g, y^g) = (10, 10)$, respectively. The ratios capture the evader's tradeoff of forming a deceptive trajectory, reaching the true target, and evading the pursuit. (a) Ratio represents $d_{2,g}^k(\theta_2^b)/d_{2,b}^k(\theta_2^b)$. (b) Ratio represents $d_{2,b}^k(\theta_2^b)/d_{2,b}^K(\theta_2^b)$. (c) Ratio represents $d_{21}^k(\theta_2^b)/d_{2,b}^K(\theta_2^b)$.

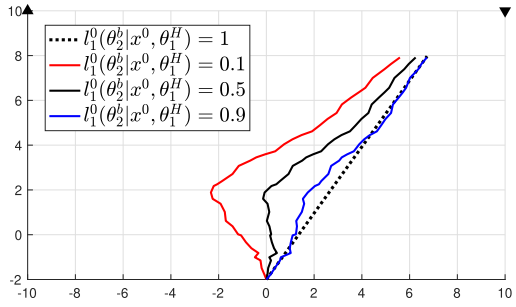

Fig. 2. Pursuer's trajectories under different initial beliefs.

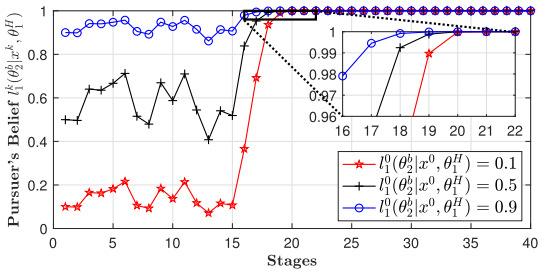

Fig. 3. Pursuer's belief update over K stages under three different initial beliefs and the same noise sequence $[w^k]_{k\in\mathscr{K}}$. The inset black box magnifies the selected area.

around stage $k = 16$, she can head toward the true target in the cost-efficient way; i.e., she attempts to keep a uniform linear motion under the external noise as shown in the upper right region of Fig. 2.

1) Finite-Horizon Analysis of Bayesian Update: In this section, we illustrate the compatibility, efficiency, and robustness of the finite-horizon Bayesian update in (3) to reduce the initial belief mismatch. The pursuer is of high-maneuverability and the evader's true type is θ_2^b. Define the likelihood function of θ_2^b and θ_2^g as $a^k := \Pr(x^{k+1}|\theta_2^b, x^k, \theta_1^H)$ and $c^k := \Pr(x^{k+1}|\theta_2^g, x^k, \theta_1^H)$, respectively. As $w^k \in \mathbb{R}^{n\times 1}$, a^k and c^k are positive. With an initial belief $l_1^0 \in (0, 1)$ and a finite likelihood ratio $e^k := c^k/a^k \in (0, \infty)$, we can represent (3) in the following form with three properties:

$$l_1^{k+1} = \frac{l_1^k \cdot a^k}{l_1^k \cdot a^k + (1 - l_1^k) \cdot c^k} = \frac{1}{1 + \left(\frac{1}{l_1^0} - 1\right) \prod_{\bar{k}=0}^{k} e^{\bar{k}}} \in (0, 1).$$

1) *Compatibility:* For all $l_1^k \in (0, 1)$, the belief update at stage k is compatible to the evidence represented by the ratio e^k. In particular, if $e^k < 1$, then $l_1^{k+1} > l_1^k$; if $e^k > 1$, then $l_1^{k+1} < l_1^k$; if $e^k = 1$, then $l_1^{k+1} = l_1^k$.
2) *Efficiency:* If the evidence of state observation x^{k+1} indicates that the type is more likely to be the true type θ_2^b, i.e., $e^k < 1$, then the function $l_1^{k+1}/l_1^k = 1/(l_1^k + (1 - l_1^k)e^k)$ at stage k is monotonically decreasing over l_1^k. If the evidence indicates that the type is more likely to be the misleading type θ_2^g, i.e., $e^k > 1$, then the function l_1^{k+1}/l_1^k is monotonically increasing over l_1^k.
3) *Robustness:* The order of the evidence sequence $e^{\bar{k}}$, $\bar{k} = 0, \ldots, k$, has no impact on the belief l_1^{k+1}.

Property one shows that although the external noise can result in the fluctuations of the belief update, the belief mismatch, i.e., $1 - l_1^k$, will decrease when $e^k < 1$, regardless of the prior belief $l_1^k \in (0, 1)$. Property two shows the efficiency of the belief update. The belief changes more under a larger belief mismatch, which results in a quick correction. Property three shows the robustness of the belief update. The erroneous belief update caused by a heavy noise can be corrected in the later stages when the noise fades.

