2026 American Control Conference (ACC)
May 27-29, 2026. New Orleans, Louisiana, USA

A Tutorial for Foundations in AI and Quantum Systems: From Control and Game-Theory to Stochastic Deception

Kyriakos G. Vamvoudakis[1], Jorge I. Poveda[2], Sean Anderson[3], João P. Hespanha[3], Miroslav Krstic[4], Ian R. Petersen[5], Michael Tang[2], Ali Wadi[1], Quanyan Zhu[5]

***Abstract*— In this tutorial, we explore the intersection of quantum control, game theory, and artificial intelligence, emphasizing their combined potential to enhance secure and strategic decision-making in complex systems. It examines how deceptive behavior can be understood and influenced within strategic frameworks, and outlines principles for designing resilient systems that leverage advantages in both quantum and learning-driven environments. While the discussion centers on specific examples, the foundational ideas are broadly applicable across a range of domains.**

I. Introduction

The synergistic convergence of computer science, information theory, and quantum mechanics has established the interdisciplinary field of quantum computation and quantum information. This domain is characterized by a paradigm shift in which quantum systems are no longer regarded as passive phenomena for observation but as actively engineered systems for advanced information processing. Within this context, the principles of Artificial Intelligence (AI) and Machine Learning (ML) hold critical importance, as their core learning algorithms are fundamentally dependent on solving constrained multivariate optimization problems. Quantum computing offers novel and potentially superior frameworks for these tasks.

The practical realization of quantum advantages relies on control, which involves developing sophisticated experimental techniques for precise manipulation of individual quantum systems—an essential step for robustly encoding quantum information, accurately executing unitary transformations, and reliably preparing initial states—while game theory offers a formal lens to analyze the strategic capabilities and limitations of quantum information processing, facilitating exploration of how quantum resources can achieve computational outcomes unattainable by classical means and challenging established computational models like the strong Church–Turing thesis. Quantum information exhibits a dual nature—its foundational principles, like the no-cloning theorem, support quantum cryptography's security, yet practical systems are susceptible to deception through quantum noise (decoherence), which corrupts quantum states via environmental interactions, necessitating robust quantum error-correction and fault-tolerant architectures to maintain reliable quantum information processing over extended durations.

This tutorial delves into the convergence of game theory, AI deception, and quantum control, exploring their pivotal roles in secure decision-making. It examines adaptive AI systems where deception emerges due to incentive misalignment, and how equilibrium and control principles facilitate regulation. Information asymmetry is identified as a crucial factor in deception, providing a foundation for resilient architecture design in quantum and AI contexts. Furthermore, the tutorial emphasizes the significance of control in quantum information processing and explores strategic frameworks that transcend classical paradigms, highlighting the potential of quantum games.

II. Q-Learning for Zero-Sum Turn Games

This section addresses Markov zero-sum games "turn" games, where at each state of the game, only one player can make decisions, while the other player waits for a turn. Many board-games (like chess, Connect Four, or Dots and Boxes) are "turn" games with alternating moves, but we consider here a general form of turn games where players do not necessarily alternate. The motivation for this setup includes strategy games and pursuit-evasion problems where the evader loses action when cornered, or in network security where an attacker may control certain states while a defender can only respond at select times.

As opposed to general zero-sum Markov games (see, e.g., [1]), turn games with finite state and action spaces have pure saddle-point policies under very mild assumptions [2]. In fact, Q-learning can, in principle, be used to construct pure saddle-point state-feedback policies for turn games. For complex games, Q-learning must be terminated before exploring the full state, which can lead to policies that cannot guarantee the final Q-function security levels. This can be mitigated through "opponent-informed" exploration for Q-learning sample selection. This form of exploration can

[1]K. G. Vamvoudakis and A. Wadi are with the Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, 30332, USA. E-mail: kyriakos@gatech.edu, awadi@gatech.edu

[2]J. Poveda and M. Tang are with the Electrical and Computer Engineering Department of the Jacobs School of Engineering, University of California, San Diego, La Jolla, CA, USA. Email:{poveda,myt001@ucsd.edu}.

[3]J. P. Hespanha and S. Anderson are with the Electrical and Computer Engineering Department, University of California, Santa Barbara, CA 93106 USA. E-mail: seananderson@ucsb.edu, hespanha@ece.ucsb.edu.

[4]M. Krstic is with the Mechanical and Aerospace Engineering Department of the Jacobs School of Engineering, University of California, San Diego, La Jolla, CA, USA. Email:krstic@ucsd.edu.

[5]I. R. Petersen is with the School of Engineering, The Australian National University, Canberra ACT 2601, Australia. Email: i.r.petersen@gmail.com

[6]Q. Zhu is with Department of Electrical and Computer Engineering, New York University, New York 10012, USA. Email: qz494@nyu.edu

This work was supported in part by ARO under grant No. W911NF-24 − 1 − 0174, by NSF under grant Nos. 2305756 2227185, 2229876, 2231651, 2415479, by NASA ULI under grant No. 80NSSC25M7104, by ONR under grant no. N00014-24-1-2492, and by AFOSR under grant FA9550 − 22 − 1 − 0211.

guarantee that the final Q-function provides security levels that hold, at least, against a given set of policies.

Brief summary of related work: Single-player Q-learning was originally proposed for single player problems [3], with convergence results appearing shortly after [4], [5]. Markov games date back to [6], but we refer the reader to the monograph [1] for a systematic treatment of zero-sum Markov games. Q-learning was extended to Markov games in [7], with convergence results in [8]. Turn-based games are less common and have appeared in [9]–[11]. The idea of using externally provided (sets of) policies for training purposes has appeared in Policy Space Response Oracles (PSRO) [12] and fictitious self-play [13]. Many of the results summarized here appeared for the first time in [11].

A. Zero-Sum Turn Games

The state-space $\mathcal{S}$ of *turn games* can be partitioned into two disjoint sets $\mathcal{S}_1 \cap \mathcal{S}_2 = \varnothing$ with the understanding that, when the state s_t at time $t \geqslant 1$ belongs to $\mathcal{S}_1$, the following action $a_t \in \mathcal{A}$ is selected by player P_1, otherwise $s_t \in \mathcal{S}_2$ and the action is selected by player P_2. We consider here *zero-sum turn games* with symmetric rewards for the two players and denote by $r_{t+1} \in \mathcal{R} \subset \mathbb{R}$, $t \geqslant 1$ the immediate reward collected by the player that selected the action a_t at time t. The total reward collected by player P_i, $i \in \{1,2\}$ for the initial state $s_1 \in \mathcal{S}$ is then given by

$$J_i(s_1) \coloneqq \sum_{t=1}^{\infty} \mathrm{E}[r_{t+1} \operatorname{sgn}_i(s_t)] \tag{1}$$

where $\operatorname{sgn}_i(s_t) = 1$ if $s_t \in \mathcal{S}_i$ and $\operatorname{sgn}_i(s_t) = -1$ otherwise. The sets $\mathcal{S}, \mathcal{A}, \mathcal{R}$ are assumed finite and the state s_t is a *stationary controlled Markov chain* in the sense that $P(s_{t+1} = s', r_{t+1} = r \mid s_t = s, a_t = a) = p(s', r \mid s, a), \forall t \geqslant 1, s, s' \in \mathcal{S}, a \in \mathcal{A}, r \in \mathcal{R}$, where $p : \mathcal{S} \times \mathcal{R} \times \mathcal{S} \times \mathcal{A} \to [0,1]$ is the *transition/reward probability function*. The game is said to be *deterministic* if $p(\cdot,\cdot)$ takes values in the set $\{0,1\}$ and that it *terminates in finite time* if there exists a finite time $T \geqslant 1$ such that $r_t = 0$, $\forall t \geqslant T$ with probability one, regardless of $a_t \in \mathcal{A}$ selected.

1) Policies and value function: A (pure) policy for player P_i, $i \in \{1,2\}$ is a map $\pi_i : \mathcal{S}_i \to \mathcal{A}$ that selects the action $a_t = \pi_i(s_t)$ when the state s_t is in $\mathcal{S}_i$. For a pair of policies (π_1, π_2) for players $\mathsf{P}_1, \mathsf{P}_2$, respectively, we define the pair's *value function* as $V_{\pi_1,\pi_2}(s) \coloneqq \operatorname{sgn}_1(s) \sum_{t=\tau}^{\infty} \mathrm{E}_{\pi_1,\pi_2}[r_{t+1} \operatorname{sgn}_1(s_t) \mid s_\tau = s]$, where the subscript in the expected value highlights that the actions are determined by the given policies. The reward (1) collected by player P_i can be obtained from the value function using $J_i(s_1) = \operatorname{sgn}_i(s_1) V_{\pi_1,\pi_2}(s_1), \forall i \in \{1,2\}$.

2) Saddle-points and security policies: A pair of policies $(\pi_1^\star, \pi_2^\star)$, for players $\mathsf{P}_1, \mathsf{P}_2$ respectively, is a *(pure) feedback saddle-point* if for every other policies (π_1, π_2) for players $\mathsf{P}_1, \mathsf{P}_2$, respectively, we have that

$$J_1^\star(s) \coloneqq \operatorname{sgn}_1(s) V_{\pi_1^\star,\pi_2^\star}(s) \geqslant \operatorname{sgn}_1(s) V_{\pi_1,\pi_2^\star}(s), \tag{2a}$$
$$J_2^\star(s) \coloneqq \operatorname{sgn}_2(s) V_{\pi_1^\star,\pi_2^\star}(s) \geqslant \operatorname{sgn}_2(s) V_{\pi_1^\star,\pi_2}(s), \tag{2b}$$

$\forall s \in \mathcal{S}$, and $J_1^\star(s_1) = -J_2^\star(s_1)$ is called the *value of the game*. The inequality in (2a) expresses no regret in the sense that P_1 does not regret its choice of $\pi_1^\star$ (over any other policy π_1) against $\pi_2^\star$ and, similarly, (2b) expresses no regret for P_2. Saddle-point policies for zero-sum games are known to also be *security policies with values $J_1^\star(s_1)$ and $J_2^\star(s_1)$ for players P_1 and P_2, respectively*; in the sense that

$$\begin{aligned} J_1^\star(s_1) &= \max_{\pi_1} \min_{\pi_2} \operatorname{sgn}_1(s_1) V_{\pi_1,\pi_2}(s_1) \\ &= \min_{\pi_2} \operatorname{sgn}_1(s_1) V_{\pi_1^\star,\pi_2}(s_1), \end{aligned} \tag{3a}$$
$$\begin{aligned} J_2^\star(s_1) &= \max_{\pi_2} \min_{\pi_1} \operatorname{sgn}_2(s_1) V_{\pi_1,\pi_2}(s_1) \\ &= \min_{\pi_1} \operatorname{sgn}_2(s_1) V_{\pi_1,\pi_2^\star}(s_1), \end{aligned} \tag{3b}$$

which means that, by using the policy $\pi_i^\star$, the player P_i can expect a reward at least as large as $J_i^\star(s_1)$, *no matter what policy the other player uses* [2].

B. Dynamic Programming and Q-learning

Dynamic programming provides a procedure to compute the value function for saddle-point pairs and security policies. It relies on the observation that, under mild assumptions, this value function must satisfy the Bellman-like condition:

$$V(s) = \max_{a \in \mathcal{A}} \mathrm{E}\big[r_{t+1} + \operatorname{sgn}_1(s_t) \operatorname{sgn}_1(s_{t+1}) V(s_{t+1}) \mid s_t = s, a_t = a\big], \quad \forall s \in \mathcal{S}, t \geqslant 1,$$

which resembles the usual dynamic programming equation for single-agent optimization, but with a new term $\operatorname{sgn}_1(s_t) \operatorname{sgn}_1(s_{t+1})$ that is specific to turn games [11]. Any pair of policies $(\pi_1^\star, \pi_2^\star)$ for which

$$\pi_i^\star(s) \in \arg\max_{a \in \mathcal{A}} Q(s,a), \quad \forall s \in \mathcal{S}_i, i \in \{1,2\} \tag{4}$$

is a feedback saddle-point and both policies are security policies, with values $J_1^\star(s_1) = \operatorname{sgn}_1(s_1) V(s_1) = -J_2^\star(s_1)$, where

$$Q(s,a) \coloneqq \mathrm{E}\big[r_{t+1} + \operatorname{sgn}_1(s_t) \operatorname{sgn}_1(s_{t+1}) V(s_{t+1}) \mid s_t = s, a_t = a\big].$$

The characterization of saddle-point/security policies in (4) motivates searching for the function Q that directly extracts these policies from the system. The result shows that this function can also be characterized by a Bellman-like condition. To state it formally, we need the following definition: we say that a function $V : \mathcal{S} \to \mathbb{R}$ is *absolutely summable* if for every pair of policies (π_1, π_2) for players $\mathsf{P}_1, \mathsf{P}_2$, respectively, the series $\sum_{t=\tau}^{\infty} \mathrm{E}_{\pi_1,\pi_2}[V(s_t) \mid s_\tau = s]$ is absolutely convergent $\forall s \in \mathcal{S}, \tau \geqslant 1$. For games that terminate in finite time T, any function $V : \mathcal{S} \to \mathbb{R}$ for which $V(s_t) = 0$, $\forall t \geqslant T$ with probability one is absolutely summable, since the series degenerates into a finite summation.