2) Comparison With Heuristic Policies: We compare the proposed pursuer's control policy with two heuristic ones to demonstrate its efficacy in counter-deception.[4] The first heuristic policy is to repeat the attacker's trajectory with a one-stage delay; i.e., the pursuer applies the action so that $x_1^{k+1} = x_2^k, \forall k \in \mathscr{K} \setminus \{K\}$. The pursuer does not need to apply Bayesian learning and we name this policy as direct following. The second heuristic policy for the pursuer is to stay at the initial location until her truth-revealing stage k_1^{tr} and then head toward the evader's expected final-stage location in the remaining stages. The second policy is conservative because the pursuer does not take proactive actions until she identifies the evader's type.

Let player i's ex-post cumulative cost $\hat{V}_i^{0:k} := \sum_{h=0}^{k} g_i^h, \forall k \in \mathscr{K}$, be a real-time evaluation of the online algorithm. Although a pursuer under both heuristic policies manages to stay close to the evader at the final stage, Fig. 4 shows that both heuristic policies are more costly than the proposed equilibrium strategy in the long run. The conservative policy avoids potential trajectory deviations under deception but results in less planning stages for the pursuer to achieve the capture goal. We visualize

[4]The supplementary materials include a video demo that compares the proposed policy's trajectory and performance with two heuristic policies.

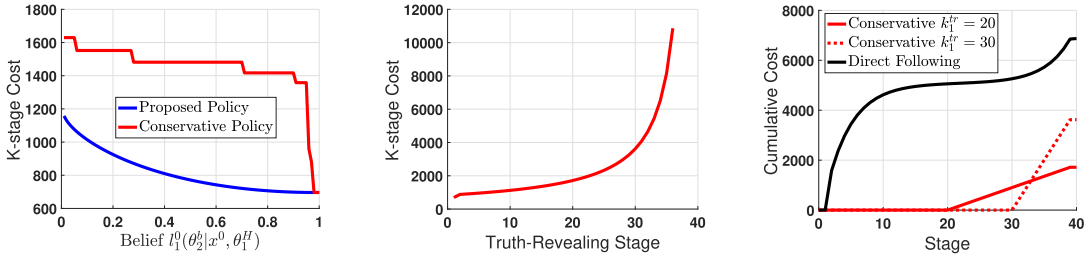

Fig. 4. Pursuer's ex-post cumulative cost under two heuristic policies and the proposed policy. (a) K-stage cumulative cost $\hat{V}_i^{0:K}$ versus different initial beliefs. (b) K-stage cumulative cost $\hat{V}_i^{0:K}$ versus k_1^{tr} under the *conservative policy*. (c) Accumulation of the pursuer's cost $\hat{V}_i^{0:k}, \forall k \in \mathscr{K}$, along with stages.

the accumulation of the pursuer's cost in Fig. 4(c). The red lines show that the pursuer who adopts the conservative policy spends no action costs before the truth-revealing stage k_1^{tr}, i.e., $(u_1^k)' F_{11}^k(\theta_1) u_1^k = 0, \forall k \leq k_1^{tr}$, but huge costs in the remaining stages to fulfill her capture goal. The total cumulative cost $\hat{V}_i^{0:K}$ at the final stage increases exponentially with the value of k_1^{tr} as shown in Fig. 4(b). The black line in Fig. 4(c) illustrates the accumulation of $\hat{V}_i^{0:k}$ when the pursuer direct follows the evader's trajectory. Only under extreme deception scenarios where $k_1^{tr} > 34$, the direct following policy results in a lower cost than the conservative policy does. Since the initial belief l_1^0 affects both the truth-revealing stage and the proposed policy, we plot $\hat{V}_i^{0:K}$ versus l_1^0 under the conservative policy and the proposed policy in Fig. 4(a). When there is no belief mismatch $l_1^0(\theta_2^b|x^0,\theta_1^H) = 1$, we have $k_1^{tr} = 1$ and the conservative policy is equivalent to the proposed policy. As the belief mismatch increases, the cost $\hat{V}_i^{0:K}$ under the proposed policy (resp. the conservative policy) increases due to the larger deviation along the x-axis (resp. the larger k_1^{tr}). The proposed policy always results in a lower cost $\hat{V}_i^{0:K}$ than the conservative policy does. The results in Fig. 4 lead to the following two principles for the pursuer to behave under deception. First, Bayesian learning is a more effective countermeasure than the direct following of the evader's deceptive trajectory. Second, if learning the evader's type takes a long time, the pursuer is better to act proactively based on her current belief than to delay actions until the truth-revealing stage.

B. Dynamic Game for Deception and Counter-Deception

In this section, the evader has a coupled cost[5] defined in Definition 5 and the level of evasion determination increases with a constant rate $\alpha > 0$; i.e., $d_{21}^k(\theta_2) = \alpha k, \forall \theta_2 \in \Theta_2$, $\forall k \in \mathscr{K}$. The evader deceives the pursuer by hiding his true target. The pursuer can adopt the following two countermeasures to reduce her cost under the evader's deception. Section IV-B1 investigates the effectiveness of adaptive learning. We find that the pursuer manages to approach the true target at the final stage by updating her belief and taking actions accordingly based on the real-time trajectory observation. Section IV-B2 further allows the pursuer to introduce additional deception, i.e., obfuscate her maneuverability, to counteract the evader's information advantage and his deception impact.

[5]A video demo of two UAVs' real-time trajectories and belief updates under the coupled structure is included in the supplementary materials.

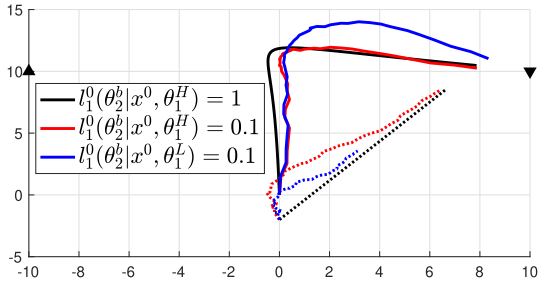

Fig. 5. K-stage trajectory of the evader and the pursuer in solid and dashed lines, respectively. If the evader's type is common knowledge and the pursuer is of high-maneuverability, we represent their noise-free trajectories in black. If the evader's type is private and the pursuer's initial belief mismatch is 0.9, two UAVs' trajectories are in red (resp. blue) when the pursuer's maneuverability is high (resp. low).

1) Pursuer With a Public Type: When the pursuer's type is common knowledge, we plot both UAVs' trajectories under two initial beliefs and two types of pursuers in Fig. 5. The solid lines show that the evader with the coupled cost detours to stay further from the pursuer. The initial belief mismatch causes a deviation along the x-axis for both high- and low-maneuverability pursuers as shown in red and blue, respectively. However, the deviation has a smaller magnitude and lasts shorter than the one represented by the red line in Fig. 2 due to the coupled cost structure of the evader. The pursuer with high maneuverability stays closer to the evader at the final stage.

2) Deception to Counteract Deception: When the pursuer's type is also private, Fig. 6 shows that she can manipulate the evader's initial belief l_2^0 to obtain a smaller k_1^{tr} and a belief update with less fluctuation. The red line with stars is the same as the one in Fig. 3. It shows that the pursuer's belief learning is slower and fluctuates more when she interacts with the evader who has a decoupled cost. The reason is that her manipulation of the initial belief l_2^0 does not affect the evader's decision making as shown in Corollary 1. A comparison between Fig. 6(a) and (b) shows that it is beneficial for a low-maneuverability pursuer to disguise as a high-maneuverability pursuer but not vice versa. Thus, introducing additional deception to counteract existing deception is not always effective.

C. Multi-Dimensional Deception Metrics

The impact of the evader's deception can be measured by metrics such as the endpoint distance $x_2^{fd} := ||x_2^K - \gamma(\theta_2)||_2$ between the evader and the true target, the endpoint

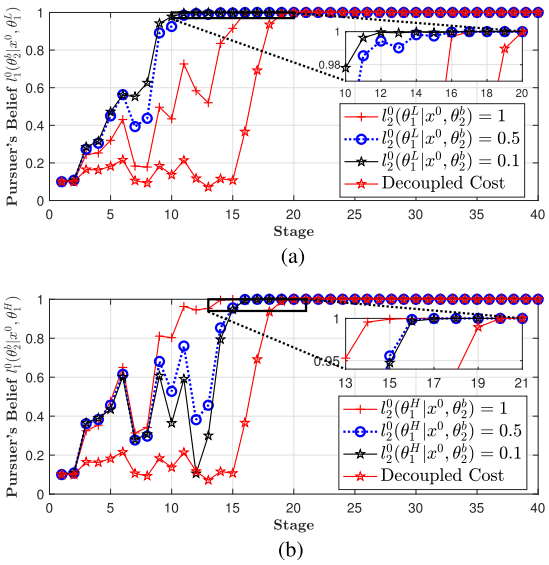

Fig. 6. Pursuer's belief update over K stages with the same initial belief $l_1^0(\theta_2^b|x^0,\theta_1)=0.1$. The inset black box magnifies the selected area. (a) Low-maneuverability pursuer's belief update. (b) High-maneuverability pursuer's belief update.