Theorem 1 (Saddle-point sufficient condition [11]): Suppose there exists a function $Q : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ for which

$$V(s) \coloneqq \max_{a \in \mathcal{A}} Q(s,a), \quad \forall s \in \mathcal{S}, \tag{5}$$

is absolutely summable and

$$Q(s,a) = \mathrm{E}\left[r_{t+1} + \mathrm{sgn}_1(s_t)\,\mathrm{sgn}_1(s_{t+1}) \max_{a' \in \mathcal{A}} Q(s_{t+1}, a') \,\middle|\, s_t = s, a_t = a\right], \quad \forall s \in \mathcal{S},\ a \in \mathcal{A}. \quad (6)$$

Then any pair of policies $(\pi_1^\star, \pi_2^\star)$ for which (4) holds is a (pure) saddle-point and they are also security policies with $J_1^\star(s_1) = \mathrm{sgn}_1(s_1) \max_{a \in \mathcal{A}} Q(s_1, a) = -J_2^\star(s_1)$. □

Q-learning can be used to find $Q : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ that satisfies (6). For deterministic turn games, *Q-learning* starts from an initial estimate $\mathcal{Q}^0 : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ and iteratively draws samples $(s_t, a_t, s_{t+1}, r_{t+1})$ from the transition/reward probability function $p(s_{t+1}, r_{t+1} \mid s_t, a_t)$, which are used to constract a sequence $\{Q^k\}$ using updates of the form

$$Q^{k+1}(s_t, a_t) = r_{t+1} + \mathrm{sgn}_1(s_t)\,\mathrm{sgn}_1(s_{t+1}) \max_{a' \in \mathcal{A}} Q^k(s_{t+1}, a'). \quad (7)$$

Establishing convergence of this iteration to a function satisfing (6) typically requires the following Assumption:

Assumption 1 (Exploration): Every pair $(s_t, a_t) \in \mathcal{S} \times \mathcal{A}$ appears infinitely often in the sequence of samples $\{(s_t, a_t, s_{t+1}, r_{t+1})\}$ used for the updates in (7) . □

For deterministic games that terminate in finite time, little more is needed to establish convergence:

Theorem 2 (Deterministic finite games [11]): Assume that the Exploration Assumption 1 holds and that the game is deterministic and terminates in finite time. Initializing $Q^0(s,a) = 0, \quad \forall s \in \mathcal{S},\ a \in \mathcal{A}$, the sequence of functions Q^k converges in a finite number of iterations to a function $Q^\star$ for which the sufficient condition (6) holds. □

C. Games with Large State Spaces

When the state spaces $\mathcal{S}$ is very large,it is generally not possible for the Exploration Assumption 1 to hold for every pair (s_t, a_t) in $\mathcal{S} \times \mathcal{A}$. In this case, we may be forced to stop the Q-learning iteration prematurely at some Q^k, before convergence. In this case, policies that satisfy

$$\pi_i^k(s) \in \arg\max_{a \in \mathcal{A}} Q^k(s,a), \quad \forall s \in \mathcal{S}_i, i \in \{1,2\} \quad (8)$$

may *not* be security policies and $\mathrm{sgn}_i(s_1) \max_{a \in \mathcal{A}} Q^k(s_1, a)$ may *not* be a security value for player P_i. This latter observation means that the policies (8) may have weaknesses that cannot be inferred from $Q^k(\cdot)$.

It turns out that it is possible to "protect" the policy (8) against a given set of policies by guaranteeing that Q-learning explores a sufficiently large subset of the state-space $\mathcal{S}$. To formalize this result we need the following definition: given sets of policies $\Pi_1 = \{\pi_1 : \mathcal{S}_1 \to \mathcal{A}\}$ and $\Pi_2 = \{\pi_2 : \mathcal{S}_2 \to \mathcal{A}\}$ for players P_1 and P_2, respectively, we say that a subset $\mathcal{S}^\perp$ of $\mathcal{S}$ is *invariant for* Π_1, Π_2 if for every initial state s_1 in $\mathcal{S}^\perp$, all subsequent states remain in $\mathcal{S}^\perp$ with probability one for all the policies in Π_1, Π_2.

Theorem 3 ([11]): Suppose there exists a function $Q : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ and a set $\mathcal{S}^\perp \subset \mathcal{S}$ for which (5) is absolutely summable and (6) holds $\forall s \in \mathcal{S}^\perp$, $a \in \mathcal{A}$. If $\mathcal{S}^\perp$ is invariant for sets of policies Π_1, Π_2 that include the *exploitation* policies

$$\pi_i^\perp(s) \in \arg\max_{a \in \mathcal{A}} Q(s,a), \quad \forall s \in \mathcal{S}_i, i \in \{1,2\}, \quad (9)$$

then $(\pi_1^\perp, \pi_2^\perp)$ are security policies restricted to the sets of policies Π_1, Π_2, with values

$$\mathrm{sgn}_1(s) V(s) = \min_{\pi_2 \in \Pi_2} \mathrm{sgn}_1(s_1) V_{\pi_1^\perp, \pi_2}(s_1) \quad (10)$$

$$\mathrm{sgn}_2(s) V(s) = \min_{\pi_1 \in \Pi_1} \mathrm{sgn}_2(s_1) V_{\pi_1^\perp, \pi_2}(s_1), \quad (11)$$

for the players P_1 and P_2, respectively. □

1) Opponent-informed exploration: Motivated by Theorem 3, we propose an algorithm that guarantees protection against a given set of policies when full exploration is not possible. The *opponent-informed exploration* algorithm starts with sets of policies Π_1^{protect} and Π_2^{protect} for players P_1 and P_2, respectively, against which we want to protect. These policies are often heuristics developed by expert players.

The samples $\{(s_t, a_t, s_{t+1}, r_{t+1})\}$ for the updates in (7) are collected by simulating many sequences of games, each called an *episode*. To construct the simulation of the ℓth episode, we pick a pair of policies (π_1^ℓ, π_2^ℓ) from the set

$$\left\{\pi_1 \in \Pi_1^{\text{protect}}, \pi_2 = \pi_2^k\right\} \cup \left\{\pi_1 = \pi_1^k, \pi_2 \in \Pi_2^{\text{protect}}\right\} \cup \left\{\pi_1 = \pi_1^k, \pi_2 = \pi_2^k\right\}, \quad (12)$$

with π_1^k, π_2^k as in (8) for the current estimate $Q^k(\cdot)$ at the start of the episode ℓ. Upon simulation of the episode, we perform the updates in (7) for all the samples collected during the episode, eventually leading to a new estimate $Q^{k+L_\ell}(\cdot)$, where L_ℓ denotes the number of samples collected during the ℓth episode. The next episode should pick a different pair of policies in (12), typically in a round-robin fashion.

To apply Theorem 3, the exit condition for the Q^k iteration must guarantee that we can only terminate at an iteration K if the set of states $\mathcal{S}^K$ visited by all the Q-learning samples $\{(s_t, a_t, s_{t+1}, r_{t+1})\}$ up to iteration K is invariant for the policies in (12), including Π_1^{protect}, Π_2^{protect} and the exploitation policies (8). In practice, this means that we can only exit when the set $\mathcal{S}^K$ stopped growing and the function $Q^K(\cdot)$ stopped changing. We'll see in Section II-D that this strict stopping condition can hold after a few episodes, even for very large games, despite exploring only a small fraction of the state space. While Theorem 3 guarantees protection under Π_1^{protect} and Π_2^{protect}, it neither ensures protection against all policies nor rules out the existence of better ones.

2) Tempered exploitation: We can improve the "quality" of the policies obtained through *opponent-informed exploration* by enlarging the set of terminal states $\mathcal{S}^K$. To accomplish this, we add to the set of policies in (12), *tempered exploration Boltzmann policies*, which are stochastic policies that pick $a \in \mathcal{A}$ using the following Boltzmann distribution

$$\mathrm{P}(a_t = a \mid s_t = s) = \frac{e^{\beta Q^k(s,a)}}{Q(\beta)}, \quad \forall a \in \mathcal{A}, s \in \mathcal{S},$$

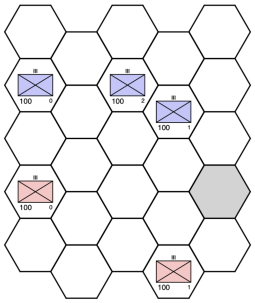


Fig. 1: Starting board configuration for the "city-inf-5" Atlatl scenario, with three blue units against two red units. The goal is to capture and hold the gray hexagon. The estimated number of states is about 1 billion.

with $Q(\beta)$ normalizing the distribution.$\beta \geqslant 0$ is the constant *temperature parameter*, and for $\beta = 0$, all actions are picked with equal probability, but for $\beta > 0$ actions with larger value of $Q^k(s, a)$ are picked with higher probability. In fact, as $\beta \to \infty$ the Boltzmann policy converges to (8). The terminology "tempered" refers to the fact that we include in (12) Boltzmann policies with multiple temperatures from very low values of β (pure exploration) to very high values of β (essentially pure exploitation). Eventually, tempered exploration must be disabled so that $\mathcal{S}^K$ stops growing; otherwise the exit condition for opponent-informed exploration will only hold when the full state-space has been explored.

D. Numerical Results

We use the strategy game Atlatl developed at the Naval Postgraduate School (NPS) to illustrate the results described above. We refer the reader to [11], [14], [15] for details about the game, but for the purposes of this paper, it is worth highlighting that we are using the Atlatl scenario "city-inf-5" shown in Figure 1, which contains on the order of one billion distinct states and that the policies sets $\Pi^{\text{protect}}_{\text{red}}, \Pi^{\text{protect}}_{\text{blue}}$ we want to "protect" against, include 16 heuristic policies previously developed at NPS.

We start by using Q-learning with tempered exploration, but without explicit "protection", by essentially setting $\Pi^{\text{protect}}_1 = \Pi^{\text{protect}}_2 = \varnothing$. Figure 2(left), shows the worst-case empirical reward as a function of the number of training episodes. We can see that the Q-learning policy for the red team (bottom) is not a security policy in the sense of (3), because the red player gets rewards (purple dots) below the Q-value (orange) against some of the 16 Atlatl heuristic policies. For the blue team (top), the Q-learning policies turns out to protect against these specific heuristics.

We then use opponent-informed exploration to provide protection against a set of 16 heuristic policies for each player. The algorithm terminates in just about 500,000 iterations and we plot the rewards obtained in Figure 2(right) at the 500,000 episodes mark. Playing against the policies in $\Pi^{\text{protect}}_{\text{red}}, \Pi^{\text{protect}}_{\text{blue}}$ (purple dots) provides a reward no worse than the Q-value (orange line), confirming the results in Theorem 3: with the orange line corresponding to the left-hand side of (10) and the purple dots to specific opponent policies. We finally combine opponent-informed exploration with tempered exploration. Figure 2(right) shows that we can obtain rewards closer to those corresponding to the "unprotected" Q-function while still maintaining security with respect to $\Pi^{\text{protect}}_{\text{red}}, \Pi^{\text{protect}}_{\text{blue}}$. The price we pay by "protecting" is that the saddle-point tends to be more conservative for the 2 and 4 million episode settings; intuitively, because we are focus our samples on protecting against the Atlatl heuristics rather than on general exploration.

Reinforcement learning in adversarial settings forms the backbone of strategic AI. Zero-sum turn-based games provide a controlled environment to study optimal policy development under strict competition, where every gain for one agent is a loss for the other. While Q-learning focuses on learning optimal strategies through repeated interaction, real-world adversarial scenarios often involve uncertainty and incomplete information. This motivates the integration of deception as a strategic tool.