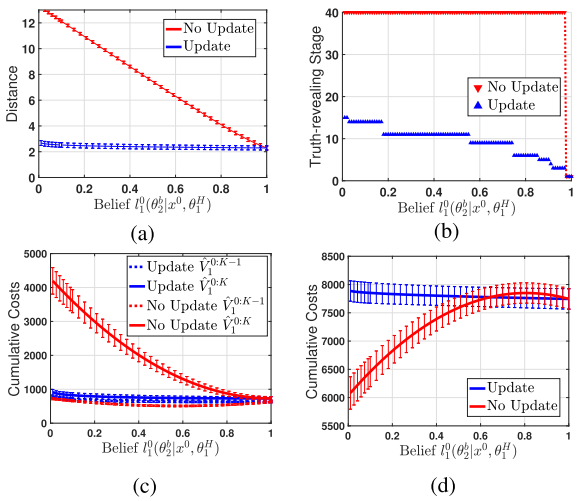

Fig. 7. Influence of the initial belief mismatch on deception. Error bars represent variances of the random variables. (a) Distance x_1^{fd} with its variance magnified by 100 times. (b) Realization of the pursuer's truth-revealing stage k_1^{tr}. (c) Costs $\hat{V}_1^{0:K-1}$ and $\hat{V}_1^{0:K}$ of the pursuer under type θ_1^H. (d) Evader's K-stage ex-post cumulative cost $\hat{V}_2^{0:K}$.

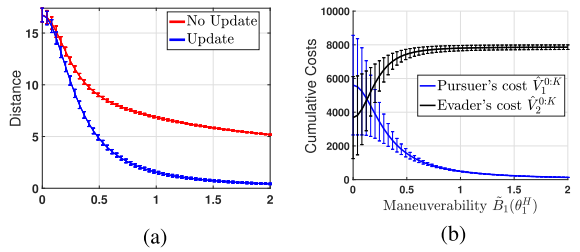

Fig. 8. Influence of the pursuer's maneuverability on deception. Error bars represent variances of the random variables.(a) Distance x_1^{fd} with its variance magnified by 100 times. (b) Two UAVs' K-stage costs $\hat{V}_1^{0:K}$ and $\hat{V}_2^{0:K}$.

distance $x_1^{fd} := ||x_2^K - x_1^K||_2$ between two UAVs, both UAVs' truth-revealing stages k_i^{tr}, and their ex-post cumulative costs $\hat{V}_i^{0:k}, \forall k \in \mathscr{K}$. In this pursuit-evasion case study, we define ϵ-reachability and ϵ-capturability in Definition 8. Although $x_i^{fd}, \forall i \in \{1,2\}$, is a random variable, we can obtain a good estimate of the reachability and capturability due to the negligible variance of x_i^{fd} as shown in Figs. 7(a) and 8(a).

Definition 8 (Reachability and Capturability): Consider the proposed pursuit-evasion scenario with a given $\epsilon \geq 0$, a threshold $\bar{x}^{fd} \geq 0$, and all initial beliefs $l_i^0 \in (0,1)$. The target is said to be ϵ-reachable if $\Pr(x_2^{fd} \geq \bar{x}^{fd}) \leq \epsilon$. The evader is said to be ϵ-capturable if $\Pr(x_1^{fd} \geq \bar{x}^{fd}) \leq \epsilon$.

In Section IV-C1, we investigate how the evader can manipulate the pursuer's initial belief $l_1^0(\theta_2^b|x^0,\theta_1^H)$ to influence the deception. In Section IV-C2, we investigate how the pursuer's maneuverability plays a role in the deception. In both sections, the evader has a coupled cost structure. The pursuer either applies the Bayesian update or not, which is denoted by blue and red lines, respectively, in both Figs. 7 and 8. In Section IV-C3, we study other metrics, such as deceivability, distinguishability, and PoD.

1) Impact of the Evader's Belief Manipulation: Both UAVs determine their initial beliefs based on the intelligence collected before their interactions. By falsifying the pursuer's intelligence, the evader can manipulate the pursuer's initial belief l_1^0 and further influence the deception as shown in Fig. 7. In the x-axis, an initial belief $l_1^0(\theta_2^b|x^0,\theta_1^H)$ closer to 1 indicates a smaller belief mismatch. Fig. 7(a) shows that the pursuer's distance to the evader at the final stage decreases as the belief mismatch decreases regardless of the existence of Bayesian learning. However, the initial belief manipulation has a much less influence on the endpoint distance x_1^{fd} when Bayesian learning is applied. Fig. 7(b) shows that for each realization of the noise sequence w^k, the pursuer's truth-revealing stage steps down as the belief mismatch decreases when Bayesian update is applied. Fig. 7(c) illustrates the pursuer's ex-post cumulative cost $\hat{V}_1^{0:K}$ and $\hat{V}_1^{0:K-1}$ at the last and the second last stage, respectively. Without Bayesian update, the evader's deception significantly increases the pursuer's cost at the second last stage due to the large endpoint distance x_1^{fd}. The red lines show that the cost increase is higher under a larger belief mismatch. Fig. 7(d) illustrates the evader's ex-post cumulative cost at the last stage. If the pursuer does not apply Bayesian learning, then the evader can decrease his cost by increasing the pursuer's belief mismatch. If the pursuer applies Bayesian learning, then the evader's cost increases slightly if the pursuer's belief mismatch is increased. When the belief mismatch is small (i.e., $1 - l_1^0 \in (0, 0.35)$), we observe a win-win situation; i.e., Bayesian learning not only reduces the pursuer's ex-post cumulative cost, but also the evader's.

2) Impact of the Pursuer's Maneuverability: The pursuer's maneuverability can also affect deception as shown in Fig. 8. The pursuer has an initial belief $l_1^0(\theta_2^b|x^0,\theta_1^H) = 0.5$ and the evader knows the pursuer's type. Fig. 8(a) illustrates that the pursuer can exponentially decrease her distance to the evader at the final stage as her maneuverability increases. Fig. 8(b) demonstrates that the maneuverability increase can decrease and increase the pursuer's and the evader's ex-post cumulative costs at the final stage, respectively. The variance

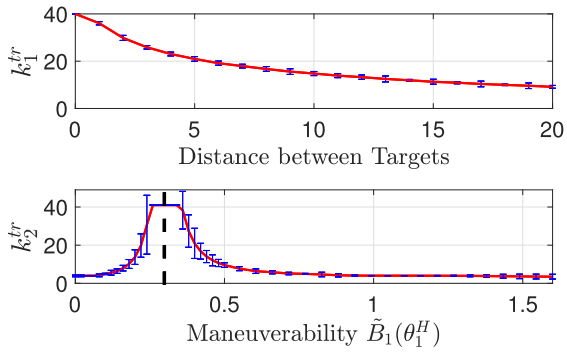

Fig. 9. Plot of the deceived robot's truth-revealing stage versus the deceiver's type distinguishability. Error bars represent their variances, which are magnified by five times.

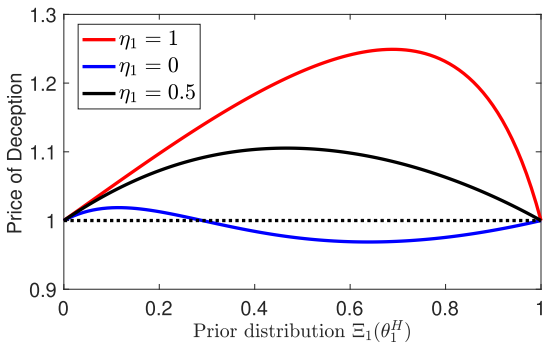

Fig. 10. PoD versus prior type distribution for three values of η_1.

grows as maneuverability decreases because the pursuer's trajectory will become largely affected by the external noise. In both figures, we observe the phenomenon of the marginal effect; i.e., the change rates of both the endpoint distance x_1^{fd} and the cost $\hat{V}_i^{0:K}$ decrease as the maneuverability increases. Thus, we conclude that higher maneuverability can improve the pursuer's performance under the evader's deception as measured by the distance x_1^{fd} and the cost $\hat{V}_1^{0:K}$. Moreover, the improvement rate is higher with low maneuverability.