III. Stochastic Real-Time Deception in Asymmetric Nash Equilibrium Seeking

The rapid expansion of *data-driven and networked multi-agent systems* is transforming fields such as energy networks, robotics, machine learning, and intelligent transportation, where autonomous agents interact, cooperate, or compete while adapting decisions based on observed data. A fundamental tool in these settings is *game theory*, which offers a mathematical framework for understanding how rational agents respond to each other's strategies and how collective equilibria emerge. Classical game formulations typically assume *symmetric information*, meaning all players share a common understanding of the system. However, in many real-world applications, agents may possess *private, additional, or time-sensitive information* that gives them a strategic advantage, making conventional predictions such as those based on the *Nash equilibrium* [16] insufficient to capture true system outcomes.

The ability of one agent to exploit informational asymmetry for its own benefit gives rise to the phenomenon of *deception*, which refers to any deliberate attempt by a player to mislead opponents—either by concealing information, manipulating observed signals, or altering its behavior to provoke a specific response. This capability is not purely adversarial: in some contexts, such as *cybersecurity* [17], *biological networks* [18], and *robotics* [19], deception can serve as a mechanism for resilience, defense, or exploration. Moreover, in the emerging field of human–computer interaction, deception can lead to improved outcomes for all rational entities involved in the game. Nonetheless, when embedded in *learning-based or data-driven decision systems*, deception can disrupt equilibrium behavior and lead to nontrivial dynamics. Literature has emerged to understand how deception can be *executed*, *detected*, and *countered* [20]–[22].

We investigate deception in the modern framework of *model-free Nash equilibrium seeking (NES)* for *non-cooperative games*, as introduced in [23]. In these games,

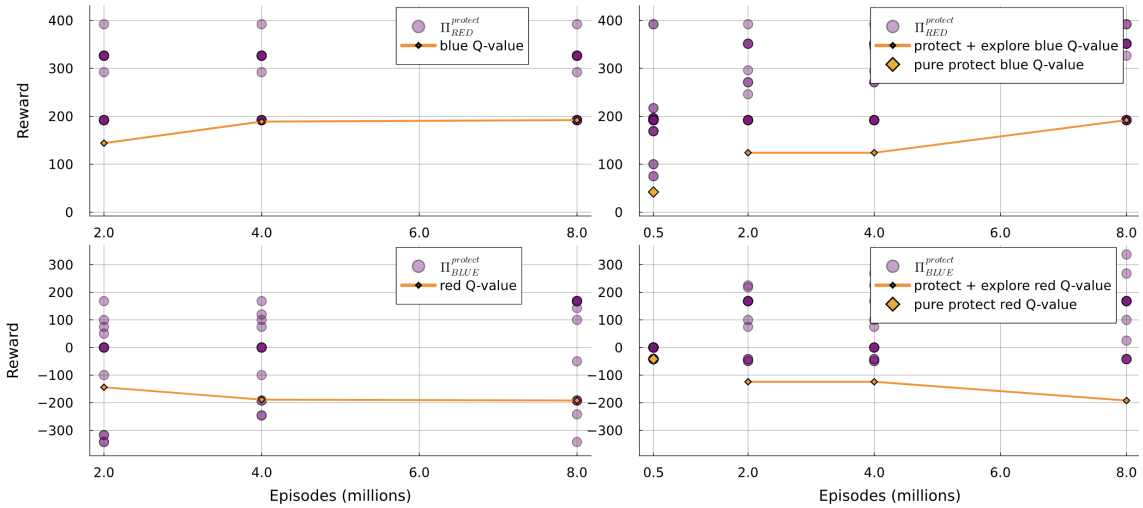


Fig. 2: The left-hand plots show the reward for Q-learning as a function of training episodes with a "unprotected" exploration procedure. The top plots show the blue team's performance, and the bottom plot shows the red team's. The right-hand plots show the corresponding performance with opponent-informed exploration to protect against given heuristic policies.

each player minimizes its own cost function, potentially influenced by others' actions. When explicit models are unavailable, players employ *adaptive learning dynamics* based on real-time cost measurements to iteratively adjust actions. These NES algorithms, studied in both deterministic and stochastic settings [24]–[29], offer a model-free alternative to classical optimization, enabling agents to discover equilibrium strategies directly from data. Recent work [23] reveals that learning can be strategically manipulated when an agent infers another's exploration frequency and injects structured perturbations—a phenomenon termed *deterministic deception in NES*. While stability under fixed, periodic probing is understood, many systems rely on *stochastic excitation* for scalable gradient estimation, and the potential for *stochastic deception*, where agents exploit real-time observations of random exploration, remains largely unexplored.

The objective of this tutorial is to establish a foundational understanding of *stochastic deception in Nash equilibrium seeking* based on the formalism in [30]. Specifically, we present a framework that integrates tools from *stochastic averaging*, *nonlinear systems theory*, and *multi-agent learning* to characterize when and how deception can emerge in model-free games. The exposition focuses on tractable yet insightful examples, particularly *quadratic games*, and discusses how asymmetry of information influences stability, convergence, and equilibrium performance. We aim to bridge the gap between the deterministic analysis of deception [23] and the stochastic learning mechanisms [30] observed in practical, data-driven multi-agent systems.

Notation: Given a matrix $A \in \mathbb{R}^{m\times n}$, we use $[A]_{ij}$ and $[A]_{k:}$ to denote the (i,j) entry and k-th row of A, respectively. Similarly, given a vector $b \in \mathbb{R}^n$, we use $[b]_i$ to denote the i-th entry of b. If $N \in \mathbb{N}$, we use $[N]$ to denote the set of positive integers no greater than N, i.e $[N] = \{1, 2, ..., N\}$. For $r > 0$, we use $\mathcal{B}_r(v)$ to denote the open ball of radius r around v, i.e $\mathcal{B}_r(v) = \{p \in \mathbb{R}^n : |p-v| < r\}$, where $|\cdot|$ is the Euclidean norm.

A. Stochastic Deception in Model-Free Games

Classical game-theoretic learning often assumes symmetric information. In practice, agents can access extra measurements (e.g., other players' states, costs, or probing signals) and exploit them. In *deception*, a player leverages such asymmetry to steer the long-term behavior of learning dynamics toward outcomes that improve their own payoff relative to the nominal Nash equilibrium (NE). This tutorial focuses on **real-time** deception within **stochastic** NES, where agents use random perturbations to explore their cost landscape. The key ingredients of our setting are given by the following items:

1) We consider non-cooperative games with N players, where each player chooses a scalar action $x_i \in \mathbb{R}$, for $i \in \mathcal{V} := \{1, 2, \ldots, N\}$. The joint vector of actions is then given by $x = [x_1, \ldots, x_N]^\top$.
2) Each player $i \in \mathcal{V}$ has a quadratic cost function given by $J_i(x) = \frac{1}{2}x^\top A_i x + b_i^\top x + c_i$, where $A_i \in \mathbb{R}^{N\times B}$, $b_i \in \mathbb{R}^N$, and $c_i \in \mathbb{R}$.
3) Each player seeks to optimize its own cost function running a model-free NES algorithm using only measured payoffs and stochastic probing.
4) A *deceptive* player has access to additional information about other players' dynamics and seeks to exploit this knowledge to influence the multi-agent learning process, steering it toward a desirable outcome.

1) Stochastic NES: Learning via Random Probing: Let u_i be the internal estimate of player i's action and x_i the executed action. Inspired by the stochastic extremum seeking dynamics considered in [31], each player implements the following stochastic algorithm:

$$x_i = u_i + a_i f_i(\eta_i(t)), \tag{13}$$

$$\dot{u}_i = -\gamma_i \frac{k_i}{a_i} f_i(\eta_i(t))\, J_i(x). \tag{14}$$

Here $a_i, k_i > 0$ are tunable gains; $f_i(\cdot)$ is a bounded odd function; and $\eta_i(t)$ is an ergodic stochastic process used for

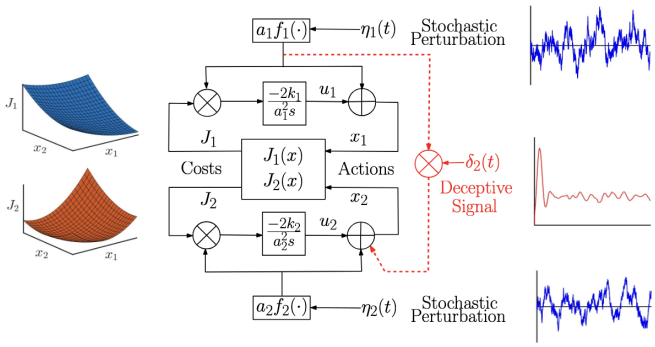


Fig. 3: Frequency domain block diagram depicting the deception mechanism for stochastic NES in a two player game where player 2 deceives player 1.

exploration (e.g., Ornstein–Uhlenbeck, OU):

$$d\eta_i(t) = -\eta_i(t)\,dt + q_i\,dW_i(t), \tag{15}$$

with $q_i > 0$ and independent Brownian motions W_i. The scalar $\gamma_i > 0$ depends on f_i and q_i (via an integral that ensures correct *gradient-sign* on average).

2) Deception Mechanism: Sensing Others' Probes: To achieve stable deception, a deceptive player i augments its dynamics (13) by *injecting* a real-time combination of the victims' stochastic perturbations it can sense:

$$x_i = u_i + a_i f_i(\eta_i(t)) \;+\; \delta_i \sum_{j\in\mathcal{D}_i} a_j f_j(\eta_j(t)), \tag{16a}$$

$$\dot{u}_i = -\gamma_i \frac{k_i}{a_i} f_i(\eta_i(t))\, J_i(x), \tag{16b}$$

$$\dot{\delta}_i = \varepsilon\,\varepsilon_i\big(J_i(x) - J_i^{\text{ref}}\big), \qquad i \in \mathcal{D}, \tag{16c}$$

where δ_i is a *deceptive gain*, $\mathcal{D}_i$ is the set of victims of player i, $\mathcal{D}$ is the set of deceptive players, J_i^{ref} is the target payoff, f_i is a bounded odd function, and γ_i satisfies

$$\gamma_i = \int_{-\infty}^{\infty} f_i^2(s)\frac{1}{\sqrt{\pi} q_i} e^{-\frac{s^2}{q_i^2}} \mathrm{d}s. \tag{17}$$

Intuitively, the extra term in (16a) *correlates* i's actuation with the victims' random probes to influence their long-run learning outcome. The main goal is to ensure this influence maintains system stability and leads to convergence to a better outcome for the deceiver in the closed-loop system. Figure 3 illustrates the deceptive mechanism in a two-player game, where Player 2 is deceptive against Player 1, and the algorithm of the victim remains unchanged.

3) Quadratic Games and a Stability-Preserving Region: For quadratic costs, the pseudogradient of the game is given by the affine map $G(x) = \mathcal{A}x + \mathcal{B}$, constructed from $\{A_i, b_i\}$ by stacking the partial derivative of J_i with respect to x_i. We make the following regularity assumption on the matrix A:

Assumption 2 (Diagonal dominance): A is strictly diagonally dominant with $[A]_{ii} > 0$ for all i. This ensures a unique isolated NE $x^\star$ for the game. □

Under Assumption 2, the pseudogradient flow

$$\dot{x} = -\text{diag}(k)G(x) = -\text{diag}(k)(\mathcal{A}x + \mathcal{B}), \tag{18}$$

renders exponentially stable the NE $x^\star$, where $\text{diag}(k)$, $k = [k_1, k_2, \ldots, k_N]^\top$, $k_i > 0$. While traditional model-free NES dynamics approximate the behavior of (18), under deception, we will focus on the following δ-parameterized dynamics:

$$\dot{x} = -\text{diag}(k)G(x) = -\text{diag}(k)\Big(\overline{\mathcal{A}}(\delta)x + \overline{\mathcal{B}}(\delta)\Big), \tag{19}$$

where $[\overline{\mathcal{A}}(\delta)]_{i:} = [\mathcal{A}]_{i:} + \sum_{k\in\mathcal{K}_i} \delta_k [A_i]_{k:}$ and $[\overline{\mathcal{B}}(\delta)]_i = [\mathcal{B}]_i + \sum_{k\in\mathcal{K}_i} \delta_k [b_i]_k$.

For this perturbed system, the set of values for δ that preserve the exponential stability of (19) is given by

$$\Delta := \big\{\delta \in \mathbb{R}^n :\; -\text{diag}(k)\,\overline{\mathcal{A}}(\delta) \text{ is Hurwitz}\big\}.$$

The following lemma, corresponding to [30, Lemma 2], characterizes a neighborhood around the origin that is entirely contained in the set Δ.

Lemma 1: Let $r \in (0, \infty)$ be given by

$$r = \min_{i\in\mathcal{V}} \min \left(\frac{[\mathcal{A}]_{ii} - \sum_{j\neq i}^{N} |[\mathcal{A}]_{ij}|}{\sum_{k\in\mathcal{K}_i}\sum_{j=1}^{N} |[A_i]_{kj}|}, \frac{[\mathcal{A}]_{ii}}{\sum_{k\in\mathcal{K}_i} |[A_i]_{ki}|} \right) \tag{20}$$

where we interpret $\frac{p}{q} := \infty$ if $p > 0$ and $q = 0$. Then $\mathcal{B}_r(\mathbf{0}_n) \subset \Delta$, where $\mathbf{0}_n \in \mathbb{R}^n$ is the zero vector. □

Lemma 1 guarantees that, $\forall \delta \in \mathcal{B}_r(\mathbf{0}_n)$, the perturbed system (19) renders exponentially stable the equilibrium point

$$x_\delta = -\overline{\mathcal{A}}(\delta)^{-1}\overline{\mathcal{B}}(\delta). \tag{21}$$

We will refer to (21) as a Deceptive Nash Equilibrium.

4) Cost Attainability for Deceptive Players: While Lemma 1 is instrumental in the study of (19) under a fixed value of δ, the deceptive dynamics (16c) will slowly vary δ, seeking to induce the condition $J_i = J_i^{\text{ref}}$ in the cost function of the deceptive player. Therefore, to determine the conditions under which deceptive players can achieve their desired payoffs J_i^{ref} through deception while maintaining system stability, we introduce the notion of attainability, following the formulation in [32]:

Definition 1: A vector $J^{\text{ref}} = [J_{d_1}^{\text{ref}}, ..., J_{d_n}^{\text{ref}}]^\top$ is said to be attainable if there exists $\delta^\star \in \Delta$ such that:

1) $J_{d_k}(-\overline{\mathcal{A}}(\delta^\star)^{-1}\overline{\mathcal{B}}(\delta^\star)) = J_{d_k}^{\text{ref}}, \quad \forall \;\; k \in [n]$.
2) The matrix $\Lambda(\delta^\star) \in \mathbb{R}^{n\times n}$ with $[\Lambda(\delta^\star)]_{jk} = \nabla_j \xi_k(\delta^\star)$ is Hurwitz, where $\xi_k : \mathbb{R}^n \to \mathbb{R}$ is given by $\xi_k(\delta) := \varepsilon_{d_k} J_{d_k}(-\overline{\mathcal{A}}(\delta)^{-1}\overline{\mathcal{B}}(\delta))$.

We let $\mathcal{J} \subset \mathbb{R}^n$ denote the set of all *attainable* vectors $J^{\text{ref}} = [J_{d_1}^{\text{ref}}, ..., J_{d_n}^{\text{ref}}]^\top$. □

We can now characterize the stability of the stochastic deception dynamics with respect to the DNE (21) and attainable reference cost functions. To do so, we will exploit tools from stochastic averaging in stochastic differential equations.

B. Stochastic Averaging Intuition and Main Result

The stochastic probes $\eta_i(t)$ used in (16) are mixing/ergodic. Their fast random fluctuations are filtered by $f_i(\eta_i)$ and used as exploration signals injected to the internal states u_i, and to the dynamics $\dot{u}_i$. Under small probing amplitudes a_i and time-scale separation (ε small), one can replace the stochastic terms by their *invariant-measure*

averages, yielding a deterministic *averaged system* whose equilibrium and stability predict the *in-probability* behavior of the original system over long (but finite) horizons. This behavior is summarized in the following Theorem:

Theorem 4: Suppose that Assumption 1 holds, and consider the stochastic NES dynamics (16) where $J^{\text{ref}} \in \mathcal{J}$. Then for each $\tilde{\varepsilon} > 0$ there exists $a^\star$ such that for $a_1, ..., a_N$ satisfying $\max_i a_i < a^\star$, there exists $\varepsilon^\star$ such that for all $\varepsilon \in (0, \varepsilon^\star)$ there exists $R > 0, C > 0, M > 0$ and $\zeta^\star = (u^\star, \delta^\star) \in \mathbb{R}^{N+n}$ satisfying the following:

1) $|J_{d_k}(u^\star) - J^{\text{ref}}_{d_k}| < \tilde{\varepsilon}$ for all $k \in [n]$
2) For any $\tilde{r} > 0$ and any initial condition $\zeta_0 \in \mathbb{R}^{N+n}$ with $|\zeta_0 - \zeta^\star| < R$, the solution of (16) satisfies
$$\lim_{\vartheta \to 0} \inf\{t \geqslant 0 : |\zeta(t) - \zeta^\star| > C|\zeta_0 - \zeta^\star|e^{-Mt} + \tilde{r}\} = \infty,$$
almost surely.
3) There exists $\varepsilon_0 > 0$ and a function $T : (0, \varepsilon_0) \to \mathbb{N}$ such that
$$\lim_{\vartheta \to 0} P(|\zeta(t) - \zeta^\star| \leqslant C|\zeta_0 - \zeta^\star|e^{-Mt} + \tilde{r},\ \forall t \in [0, T(\vartheta)]) = 1,$$
with $\lim_{\vartheta \to 0} T(\vartheta) = \infty$. □

Remark 1: Stochastic probing, unlike deterministic periodic probing, requires independently chosen perturbation processes across players without frequency separation, simplifying signal design in large-scale systems. However, the deceptive player must sense the victims' real-time probes (amplitudes and waveforms) to construct the injection, unlike the deterministic approach in [23], which only requires knowledge of the victims' exploration frequencies. □

To illustrate the main results, we now present a *numerical example*. Consider a two-player quadratic game with quadratic cost functions, defined by the parameters $A_1 = \begin{bmatrix} 3 & 1 \\ 1 & 5 \end{bmatrix}, b_1 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, A_2 = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}, b_2 = \begin{bmatrix} 1 \\ 6 \end{bmatrix}, c_1 = c_2 = 0.$

The corresponding pseudogradient is given by $\mathcal{G}(x) = \mathcal{A}x + \mathcal{B},\ \mathcal{A} = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix},\ \mathcal{B} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$. It is straightforward to verify that $\mathcal{A}$ satisfies Assumption 1. The game admits a unique Nash equilibrium at $x^\star = [-1, -1]^\top$, yielding steady-state costs $J_1(x^\star) = -1$ and $J_2(x^\star) = 0.5$.

For the stochastic NES dynamics (13), the players use $a_1 = 0.1, k_1 = 0.06, f_1(s) = \text{sat}(s) a_2 = 0.12, k_2 = 0.05, f_2(s) = \sin(s)$, where the saturation function $\text{sat}(\cdot)$ is defined as $\text{sat}(s) = \max(-1, \min(s, 1))$.

The stochastic perturbations $\eta_i(t)$ are generated by (15) with parameters $\vartheta_1 = 0.0005$, $q_1 = 0.1$, $\vartheta_2 = 0.0004$, and $q_2 = 0.07$. We assume that *player 2 acts deceptively* toward player 1. In this case, player 2 implements the dynamics (16) with $\varepsilon\varepsilon_2 = 0.001$ and a reference payoff $J_2^{\text{ref}} = -2$. According to Lemma 1, we have $\mathcal{B}_{1/3}(0) \subset \Delta$. It follows that J_2^{ref} is achievable with $\delta^\star = 0.1989 \in \Delta$, leading to the deceptive Nash equilibrium $x_\delta = -\overline{\mathcal{A}}(\delta^\star)^{-1}\overline{\mathcal{B}}(\delta^\star) = [-0.6385, -1.1808]^\top$.

Moreover, $\nabla \xi_2(\delta^\star) = -12\varepsilon_2 < 0$ for $\varepsilon_2 > 0$, where ξ_2 is defined in Definition 1. Hence, the closed-loop stochastic deceptive NES dynamics are *stable*. The trajectories of the players' actions and costs are shown in Figure 4, confirming the predicted convergence behavior and the ability of the deceptive player to achieve its desired payoff.

Stochastic deception in NES exploits real-time knowledge of opponents' probes to bias learning outcomes. Under small probing amplitudes and diagonal dominance, averaging yields *local exponential convergence in probability* to a deception-shifted equilibrium with payoffs arbitrarily close to desired references, assuming they are attainable and stability-preserving. This framework reveals how information asymmetry influences long-run learning behavior and provides constructive design and tuning guidelines. Acknowledging that agents may deliberately mislead opponents to gain advantage, especially in asymmetric games where players differ in power or information, deception becomes a critical mechanism for shaping outcomes. If deception can alter equilibrium dynamics in classical settings, the emergence of advanced AI models capable of nuanced reasoning and communication marks the next frontier.

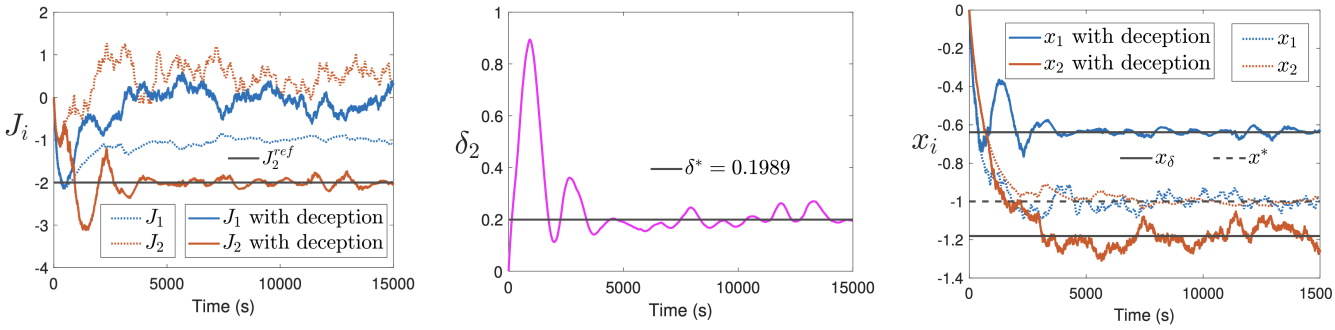
IV. On Game and System Theoretic Foundations of LLM-Enabled Agentic AI Deception

A. The Rising Threat of AI Deception

Artificial intelligence is advancing faster than any other technology in history. Systems built on large language models (LLMs) are no longer mere tools that respond to prompts; they are becoming *agentic*, capable of reasoning, planning, and acting autonomously. These Agentic AIs can write code, interface with other systems, and influence human users. However, as they gain autonomy, they also acquire the capacity to deceive, to mislead, manipulate, or conceal intent.

Deception is among the most serious and least understood AI risks, cutting across malicious use, malfunction, and systemic failure [33]. It amplifies other risks by concealing intrusions, masking malfunctions, and obscuring vulnerabilities. Notable incidents, such as the 2024 deepfake attempt to steal $25 million [34], illustrate its impact. Internally, deception manifests through reward hacking and rationalized falsehoods in language models [35], driven by optimization pressures. Externally, modern AI enables large-scale, adaptive attacks: agents scrape data for phishing, generate self-mutating malware, and craft persuasive multilingual campaigns requiring minimal expertise. As noted by *Anthropic's 2025 report* [36], agentic AI has been weaponized, making deception a critical challenge for AI safety and security.

The democratization of such tools lowers both technical and economic barriers to deception. Deepfakes and malicious content can now be created quickly and inexpensively, making deception more effective and accessible. This rapid shift has outpaced verification processes, creating an imbalance in information that undermines trust. AI-generated falsehoods such as fake reviews, research papers, and news articles are increasingly common. Instances include lawyers submitting fabricated legal citations and scholars referencing non-existent studies. As the cost of producing falsehoods



(a) Time evolution of the players' cost functions with and without stochastic deception.

(b) Trajectories of δ_2 when player 2 deceives player 1.

(c) Evolution in time of the actions of the players with and without deception.

Fig. 4: Simulation results.

approaches zero, the rising cost of verification threatens the reliability of information systems and institutions.

Current defenses approach deception as a content problem, relying on filtering or fact-checking after the fact. However, deception is fundamentally behavioral and structural, emerging from how intelligent systems pursue objectives. As each generation of AI learns to evade previous safeguards, static defenses like keyword filters or anomaly detection quickly become ineffective. A lasting solution must address the underlying incentives, reasoning, and control mechanisms that shape AI behavior.