3) Deceivability, Distinguishability, and PoD: Deceivability defined in Definition 7 is highly related to the distinguishability among different types. In this case study, a larger distance between targets, i.e., $||\gamma(\theta_2^g) - \gamma(\theta_2^b)||_2$, makes it easier for the pursuer to distinguish between evaders of type θ_2^b and type θ_2^g. A larger maneuverability difference $|\tilde{B}_1(\theta_1^H) - \tilde{B}_1(\theta_1^L)|$ makes it easier for the evader to distinguish between pursuers of type θ_1^H and type θ_1^L. We visualize two UAVs' truth-revealing stages k_i^{tr} versus the distance between targets and the maneuverability difference in Fig. 9. The evader has a coupled cost and both players' initial belief mismatches are 0.5. The black dashed line indicates $\tilde{B}_1(\theta_1^L) = 0.3$. When the maneuverability difference is negligible $\tilde{B}_1(\theta_1^H) \in (0.26, 0.36)$, the pursuer's type cannot be learned correctly in K stages; i.e., the pursuer is $(K+1)$-stage 0-deceivable. When the maneuverability difference is small, i.e., $\tilde{B}_1(\theta_1^H) \in (0.1, 0.5)$, yet not negligible, i.e., $\tilde{B}_1(\theta_1^H) \notin (0.26, 0.36)$, the variance of k_2^{tr} is large.

Let $\theta_2 = \theta_2^b$ be common knowledge and assume that the evader's belief confirms to the prior distribution of the pursuer's type for all stages, i.e., $l_2^k(\theta_1|h^k, \theta^b) = \Xi_1(\theta_1)$, $\forall \theta_1 \in \Theta_1, \forall k \in \mathscr{K}$. Then, Fig. 10 illustrates how the prior distribution of the pursuer's type affects the value of PoD under three scenarios.

1) $\eta_1 = 1$, i.e., the central planner only evaluates UAV 1's performance under deception.
2) $\eta_1 = 0$, i.e., the central planner only evaluates UAV 2's performance under deception.
3) $\eta_1 = 0.5$, i.e., the central planner evaluates the average performance of two UAVs under deception.

When the pursuer's type is also common knowledge, i.e., $\Xi_1(\theta_1^H) = 0$ (i.e., the pursuer has type θ_1^L) and $\Xi_1(\theta_1^H) = 1$ (i.e., the pursuer has type θ_1^H), the game is of complete information and the value of PoD equals 1. Since PoD takes continuous values over $\Xi_1(\theta_1^H) \in [0, 1]$ and has a value of 1 at two endpoints for all feasible η_1, we refer to the plots in Fig. 10 as jump rope plots. They corroborate that the PoD can be bigger than 1; i.e., deception among players may not only benefit the deceiver but also the deceivee.

V. Conclusion and Future Work

We have investigated a novel class of rational robot deception problems where intelligent robots hide their heterogeneous private information to achieve their objectives in finite stages with minimum costs. We have proposed an N-player dynamic game framework to quantify the impact of deception and design long-term optimal actions for deception and counter-deception. Robots form their own initial beliefs on others' private information and update their beliefs at each stage based on extrinsic or intrinsic information. Satisfying the properties of sequential rationality and belief consistency, PBNE can be used to predict N robots' actions and costs over the K stages. We have studied a class of games in the LQ form with extrinsic belief dynamics to obtain a unique affine state-feedback control policy and a set of extended Riccati equations. The cognitive coupling resulted from the deception of types demonstrates a distinct feature of rational deception where each player's action hinges on not only his own belief but also all other players' beliefs. The concepts of deceivability, distinguishability, and reachability have been defined to characterize the fundamental limits of deception. Meanwhile, the PoD serves as a crucial evaluation and design metric.

We have investigated a target protection problem where the evader aims to deceptively reach the true target and the pursuer keeps her maneuverability as private information. The pursuer achieves a lower ex-post cumulative cost under the proposed policy than under the direct-following and conservative policies. We have proposed multi-dimensional metrics such as the stage of truth revelation and the endpoint distance to measure the deception impact throughout stages. We have concluded that Bayesian learning can largely reduce the impact of initial belief manipulation and sometimes result in a win-win situation. The increase of the pursuer's maneuverability can also reduce the endpoint distance and her ex-post cumulative cost yet has a marginal effect. A robot is more deceivable, i.e., less learnable when its potential type is less distinguishable. Finally, we have found that introducing additional deception to counteract existing deception is not always effective. Moreover, deception among multiple players may not only benefit the deceiver but also the deceivee.

REFERENCES

[1] D. L. Smith, *Why We Lie: The Evolutionary Roots of Deception and the Unconscious Mind*. New York, NY, USA: Macmillan, 2004.

[2] M. Howard and M. E. Howard, *Strategic Deception in the Second World War*, vol. 5. New York, NY, USA: WW Norton & Company, 1995.

[3] L. Cowen, T. Ideker, B. J. Raphael, and R. Sharan, "Network propagation: A universal amplifier of genetic associations," *Nature Rev. Genet.*, vol. 18, no. 9, p. 551, 2017.