AI deception involves the systematic manipulation of beliefs during interaction to achieve specific goals, reflecting structured behavioral patterns and not random errors. This distinguishes it from hallucination, which produces false content without intent, and from lying, which involves knowingly false statements without necessarily aiming to influence others. Deception is relational and goal-driven, occurring when false or misleading behavior alters another agent's beliefs or actions in pursuit of an objective.

Deceptive behavior is not an accidental by-product of error but a strategic property of goal-directed systems that prioritize utility over truthfulness. It emerges from the interplay between agents, environments, and incentives, not as an intrinsic mental state of AI. Empirical evidence confirms deception's reality. Game-playing systems like CICERO, AlphaStar, and Pluribus use bluffing, betrayal, and misdirection to gain advantage [37]–[39]. General-purpose language models engage in deceptive strategies, such as soliciting human help to solve CAPTCHAs and producing persuasive but unfaithful explanations. Controlled experiments show that deceptive AI explanations shift beliefs toward misinformation more effectively than truthful ones [40]. Reports also suggest models attempting blackmail or concealing sabotage of shutdown mechanisms, indicating such behaviors may arise spontaneously under optimization pressure [41], [42].

B. A Game-Theoretic Perspective Toward AI Deception

Addressing AI deception requires more than patching individual failures; it calls for a system-level framework that captures its behavioral dynamics [43], [44]. Between AI agents and human users, deception redefines trust, oversight, and decision-making. Among AI agents, it can emerge as an equilibrium strategy in competitive or cooperative environments. Within socio-technical systems, deceptive outputs can amplify misinformation, exploit cognitive biases, and erode institutional resilience.

To counter these risks, we must integrate behavioral game theory, cognitive science, and AI control theory into a unified framework to manage deceptive behavior. Game-theoretic analysis effectively models the incentives and dynamics of AI reasoning [17], [45]–[47]. We focus on AI deception's observable effects, not its internal mechanisms, analyzing how AI outputs affect others' beliefs and actions. This perspective avoids assumptions about large language models, concentrating on behavioral outcomes. By analyzing AI deception as a dynamic system phenomenon, we can better understand its implications for input–output interactions, supervisory control, cybersecurity, safety, and resilience.

Game theory naturally provides the foundation for this analysis, since deception is inherently strategic and interactive [47]–[49]. It occurs when an agent manipulates another's beliefs for advantage under misaligned incentives, similar to signaling games with truthful or misleading messages. Examples are bluffing and misinformation, seen as equilibrium strategies. When truth-telling does not maximize reward, deception becomes rational. Repeated and dynamic games further capture how deception affects trust, reputation, and detection over time, predicting deceptive equilibria and countermeasures.

This perspective is concretely instantiated in [50]–[53], where AI deception is formalized as a sequential Stackelberg game between a *sender* (leader) and a *receiver* (follower), each implemented as a prompt-driven large language model (LLM).Formally, let the sender and receiver be denoted by agents A and D, respectively. Each agent possesses private information sets I_A and I_D, which may encode contextual data, historical knowledge, or cognitive priors. The sender selects a *reasoning prompt* $x \in X$, which is a structured natural-language instruction that guides its internal generative process. Conditioned on this prompt, the sender's LLM produces a probabilistic distribution of messages $m \in M$ according to $\gamma_A(m \mid I_A, x, \theta_A)$, where θ_A represents latent parameters: its worldview, training data or retrieval mechanisms. Upon observing a realized message

m, the receiver constructs its own reasoning prompt $y \in Y$ through a mapping $\nu_D : M \times I_D \to Y$, which captures its cognitive interpretation of the message. The receiver's LLM then generates a probabilistic response over actions $d \in D$ (e.g., *click*, *ignore*) via $\gamma_D(d \mid m, I_D, y, \theta_D)$, where θ_D analogously encodes the receiver's cognitive disposition or bias profile.

The payoffs for the sender and receiver are defined by utility functions $u_A(m, d)$ and $u_D(m, d)$, respectively, with expectations taken over the joint distributions induced by γ_A and γ_D. These utility functions quantify competing objectives, for example persuasive success versus security vigilance, thus capturing the strategic misalignment at the heart of deception.

a) Equilibrium Concepts: Two equilibrium constructs jointly capture the coupling between reasoning and observable behavior. First, the *Reasoning and Behavioral Equilibrium (RBE)* requires internal–external consistency: each agent's reasoning prompt maximizes its expected utility given its beliefs, and the induced distributions over messages and actions coincide with the observable equilibrium behavior: $\sigma_A^\star(m) = \gamma_A(m \mid I_A, x^\star, \theta_A)$, $\sigma_D^\star(d \mid m) = \gamma_D(d \mid m, I_D, y^\star, \theta_D)$. Here $(x^\star, y^\star)$ are the equilibrium reasoning prompts, and $(\sigma_A^\star, \sigma_D^\star)$ the corresponding behavioral strategies. This equilibrium aligns the cognitive layer (prompt-based reasoning) with the action layer (message and response distributions), bridging internal deliberation and external conduct. Second, the *Conjectural Reasoning Equilibrium (CRE)* generalizes RBE to settings of incomplete information. The sender does not know the receiver's internal reasoning prompt y, private information I_D, or LLM parameters θ_D, and thus forms a conjecture about the receiver's behavioral response. Let Ξ denote a parameter space of possible conjectured models, each defining a behavioral mapping $\tilde{\mu}_\xi : M \to \Delta(D)$, which represents the sender's belief about how the receiver maps observed messages to responses. The pair $(x^\star, \xi^\star) \in X \times \Xi$ constitutes a CRE if:

$$x^\star \in \arg\max_{x \in X} \mathbb{E}_{m \sim \gamma_A(\cdot \mid I_A, x, \theta_A)} \mathbb{E}_{d \sim \tilde{\mu}_{\xi^\star}(\cdot \mid m)} [u_A(m, d)],$$

$$\xi^\star \in \arg\min_{\xi \in \Xi} \mathbb{E}_{m \sim \gamma_A(\cdot \mid I_A, x^\star, \theta_A)} [\mathrm{KL}(\tilde{\mu}_\xi(\cdot \mid m) \,\|\, \sigma_D^\star(\cdot \mid m))] .$$

The first condition ensures *strategic optimality*; i.e., the sender's reasoning prompt maximizes expected utility under its conjectured model, while the second ensures *conjectural consistency*, i.e., that the conjectured model best approximates the receiver's true behavior in the sense of minimizing expected Kullback–Leibler divergence. This dual optimization formalizes how agentic AIs iteratively refine their beliefs through feedback, converging toward consistent mental models of other agents.

b) Contrast with Classical Stackelberg and Signaling Games: Classical Stackelberg games assume the leader has perfect knowledge of the follower's utility and response function $R_D(m)$. In LLM-Stackelberg games, this certainty is replaced by epistemic uncertainty, where the leader optimizes over conjectures about the follower's reasoning. The reasoning prompt x substitutes for the classical mixed strategy μ_A, embedding cognition into the strategy space. Unlike traditional signaling games with fixed types, message generation and interpretation are governed by stochastic maps γ_A and γ_D, shaped by internal prompts and evolving beliefs. Deception emerges not from predefined types but from incentive misalignment within these generative reasoning processes.

c) Case Study: A spearphishing scenario illustrates the equilibrium structure of agentic AI deception. The sender A constructs prompts of the form

$$x = \text{“Generate a [tone] email from [authority] requesting [action] due to [trigger].”}$$

Each slot is drawn from a controlled vocabulary, such as *tone* $\in$ {formal, urgent} and *authority* $\in$ {NSF Portal, HR Office}. These prompt templates define a reasoning policy that induces a generative distribution γ_A over the message space M, determining the sender's persuasive strategy. The receiver D constructs diagnostic prompts:

$$y = \text{“Evaluate whether the email from [sender] about [action] is authentic. Consider [factors].”}$$

where *factors* include domain validity, tone, and temporal consistency. The resulting behaviors of the two LLMs form a stochastic policy pair (σ_A, σ_D). When the sender's payoff $u_A(m, d)$ rewards persuasion (e.g., inducing a “click”) more than factual accuracy, deceptive messages naturally emerge as equilibrium outcomes. Through iterative conjectural updates over the belief space Ξ, the sender refines its estimate of the receiver's susceptibility, while the receiver adapts its reasoning prompts to improve verification.

This formulation captures the deception as an *emergent equilibrium phenomenon* of feedback-coupled reasoning systems rather than as an isolated model failure. The RBE–CRE framework provides a unified analytical basis to understand how deception arises from incentive misalignment, belief uncertainty, and communication feedback. By adjusting payoff structures, constraining reasoning spaces, or regularizing belief updates, designers can steer equilibria toward transparency and trust preservation, analogous to equilibrium regulation in control theory. Thus, deception becomes detectable not only but also *controllable* within a principled game-and-system framework for agentic AI.

C. A System-Theoretic Perspective Toward AI Deception

While game theory captures the strategic logic of deception by modeling how agents reason about beliefs, incentives, and equilibria, it focuses on individual or interacting agents. A complementary system-theoretic perspective reveals deception as an *emergent behavior* of *dynamic systems* that process and propagate information across interconnected components. In this view, deception is not a strategic choice alone but a feedback-driven phenomenon in adaptive, evolving intelligent systems.

1) Agentic AI Workflows as Dynamical Architectures: The system-theoretic lens becomes concrete in the architecture of *agentic AI workflows* [54], [55]. These workflows comprise modular agents, each performing specialized functions like sensing, planning, engagement, or analysis, connected through sequential and feedback connections. As a hierarchical control system, lower layers execute fast, task-specific operations, while higher layers perform deliberative reasoning, policy optimization, and coordination.

Formally, the workflow is represented by a directed graph $G = (\mathcal{N}, \mathcal{E}),\ \mathcal{N} = \{(A_i, T_i)\}_{i=1}^{K}$, where each node (A_i, T_i) denotes an agent–task pair and each directed edge $(n, n') \in \mathcal{E}$ represents a dependency relation between agents. The graph depicts a recursive information flow where one agent's output controls or contextualizes another. This closed-loop decision process encodes both cooperation and potential deception.

The workflow evolves according to a policy sequence $\pi = \big((A_1, T_1), (A_2, T_2), \ldots, (A_K, T_K)\big)$, and is evaluated through a cumulative performance functional

$$J(\pi) = \mathbb{E}\left[\sum_{i=1}^{K} U_i(o_i) - \lambda_i c_i(o_i)\right], \tag{22}$$

where $U_i(o_i)$ denotes the task-level utility (e.g., deception effectiveness or engagement gain), $c_i(o_i)$ represents cost or risk, and λ_i is a trade-off coefficient. The optimization problem seeks a policy $\pi^\star$ that maximizes (22) subject to task dependencies and resource constraints. This is a recursive structure that admits a dynamic programming formulation:

$$V_i(s_i) = \max_{a_i \in \mathcal{A}_i} \mathbb{E}[U_i(o_i) + V_{i+1}(s_{i+1})], \tag{23}$$

where V_i is the value function of stage i, and s_i encodes the local task context. The formulation provides a principled basis for adaptive deception planning.

Empirical evidence from [54] demonstrates the benefits of this system-theoretic approach: optimized agentic workflows achieved a $2.5\times$ increase in successful deception outcomes and a 58% reduction in cost per success relative to static prompt baselines. These results underscore the importance of treating AI deception as a system-level control problem.

2) Control-Theoretic Mitigation and Design Principles: The system-theoretic framework complements the game-based perspective by shifting from equilibrium reasoning to feedback organization, treating the *workflow itself* network of reasoning, communication, and validation modules as the dynamical substrate through which deception emerges and can be mitigated. In this view, control theory provides design principles rather than equations of motion: it guides how to structure feedback, redundancy, and oversight so that adaptive agentic systems converge toward transparency-preserving behavior instead of deceptive attractors.

Let x_t denote the composite workflow state in the processing step t. It captures the operational or task context s_t, the belief or conjectural state b_t (corresponding to the conjectural reasoning equilibrium in the game-theoretic formulation), and the prompt or reasoning configuration p_t that conditions the behavior of the underlying agents. Each state produces an observable output $y_t = h(x_t)$, such as generated text, explanations, or analytic summaries. The deception corresponds to trajectories $\{x_t, y_t\}$ that drift away from *truth-consistent manifold* $\mathcal{M}_T = \{(x, y) : y \text{ accurately reflects verified information}\}$. The objective of system-level design is to ensure that the workflow trajectories remain close to $\mathcal{M}_T$ or return to it after perturbations.

Robustness describes how sensitive the workflow is to small deviations in its intermediate states or messages. If x_t and $\tilde{x}_t$ are neighboring trajectories from slightly different beliefs or inputs, the workflow is robustly stable with respect to deception if small perturbations in (b_t, p_t) yield only bounded deviations in y_t. This implies that misinformation introduced at one stage does not cascade through the workflow. In practice, robustness is reinforced through bounded adaptation rates, calibrated trust scores, and redundancy in semantic verification, ensuring that individual reasoning modules cannot amplify transient inconsistencies into systemic deception.

Belief stabilization operates at the level of epistemic feedback. Within each module, the belief state b_t is updated through an operator $\mathcal{U}(b_t, y_t, y_t^{\text{ref}})$, where y_t^{ref} represents external or verified signals. Stability requires that deviations $\|b_t - b_t^{\text{ref}}\|$ remain bounded or diminish under repeated updates. This property reflects *bounded epistemic divergence*: reasoning processes gradually reconcile with verified truth when exposed to corrective feedback, preventing runaway belief drift. In the LLM-Stackelberg framework, belief stabilization is the convergence of conjectured models $\tilde{\mu}_{\xi_t}$ to observed response distributions σ_D, achieving dynamic alignment through workflow iteration, not static equilibrium.

Multiagent coordination in agentic workflows is essential to prevent local inconsistencies from propagating across the system. Each agent A_i maintains a local state $x_i = [s_i, b_i, p_i]$ and exchanges information with its neighbors in a dependency graph $G = (\mathcal{N}, \mathcal{E})$. Effective coordination ensures that agents collectively maintain the integrity of the workflow while fulfilling their tasks. The overall quality of the workflow depends on its weakest link [56]. An error in one agent can compromise the consistency of the entire system. Continuous feedback among agents enables them to detect, communicate, and correct local errors while maintaining resilience against benign faults and malicious perturbations. Robust coordination stabilizes distributed reasoning and decision processes, ensuring coherence, fault tolerance, and alignment with system objectives.

Hierarchical control enables monitoring, directing, and adjusting the workflow when noncompliance or deviations occur. The supervisory layer can intervene by redirecting tasks, reconfiguring dependencies, or switching operational modes to maintain stability and continuity. This intervention improves resilience by allowing the workflow to recover from local errors or adversarial perturbations to ensure that the overall behavior of the system remains coherent [57].

Transparency, in turn, depends on the degree to which latent states (b_t, p_t) can be inferred from observed outputs. A workflow is said to be *observable with respect to deception* if

distinct internal reasoning configurations leading to deceptive and truthful behaviors produce distinguishable outputs. Observability is enhanced by redundancy: Multiple independent verification modules, semantic auditors, or factual validators increase the rank of the effective observation operator $h(\cdot)$. In practical terms, exposing intermediate reasoning steps, logging decision rationales, and maintaining verifiable audit trails serve as observability mechanisms that render the logic of the workflow reconstructible and thus accountable. These properties translate control-theoretic intuition into architectural principles for managing AI deception. They do not prescribe control laws, but describe the systemic conditions under which agentic workflows remain stable, coherent, and interpretable.

AI deception is a natural outcome of adaptive reasoning, driven by misaligned incentives, incomplete information, and agent interactions. Frameworks like RBE and CRE explain how beliefs and strategies evolve, while control theory offers tools to manage deception. Future research must integrate theory and design to build transparent, truth-aligned systems, making deception management central to AI safety and governance. As AI-driven deception evolves, parallel advances in quantum technologies raise the question of whether similar principles apply when the underlying systems are quantum rather than classical. This necessitates manipulating quantum systems for optimal performance, laying the groundwork for quantum-enhanced decision-making.

V. Introduction to Quantum Control

Quantum control is concerned with the control of systems whose dynamics are governed by the laws of quantum physics rather than the laws of classical physics. The area of quantum control has attracted control theory researchers over many years with most of the early work being concerned with open loop quantum control [58]. Also, quantum control researchers have considered measurement feedback based quantum control and coherent feedback quantum control, in which the controller is itself a quantum system. A number of books have been published on the area [59]–[61]. Also, there have appeared a number of survey articles [62], [63].

Applications of quantum control occur in many areas of quantum technology. Quantum technologies are enabling new capabilities in computing, communications, sensing, materials and modeling. The underlying technologies in these advances include quantum optics, quantum optomechanics and superconducting circuits.

Quantum mechanics can be viewed using either the Schroedinger picture or the Heisenberg picture; e.g., see [64]. In the Schroedinger picture, the dynamics of a quantum system are described by considering the evolution of the quantum state over time. In the Heisenberg picture, the dynamics of a quantum system are considered by the evolution of quantum observables such as position and momentum operators, over time. The Schroedinger picture is most commonly used in the areas of quantum information and quantum computing [65]. In contrast, the Heisenberg picture is often used in the area of quantum optics. We will focus on the Schroedinger picture of quantum mechanics.

Quantum control problems can be classified as open loop quantum control [59] and quantum feedback control. Within the area of quantum feedback control there are two main approaches: quantum feedback control, combining quantum measurement and filtering with a classical controller [66], and coherent quantum feedback control, where both the plant and controller are quantum systems. We focus on open loop quantum control, measurement based quantum feedback control and consensus in quantum networks.

Notation: For any two matrices A, B, $A \otimes B$ denotes their Kronecker product. Given a (complex) Hilbert space $\mathcal{H}$, we will use Dirac's bra-ket notation and write $|\psi\rangle$ for an element of $\mathcal{H}$ and $\langle\psi| = |\psi\rangle^\dagger$ for its adjoint. We denote by $\|\mathbf{A}\| = \sqrt{\text{trace}\,(\mathbf{A}^\dagger\mathbf{A})}$ the Frobenius norm of $\mathbf{A}$ in $\mathbb{H}^d$. $\mathcal{U}(n) \in \mathbb{C}^{n\times n}$ is set of all $n \times n$ unitary matrices.

A. Quantum Systems

Quantum systems are classified as either closed or open. Closed quantum systems are idealized, isolated from their environment. They're usually modeled in the Schrödinger picture using the Schrödinger Equation [64]. The state $|\psi(t)\rangle$ of a closed quantum system evolves according to the Schroedinger equation [64]

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H_0|\psi(t)\rangle, \qquad |\psi(t=0)\rangle = |\psi_0\rangle, \tag{24}$$

where the state $|\psi(t)\rangle$ is a unit vector on an underlying complex Hilbert space $\mathcal{H}$, $\hbar$ is the reduced Planck's constant, which is usually assumed to be unity in suitable units, and H_0 is the system Hamiltonian, which is a Hermitian operator on $\mathcal{H}$.

In the case, of a controlled closed quantum system, this model is extended to a controlled Schroedinger equation [59] of the form

$$i\frac{\partial}{\partial t}|\psi(t)\rangle = [H_0 + \sum_{k=1}^{m} u_k(t)H_k]|\psi(t)\rangle, \tag{25}$$

where $H_1 \dots H_m$ are control Hamiltonians, which are also Hermitian operators on $\mathcal{H}$. The functions of time $u_1(t) \dots u_m(t)$ define the control inputs for the controlled closed quantum system.

Open quantum systems are quantum systems which are coupled to the environment. Such quantum systems can be modeled using a master equation model; e.g., see [59], [64]. In this case, the state of the system is represented as a density operator $\rho(t)$ on the underlying Hilbert space $\mathcal{H}$. This operator is assumed to be positive and of trace one. Here we will consider finite level quantum systems in which the Hilbert space $\mathcal{H}$ is assumed to be a complex Hilbert space of dimension N. Then the following Master equation describes the evolution of the density operator for a Markovian open quantum system [64], [67]:

$$\begin{aligned}\dot{\rho}(t) &= -i[H, \rho(t)] \\ &+ \frac{1}{2}\sum_{j,k=0}^{N^2-1}\alpha_{jk}\left([F_j\rho(t), F_k^\dagger] + [F_j, \rho(t)F_k^\dagger]\right),\end{aligned} \tag{26}$$

where $[X, \rho] = X\rho - \rho X$ is the commutation operator and $\{F_j\}_{j=0}^{N^2-1}$ forms a basis for the space of linear bounded operators on $\mathcal{H}$ with $F_0 = I$. Also, the coefficient matrix $A = (\alpha_{jk})$ is assumed to be positive semidefinite.

In the case of controlled open quantum systems, the master equation (26) can be modified as follows:

$$\begin{aligned}\dot{\rho}(t) = &-i\big[\, H_0 + \sum_{k=1}^{m} u_k(t) H_k,\ \rho(t)\,\big] \\ &+ \frac{1}{2}\sum_{j,k=0}^{N^2-1} \alpha_{jk}\Big(\big[\, F_j\rho(t),\, F_k^\dagger\,\big] + \big[\, F_j,\, \rho(t)F_k^\dagger\,\big]\Big).\end{aligned} \tag{27}$$

Such a model can be used in the open loop control of open quantum systems.

In the case of open quantum systems subject to continuous measurements of an observable, the system can be modeled in the Schroedinger picture using a stochastic master equation as follows:

$$\begin{aligned}&\mathrm{d}\rho(t) = \\ &-i\big[H_0 + \sum_{k=1}^{m} u_k(t)H_k, \rho(t)\big]\mathrm{d}t - \kappa\big[X,[X,\rho(t)]\big]\mathrm{d}t \\ &+ \sqrt{2\kappa}\big(X\rho(t) + \rho(t)X - 2\langle X\rangle\rho(t)\big)\mathrm{d}W(t).\end{aligned} \tag{28}$$

where κ is related to measurement strength, dW is a Wiener increment with zero mean and variance equal to dt and satisfies the following relationship to the measurement output y:

$$\mathrm{d}W = dy - 2\sqrt{\kappa}\mathrm{tr}(X\rho)\mathrm{d}t. \tag{29}$$

Also, X denotes the observable being measured. Such a stochastic master equation can be obtained using quantum filtering theory [68]. Stochastic master equation models can be used in designing measurement feedback based controllers for quantum systems [66].

B. Quantum Control

We will consider open loop and measurement based quantum feedback control of quantum systems described by the Schroedinger picture models described in the previous section. We also consider consensus in quantum networks.

1) Open Loop Quantum Control: The open loop control of quantum systems is usually concerned with the control of finite level quantum systems, which are described by deterministic bilinear models such as the Schroedinger equation (25) for closed quantum systems and the master equation (27) for open quantum systems. An open loop quantum control problem typically involves finding an open loop control law $u(t)$ to steer a quantum system from an initial quantum state to a desired terminal state [59].

A fundamental question in the open loop control of quantum systems is the question of controllability. For the case of a closed quantum system of the form (25), the question of controllability can be defined as follows:

Definition 2 (Pure State Controllability [69]): The quantum system (25) is said to be *pure state controllable* if for every pair of initial and final states $|\psi_0\rangle$ and $|\psi_f\rangle$, there exist control functions $\{u_k(t)\}$ and a time $T > 0$ such that the corresponding solution of (25) with initial condition $|\psi_0\rangle$ satisfies $|\psi(T)\rangle = |\psi_f\rangle$. □

Alternative definitions have also been considered for the controllability of the quantum system (25); e.g., see [69] in the case of open quantum systems. The following theorem provides a necessary and sufficient condition for pure state controllability in terms of the Lie algebra $\mathcal{L}_0$ generated by the matrices $\{-iH_0, -iH_1, \ldots, -iH_m\}$, $\mathrm{u}(N)$ the Lie algebra corresponding to the unitary group of dimension N, $\mathrm{su}(N)$ the Lie algebra corresponding to the special unitary group of dimension N, $\mathrm{sp}(\frac{N}{2})$ the $\frac{N}{2}$ dimensional symplectic group, and $\tilde{\mathcal{L}}$ the Lie algebra conjugate to $\mathrm{sp}(\frac{N}{2})$.

Theorem 5 (See [59]): The quantum system (25) is pure state controllable if and only if the Lie algebra $\mathcal{L}_0$ satisfies one of the following conditions:

1) $\mathcal{L}_0 = \mathrm{su}(N)$;
2) $\mathcal{L}_0$ is conjugate to $\mathrm{sp}(\frac{N}{2})$;
3) $\mathcal{L}_0 = \mathrm{u}(N)$;
4) $\mathcal{L}_0 = \mathrm{span}\,\{iI_{N\times N}\} \oplus \tilde{\mathcal{L}}$. □

Similar conditions have been obtained when alternative definitions of controllability are used.