[4] E. Al-Shaer, J. Wei, K. W. Hamlen, and C. Wang, "Dynamic Bayesian games for adversarial and defensive cyber deception," in *Autonomous Cyber Deception*. New York, NY, USA: Springer, 2019, pp. 75–97.

[5] D. Li and J. B. Cruz, "Defending an asset: A linear quadratic game approach," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 2, pp. 1026–1044, Apr. 2011.

[6] K. Sreenath and V. Kumar, "Dynamics, control and planning for cooperative manipulation of payloads suspended by cables from multiple quadrotor robots," in *Robotics: Science and Systems*. Cambridge, MA, USA: MIT Press, 2013.

[7] J. C. Harsanyi, "Games with incomplete information played by 'Bayesian' players, I–III Part I. The basic model," *Manage. Sci.*, vol. 14, no. 3, pp. 159–182, 1967.

[8] V. L. L. Thing and J. Wu, "Autonomous vehicle security: A taxonomy of attacks and defences," in *Proc. IEEE Int. Conf. Internet Things (iThings), IEEE Green Comput. Commun. (GreenCom), IEEE Cyber, Phys. Social Comput. (CPSCom), IEEE Smart Data (SmartData)*, Dec. 2016, pp. 164–170.

[9] Y. Huang, J. Chen, L. Huang, and Q. Zhu, "Dynamic games for secure and resilient control system design," *Nat. Sci. Rev.*, vol. 7, no. 7, pp. 1125–1141, Jul. 2020.

[10] Y. Zhao, L. Huang, C. Smidts, and Q. Zhu, "Finite-horizon semi-Markov game for time-sensitive attack response and probabilistic risk assessment in nuclear power plants," *Rel. Eng. Syst. Saf.*, vol. 201, Sep. 2020, Art. no. 106878.

[11] D. Li, N. Gebraeel, and K. Paynabar, "Detection and differentiation of replay attack and equipment faults in SCADA systems," *IEEE Trans. Autom. Sci. Eng.*, early access, Aug. 25, 2020, doi: 10.1109/TASE.2020.3013760.

[12] S. Bhattacharya and T. Başar, "Game-theoretic analysis of an aerial jamming attack on a UAV communication network," in *Proc. Amer. Control Conf.*, Jun. 2010, pp. 818–823.

[13] R. Zhang and P. Venkitasubramaniam, "Stealthy control signal attacks in linear quadratic Gaussian control systems: Detectability reward tradeoff," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 7, pp. 1555–1570, Jul. 2017.

[14] Q. Zhang, K. Liu, Y. Xia, and A. Ma, "Optimal stealthy deception attack against cyber-physical systems," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3963–3972, Sep. 2020.

[15] A. Ayub, "An adaptive Markov process for robot deception," M.S. thesis, Dept. Elect. Eng., Pennsylvania State Univ., State College, PA, USA, Jul. 2017.

[16] L. Huang and Q. Zhu, "A dynamic games approach to proactive defense strategies against advanced persistent threats in cyber-physical systems," *Comput. Secur.*, vol. 89, Feb. 2020, Art. no. 101660.

[17] M. O. Karabag, M. Ornik, and U. Topcu, "Optimal deceptive and reference policies for supervisory control," in *Proc. IEEE 58th Conf. Decis. Control (CDC)*, Dec. 2019, pp. 1323–1330.

[18] M. Ornik and U. Topcu, "Deception in optimal control," in *Proc. 56th Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2018, pp. 821–828.

[19] K. Horák, Q. Zhu, and B. Bošanský, "Manipulating adversary's belief: A dynamic game approach to deception by design for proactive network security," in *Decision and Game Theory for Security*, S. Rass, B. An, C. Kiekintveld, F. Fang, and S. Schauer, Eds. New York, NY, USA: Springer, 2017, pp. 273–294.

[20] B. R. Brewer, R. L. Klatzky, and Y. Matsuoka, "Visual-feedback distortion in a robotic rehabilitation environment," *Proc. IEEE*, vol. 94, no. 9, pp. 1739–1751, Sep. 2006.

[21] J. Shim and R. C. Arkin, "Other-oriented robot deception: A computational approach for deceptive action generation to benefit the mark," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2014, pp. 528–535.

[22] A. Dragan, R. Holladay, and S. Srinivasa, "An analysis of deceptive robot motion," in *Robotics: Science and Systems*. Cambridge, MA, USA: MIT Press, Jul. 2014.

[23] J. Shim and R. C. Arkin, "A taxonomy of robot deception and its benefits in HRI," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2013, pp. 2328–2335.

[24] A. R. Wagner and R. C. Arkin, "Acting deceptively: Providing robots with the capacity for deception," *Int. J. Social Robot.*, vol. 3, no. 1, pp. 5–26, Jan. 2011.