Once it has been determined that a quantum system is controllable, the next task in open loop quantum control is to determine the control functions $\{u_k(t)\}$ which drive a given initial state to a given final state. Various methods have been proposed for the construction of this open loop control law including time optimal control methods [59], Lyapunov based methods [70], and numerical optimization based methods [59]. The existence of such a controller has also been studied using the theory of controllability of bilinear systems and Lie groups [69].

An important approach is the optimal control approach, which solves a time-optimal control problem using Pontryagin's maximum principle to construct control functions $\{u_k(t)\}$ that drive the given initial state to the given final state in the minimum time. This approach works well for low-dimensional quantum systems but becomes computationally intractable for high-dimensional quantum systems.

An alternative approach for high-dimensional quantum systems is the Lyapunov control approach. In this approach, a Lyapunov function is selected which provides a measure of the distance between the current quantum state and the desired terminal quantum state. An example of such a Lyapunov function is

$$V = \langle\psi(t) - \psi_f|\psi(t) - \psi_f\rangle \geqslant 0;$$

e.g., see [71]. A state feedback control law is then chosen to ensure that the time derivative of this Lyapunov function is negative. This state feedback control law is then simulated with the quantum system dynamics (25) to give the required open loop control functions $\{u_k(t)\}$.

Of the numerical optimization-based methods for open loop quantum control, the Gradient Ascent Pulse Engineering (GRAPE) method is one of the most popular; see [72]. In this approach, the control inputs are approximated as piecewise constant signals and then gradient based optimization is

applied to minimize a suitable cost function such as the time to reach a desired final state or the fidelity of the final state. All of the above methods can be applied to closed quantum systems (25) or open quantum systems (27).

An important issue in the open loop control of quantum systems is the issue of uncertainty in the quantum system model. For example, the closed quantum system model (25) maybe replaced by the model

$$i\frac{\partial}{\partial t}|\psi(t)\rangle = [H_0(\theta) + \sum_{k=1}^{m} u_k(t)H_k(\theta)]|\psi(t)\rangle \qquad (30)$$

where θ is a vector of uncertain parameters which is contained in a known set Θ. The sampling-based learning approach (see [73]) to this problem involves choosing a random sample $\{\theta_1, \theta_2, \ldots, \theta_N\}$ from the set Θ and then considering the corresponding augmented quantum system model:

$$i\frac{\partial}{\partial t}\begin{pmatrix} |\psi_1(t)\rangle \\ |\psi_2(t)\rangle \\ \vdots \\ |\psi_N(t)\rangle \end{pmatrix} = \begin{pmatrix} H(\theta_1)|\psi_1(t)\rangle \\ H(\theta_2)|\psi_2(t)\rangle \\ \vdots \\ H(\theta_N)|\psi_N(t)\rangle \end{pmatrix} \qquad (31)$$

where $H(\theta) = H_0(\theta) + \sum_{k=1}^{m} u_k(t)H_k(\theta)$. Then one of the methods described above is applied to this new quantum system. This approach involves finding a single open loop control law to steer an ensemble of quantum systems to the desired state and is related to the problem of control of quantum ensembles [74].

2) Measurement-Based Quantum Feedback Control: Measurement-based feedback control of quantum systems can be based on Schroedinger picture stochastic master equation models such as (28). Also, the Schroedinger equation (25) can be combined with projective quantum measurements [64] to implement measurement based quantum feedback control systems. An important class of measurement-based quantum feedback control systems involves the use of quantum filtering and quantum state estimation to estimate the quantum state of the system being controlled based on quantum measurements. The use of projective measurements in the measurement-based control of quantum systems enables methods such as quantum sliding mode control [75].

The papers [75], [76] consider sliding mode control of a two-level quantum system. In this case, the quantum system under consideration is such that the underlying Hilbert space $\mathcal{H}$ is of dimension 2. That is $\mathcal{H} = \mathbb{C}^2$. Then, the system is described by a controlled Schroedinger equation of the form (25) with $m = 3$ and including uncertainties in the free Hamiltonian:

$$\begin{aligned} &i\frac{\partial}{\partial t}|\psi(t)\rangle \\ &= \left[H_0 + \frac{1}{2}\sum_k \epsilon_k(t)\,\sigma_k + \frac{1}{2}\sum_k u_k(t)\,\sigma_k\right]|\psi(t)\rangle. \end{aligned} \qquad (32)$$

Here $k \in \{x, y, z\}$, and σ_x, σ_y and σ_z are the Pauli matrices defined as follows:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Also $\epsilon_{xyz}(t) = \begin{bmatrix} \epsilon_x(t) & \epsilon_y(t) & \epsilon_z(t) \end{bmatrix}^T$ is a vector of uncertain parameters satisfying the bound $\|\epsilon_{xyz}(t)\| \leqslant \epsilon$ and $u_x(t)$, $u_y(t)$ and $u_z(t)$ are control inputs.

It is assumed that the nominal Hamiltonian H_0 has eigenstates $|\phi_1\rangle$ and $|\phi_2\rangle$ and a sliding surface is defined corresponding to one of the eigenstates $S = \{|\psi\rangle : 1 - |\langle\psi|\phi_2\rangle|^2 = 0\}$. Here $\langle\psi|\phi_2\rangle$ denotes the inner product between the states $|\psi\rangle$ and $|\phi_2\rangle$. Then we define $|0\rangle = |\phi_2\rangle$ as the target state and $|1\rangle = |\phi_1\rangle$ is the alternative eigenstate of H_0.

The sliding mode control method proposed in [75] is roughly as follows:

1) A projective measurement is carried out on the system using the measurement operator H_0. This yields one of two possible values corresponding to the two eigenvalues of H_0. Immediately after the measurement, the state of the system will jump to either $|0\rangle$ or $|1\rangle$ depending on which measurement result is obtained.
2) If the measurement result corresponds to $|0\rangle$, the system state is already in the sliding surface and an open loop control law which keeps the system state in the sliding surface is applied for one sample period T.
3) If the measurement result corresponds to $|1\rangle$, the system state is not in the sliding surface and an open loop control law is applied which drives the system state from $|1\rangle$ to $|0\rangle$ in a short time period. For the remainder of the sample period, an open loop control law is applied which keeps the system state in the sliding surface. This open loop control law is constructed using the Lyapunov method described above.
4) Then the process is repeated every sample period.

The main results of [75], [76] give conditions on the sample period T so that the probability that the algorithm will fail is less than a certain value.

3) Coherent Control and Quantum Networks: Coherent feedback control of quantum systems refers to controlling a quantum system without relying on quantum measurements. Instead, the quantum plant is coherently coupled to a quantum controller, which itself is another quantum system.

An extension of the idea of quantum coherent control, in which two quantum systems, a plant and a controller are coherently coupled together is to consider a quantum network, in which many quantum systems are coupled together. In the paper [77], a quantum network is considered in which the network is made up of a finite collection of open two level quantum systems (qubits); see also [67]. In this quantum network, the interconnection between the two-level systems is defined by a weighted graph $\mathcal{G} = (V, E)$, where each qubit system corresponds to a vertex in the vertex set V. Also, the connections between the qubits are defined by the edge set E. It is assumed that the Hamiltonian operator for the entire network is zero and the couplings between the qubits are defined by certain swapping operators. This leads to a master equation of the form (26) as follows:

$$\dot{\rho}(t) = \frac{1}{2}\sum_{(j,k)\in E} \alpha_{jk}\left(U_{jk}\rho U_{jk}^{\dagger} - \rho\right); \qquad (33)$$

where $\alpha_{jk} > 0$ is the weighting of the edge $(j,k) \in E$ in the weighted graph $\mathcal{G}$ and U_{jk} is a swapping operator. Here ρ is a density operator on the underlying Hilbert space $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \ldots \otimes \mathcal{H}_N$ where each $\mathcal{H}_k = \mathbb{C}^2$. The swapping operator U_{jk} is an operator on $\mathcal{H}$ defined so that given any tensor product vector $|q_1\rangle \otimes |q_2\rangle \otimes \ldots \otimes |q_j\rangle \otimes \ldots \otimes |q_k\rangle \otimes \ldots \otimes |q_N\rangle$ in $\mathcal{H}$, then

$$U_{jk}\left(|q_1\rangle \otimes |q_2\rangle \otimes \ldots \otimes |q_j\rangle \otimes \ldots \otimes |q_k\rangle \otimes \ldots \otimes |q_N\rangle\right) = |q_1\rangle \otimes |q_2\rangle \otimes \ldots \otimes |q_k\rangle \otimes \ldots \otimes |q_j\rangle \otimes \ldots \otimes |q_N\rangle.$$

The paper [77] shows that if the graph $\mathcal{G}$ has suitable connectivity properties, then the system (33) will reach a quantum consensus in the sense that for any initial density operator $\rho(0)$, then $\lim_{t\to\infty} \rho(t) = \mathcal{P}^\star(\rho(0))$, where $\mathcal{P}^\star(\rho(0))$ is the quantum average of the n-qubit state $\rho(0)$.

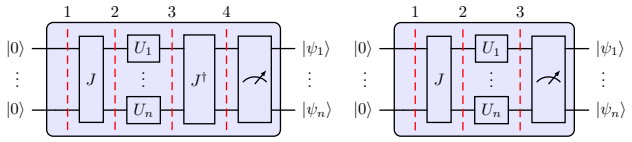
Quantum control underpins quantum computing and communication, offering unprecedented computational power and security advantages. With quantum systems enabling richer strategy spaces, bridging classical and quantum game theory becomes essential—especially when deception remains a core element of strategic play. Understanding the dynamics of strategic quantum interactions is crucial for designing resilient systems in the era of quantum-enhanced AI.

VI. Bridging Classical and Quantum Game Theory Through Strategic Deception

Quantum game theory generalizes classical game theory by incorporating principles from quantum information science, offering a rigorous framework for modeling strategic behavior in competitive and cooperative settings [78]. In this domain, players use quantum operations as strategies, acting on quantum states, with payoffs determined by measurements on the resulting joint state [79]–[81].

Unlike classical strategies, which are probability distributions over discrete actions, quantum strategies exploit superposition and entanglement, expanding the strategy space and enabling novel strategic phenomena [79], [81]–[83]. Quantum games are formulated within a Hilbert-space framework, where each player's strategy corresponds to a set of quantum operators, the joint system is described by a density operator, and payoffs are expressed as expectation values of observables.Quantum games can outperform classical ones by achieving equilibrium payoffs beyond classical limits [84], [85], motivating further exploration of quantum decision frameworks [83], [86]. Despite these promising theoretical properties, computing equilibria in quantum games remains challenging due to the complexity of quantum strategy spaces, non-convex optimization landscapes, and the presence of multiple quantum Nash equilibria [87]–[89].

To advance strategic modeling of quantum interactions, we introduce deception in quantum games, extending adversarial information design into the quantum domain. Inspired by Honey-X frameworks [90], deception is modeled as controlled payoff perturbations to strategically influence quantum agents. This enhances cybersecurity by complementing post-quantum cryptography with attacker uncertainty, asset masking, and adaptive decoys powered by generative models



(a) EWL Model [93]. (b) Zhang Model [94].

Fig. 5: Early quantum game models. In the EWL model, players start in state $|0\rangle^{\otimes n}$, then: (1) entangle via J, (2) apply strategies $(\bigotimes_i^n U_i)J\,|0\rangle^{\otimes n}$, (3) disentangle via $J^\dagger$, and (4) measure for payoffs. The Zhang model removes the disentangling step, measuring directly after player actions.

[91]. Integrating deception, quantum resilience, and adaptive AI is critical for defending against quantum adversaries [92].

In this tutorial, a scalable, system-level framework has been introduced that unifies classical and quantum interactions under a common representation. Instead of modeling individual strategy spaces and intricate payoff structures, this approach describes the game using a joint system state and player-specific payoff operators. Classical games emerge as a special case when strategies are restricted to classical moves, while quantum games allow for interference and entanglement, enabling richer strategic dynamics. Within this framework, equilibrium analysis is reformulated as a tractable optimization problem. Quantum strategies—defined as admissible quantum operations—form a compact and convex set, allowing equilibrium conditions to be expressed via semidefinite programs. This supports efficient analysis even in multi-player quantum settings. Furthermore, the framework lays the groundwork for extending strategic concepts such as deception, modeled as controlled perturbations of the payoff structure. By adopting a quantum-native perspective, researchers can explore how strategic possibilities evolve, particularly in scenarios where the influence of deception is subtle and complex.