[25] P. Masters, "Goal recognition and deception in path-planning," Ph.D. dissertation, School Sci., College Sci., Eng. Health, RMIT Univ., Melbourne VIC, USA, Feb. 2019.

[26] K. Xu, Y. Zeng, L. Qin, and Q. Yin, "Single real goal, magnitude-based deceptive path-planning," *Entropy*, vol. 22, no. 1, p. 88, Jan. 2020.

[27] W. McEneaney and R. Singh, "Deception in autonomous vehicle decision making in an adversarial environment," in *Proc. AIAA Guid., Navigat., Control Conf. Exhib.*, Aug. 2005, p. 6152.

[28] P. G. Bennett, "Hypergames: Developing a model of conflict," *Futures*, vol. 12, no. 6, pp. 489–507, Dec. 1980.

[29] X. He and H. Dai, *Dynamic Security Games With Deception*. Cham, Switzerland: Springer, 2018, pp. 61–71.

[30] J. Pawlick, E. Colbert, and Q. Zhu, "Modeling and analysis of leaky deception using signaling games with evidence," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 7, pp. 1871–1886, Jul. 2019.

[31] L. Huang and Q. Zhu, "Duplicity games for deception design with an application to insider threat mitigation," 2020, *arXiv:2006.07942*. [Online]. Available: http://arxiv.org/abs/2006.07942

[32] N. R. Sandell and M. Athans, "Solution of some nonclassical LQG stochastic decision problems," *IEEE Trans. Autom. Control*, vol. AC-19, no. 2, pp. 108–116, Apr. 1974.

[33] L. Huang and Q. Zhu, "Analysis and computation of adaptive defense strategies against advanced persistent threats for cyber-physical systems," in *Proc. Int. Conf. Decis. Game Theory Secur.*, 2018, pp. 205–226.

[34] L. Huang and Q. Zhu, "Adaptive strategic cyber defense for advanced persistent threats in critical infrastructure networks," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 46, no. 2, pp. 52–56, 2019.

[35] A. Ouammi, Y. Achour, D. Zejli, and H. Dagdougui, "Supervisory model predictive control for optimal energy management of networked smart greenhouses integrated microgrid," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 1, pp. 117–128, Jan. 2020.

[36] J. B. Cruz, M. A. Simaan, A. Gacic, and Y. Liu, "Moving horizon Nash strategies for a military air operation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 989–999, Jul. 2002.

[37] A. Liniger and J. Lygeros, "A noncooperative game approach to autonomous racing," *IEEE Trans. Control Syst. Technol.*, vol. 28, no. 3, pp. 884–897, May 2020.

[38] H. Hajieghrary, D. Kularatne, and M. A. Hsieh, "Cooperative transport of a buoyant load: A differential geometric approach," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 2158–2163.

[39] R. S. Nickerson, "Confirmation bias: A ubiquitous phenomenon in many guises," *Rev. Gen. Psychol.*, vol. 2, no. 2, pp. 175–220, 1998.

[40] D. Fridovich-Keil, V. Rubies-Royo, and C. J. Tomlin, "An iterative quadratic method for general-sum differential games with feedback linearizable dynamics," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 2216–2222.

[41] T. Basar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, vol. 23. Philadelphia, PA, USA: SIAM, 1999.

Linan Huang (Student Member, IEEE) received the B.Eng. degree (Hons.) in electrical engineering from the Beijing Institute of Technology, Beijing, China, in 2016. He is currently pursuing the Ph.D. degree with the Laboratory for Agile and Resilient Complex Systems, Tandon School of Engineering, New York University, New York, NY, USA.

His research interests include dynamic decision making of multi-agent systems, mechanism design, artificial intelligence, security, and resilience for cyber-physical systems.

Quanyan Zhu (Member, IEEE) received the B.Eng. degree (Hons.) in electrical engineering from McGill University, Montreal, QC, Canada, in 2006, the M.A.Sc. degree from the University of Toronto, Toronto, ON, Canada, in 2008, and the Ph.D. degree from the University of Illinois at Urbana–Champaign (UIUC), Champaign, IL, USA, in 2013.

After stints at Princeton University, Princeton, NJ, USA, he is currently an Associate Professor with the Department of Electrical and Computer Engineering, New York University (NYU), New York, NY, USA. He is an Affiliated Faculty Member with the Center for Urban Science and Progress (CUSP) and the Center for Cyber Security (CCS), NYU. His current research interests include game theory, machine learning, cyber deception, and cyber-physical systems.