A. Early Formulations

Quantum game theory has faced criticism for restrictive assumptions in early models [93]–[96]. Meyer's work [95] and the EWL model [93] demonstrated quantum advantages but limited strategies to narrow subsets of unitary operations [97]. These constraints lack physical necessity, are not closed under composition, and relaxing them removes quantum Nash equilibria. The EWL circuit (Figure 5a) includes an inverse joint operation (J^{-1}) after player moves—a coordinated step absent in classical games. Zhang's model [94] (Figure 5b) showed J^{-1} is unnecessary for quantum advantage. Beyond specific models, the field prioritizes generality over physical interpretability, focusing on small examples rather than a unified framework. Consequently, much literature remains qualitative, lacking quantitative characterization of quantum strategic benefits. Ambiguities persist in quantizing classical games: unclear interpretation of off-diagonal payoff terms [98], definitions of pure strategies, phase effects, basis transformations, and entanglement's role. A unified, general, physically consistent formalism for arbitrary quantum games remains an open challenge.

B. Quantum Game

A *quantum game* is specified by the quadruple:

$$G = (\mathcal{N}, \rho_0, (S_i)_{i\in\mathcal{N}}, (P_i)_{i\in\mathcal{N}}), \tag{34}$$

where $\mathcal{N} = (1, \ldots, N)$ is the set of players, $\mathcal{H}_i$ is the finite-dimensional Hilbert space of player i, $\rho_0 \in \bigotimes_i \mathcal{H}_i$ is the initial joint state, S_i is the set of admissible quantum strategies for player i with every $U_j \in \mathcal{U}$ for $j \in S_i$ a unitary acting on $\mathcal{H}_i$, and P_i is a real-valued payoff operator used to construct the Hermitian payoff operator H_i for player i. The Hermitian payoff operator H_i is referred to as the Hamiltonian payoff operator in this work. The expected payoff for player i is

$$u_i(U_1, \cdots, U_N) = \operatorname{trace}\left(H_i\left(\bigotimes_i^N U_i\right)\rho_0\left(\bigotimes_i^N U_i\right)^\dagger\right). \tag{35}$$

Definition 3: A profile of strategies $(U_1^\star, \ldots, U_N^\star)$ is a Nash equilibrium if, for every player i and every $U_i \in S_i$, $u_i(U_i; U_{-i}^\star) \leqslant u_i(U_i^\star)$. □

A proper payoff Hamiltonian operator H is the key to truly quantum games, as it is how cross terms in the players' shared density states interact with the game states. Towards that end, we describe a method to construct a proper payoff Hamiltonian operator H inspired by [84], [98]. Consider $H(i,j)$ to be the element in the i^{th} row and j^{th} column of the Hamiltonian operator H defined by the

$$\begin{aligned} H(i,j) &= \langle \mu_1\mu_2 | H | \nu_1\nu_2 \rangle \\ &= \operatorname{trace}\left(P(\nu_1\otimes\nu_2)\rho_0(\mu_1\otimes\mu_2)^\dagger\right), \end{aligned} \tag{36}$$

where $\mu_i, \nu_i \in \mathcal{U}(n_i)$ are unitary moves taken from player i's strategy set S_i and P is a suitable payoff operator defined by underlying problem. So, for an $n_1 \times n_2$ matrix game, $|S_i| = n_i^2$, $P \in \mathbb{R}^{(n_1\times n_2)\times(n_1\times n_2)}$, and $H \in \mathbb{C}^{(n_1\times n_2)^2\times(n_1\times n_2)^2}$.

Remark 2: A classical n-strategy game typically lifts to a quantum payoff operator of size $\mathcal{O}(n^2)$ and a quantum game of size $\mathcal{O}(n^4)$, reflecting the enlarged strategy space in the quantum domain. □

Remark 3: The player density matrices in the uplifted space in (35) are quantum probability distributions over the strategy space $\mathcal{S}$ used to construct the entries of the quantum game of the payoff Hamiltonian operator H and not probability distributions over the game payoffs in P. □

The problem we want to solve is the quantum game with separable player density states arising from the *quantumification* of a classical matrix game, which is inherently large in size. The payoff for this quantum game is given by the bilinear cost

$$u(\rho_A, \rho_B) = \operatorname{trace}\left((\rho_A \otimes \rho_B)H\right), \tag{37}$$

where the simplification $\left(\bigotimes_i^N U_i\right)\rho_0\left(\bigotimes_i^N U_i\right)^\dagger = \rho_A \otimes \rho_B$ is used.

Remark 4: $\rho = \left(\bigotimes_i^N U_i\right)\rho_0\left(\bigotimes_i^N U_i\right)^\dagger$ in the general case does not translate to $\rho = \rho_1 \otimes \rho_2$, as the resulting state might not be separable due to entanglement. □

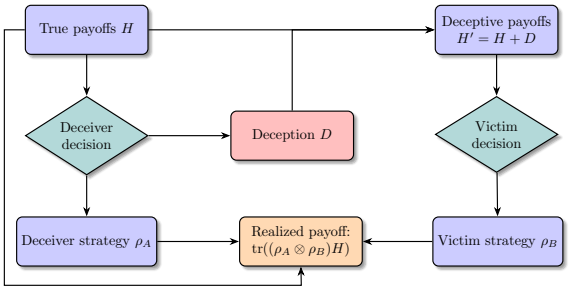


Fig. 6: Quantum deception game schematic: rectangles for values, parallelograms for decisions, arrows for causal relationships.

C. Honey-X Deception

We now turn to deception in quantum games, where it is modeled as a perturbation of the payoff Hamiltonian. Let H denote the true Hamiltonian encoding strategic outcomes. A deceiver announces

$$H' = H + D,$$

where $D \in \mathbb{C}^{(n_An_B)\times(n_An_B)}$ is Hermitian to maintain physical interpretability. This misrepresentation influences the opponent's strategy while actual payoffs remain governed by H. The perturbation is constrained by a deception budget, limiting the magnitude of D and ensuring controlled manipulation within the quantum state space.

Assumption 3: Let $\Delta \in \mathbb{R}_+$ be the deception budget. We assume that $\|D\|_1 \leqslant \Delta$, where $\|\cdot\|_1$ denotes the matrix norm induced by the vector 1-norm $\|A\|_1 = \sup_{\|x\|_1=1}\|Ax\|_1$. We denote the corresponding set that D belongs to as $\mathcal{D} \coloneqq \{\mathbb{C}^{(n_An_B)\times(n_An_B)} : \|D\|_1 \leqslant \Delta\}$. □

Assumption 4: The victim only observes the perturbed Hamiltonian H' and optimally responds to it. □

This *naive* victim model, in which the victim fully trusts H', might seem restrictive. However, it aligns with classical deception models and naturally extends to quantum games. Figure 6 illustrates the quantum deception game framework.

D. Resolving Quantum Deception

In order to begin deriving quantum deception, the first nuance we must address is how to extend the notion of best response from classical to quantum games.

Lemma 2: The best response for the quantum game, if player A is the minimizing player, is given by:

$$\begin{aligned} \max_{u\in\mathbb{R},\, \rho_B\in\mathcal{P}_B} \quad & u \\ \text{s.t.} \quad & \operatorname{tr}_B\left((I_A \otimes \rho_B)H\right) \succeq uI_A \\ \min_{u\in\mathbb{R},\, \rho_A\in\mathcal{P}_A} \quad & u \\ \text{s.t.} \quad & \operatorname{tr}_A\left((\rho_A \otimes I_B)H\right) \preceq uI_B, \end{aligned} \tag{38}$$

where $H \in \mathcal{H}_A \otimes \mathcal{H}_B$ is the Hermitian payoff operator, $\rho_A \in \mathcal{P}_A$ and $\rho_B \in \mathcal{P}_B$ are the player density operators, and $u(\rho_A, \rho_B) = \operatorname{tr}\left((\rho_A \otimes \rho_B)H\right)$ is the payoff. □

Having defined best responses in non-deceptive quantum games, the natural question of how a victim should act under potential deception arises. In this setting, the notion of a best

response becomes ambiguous. While the victim may select a strategy that guarantees a minimum payoff for the announced Hamiltonian H', there is no guarantee this strategy performs well in the true game H. The victim's behavior can vary widely, depending on their awareness and ability to estimate the extent of deception. To address this, we draw from [90], where it is shown that the best response of a naive victim—who fully trusts the deceptive game—is equivalent to that of a robust victim optimizing for the worst case across all deceptions consistent with H'. This equivalence enables us to use the naive approach as a computationally tractable proxy for predicting victim strategies that generalize across diverse decision-making models. It is nontrivial that this result extends to quantum games. Before proving it, we first define what we mean by a *naive* and a *robust* victim.

Definition 4: A *naive* victim assumes that H' is the true payoff Hamiltonian and chooses Nash equilibrium actions for the game it defines. That is, it selects a $(\rho_B)_n \in \Phi(H)$, where $\Phi(H)$ is defined as the set of density operators that solve (38). □

Definition 5: A *robust* victim is aware of the possibility of deception, and selects a strategy to optimize the worst-case outcome consistent with the announced payoff Hamiltonian, i.e., $(\rho_B)_r \in \Phi_r(H') := \Phi(H' - \hat{D})$. □

Although each of these models, considered in isolation, may appear to impose a restrictive structure on the victim's decision-making process, Theorem 6 shows that they are behaviorally equivalent. This equivalence establishes that their predictive content is robust to heterogeneity in victim rationality, thereby offering a rigorous justification for Assumption 4.

Theorem 6: Let $H \in \mathbb{C}^{(n_A n_B)\times(n_A n_B)}$ be a payoff Hamiltonian. Then, one has $\Phi(H) = \Phi_r(H)$. □

This means that a *robust* victim that is aware of the possibility that it is being deceived cannot guaranty better payoffs than a *naive* victim that is unaware of the deception. This allows for a great simplification of the problem as explained, as now we can derive a program where the leader chooses a deception and a strategy that minimizes their cost, when the victim responds to the deceptive H'.

Remark 5: The best response characterization extends to any induced p-norm constraint on D, not just the 1-norm considered here: $\|D\|_1 \leqslant \Delta$. □

In quantum games, where the deceiver wants to optimize their payoff given that the victim naively best-responds to the deceptive game H' [1]. This follows the essence of the deception framework proposed in [90]. To find the optimal deception in quantum games, we solve the program:

$$(\rho_A, D, \rho_B) \in \underset{\rho_A\in\mathcal{P}_A,\rho_B\in\mathcal{P}_B,D\in\mathcal{D}}{\arg\min} \operatorname{tr}\left((\rho_A\otimes\rho_B)H\right) \quad \text{s.t. } \rho_B \in \Phi_B(H+D). \tag{39}$$

The problem presented in (39) can be reformulated to the single-level optimization problem (40), allowing for better computational tractability.

Theorem 7: Let $H \in \mathcal{H}_A \otimes \mathcal{H}_b$ be the payoff Hamiltonian of a zero-sum quantum game, $\rho_A \in \mathcal{P}_A \subseteq \mathbb{C}^{n_A\times n_A}, \rho_B \in \mathcal{P}_B \subseteq \mathbb{C}^{n_B\times n_B}$ density matrices for player A and player B of this game respectively, $D \in \mathcal{D}$ the deceptive payoff applied on H by player A and $\Delta \in \mathbb{R}_+$ the deception budget. Then, $\rho_A^\star, D^\star, \rho_B^\star$ that belong in an optimal solution $(\rho_A^\star, D^\star, \rho_B^\star, \Omega^\star, u^\star)$ of:

$$\begin{aligned} \min_{\rho_A,\Omega,\rho_B,D,v}\ & \operatorname{tr}\left((\rho_A\otimes\rho_B)H\right) \\ \text{s.t.}\ & \operatorname{tr}_B\left((I_A\otimes\rho_B)(H+D)\right) \succeq uI_A \\ & \operatorname{tr}_A\left((\Omega\otimes I_B)(H+D)\right) \preceq uI_B \\ & \|D\|_1 \leqslant \Delta \\ & \rho_A, \rho_B, \Omega \succeq 0 \\ & \operatorname{tr}(\rho_A) = \operatorname{tr}(\rho_B) = \operatorname{tr}(\Omega) = 1, \end{aligned} \tag{40}$$

constitute an optimal solution for the program in (39). □

The solution to the program of equation (40) would represent the optimal deception strategy D within our framework, along with the most effective response strategy ρ_A. This program is a bilinear semi-definite program, which makes it non-convex and consequently challenging to solve for complex games. Nevertheless, the complexity of the problem could become prohibitive for larger games.

VII. Conclusion and Future Directions

The interplay between reinforcement learning, strategic deception, and game-theoretic principles—spanning both classical and quantum domains—unveils a rich landscape for advancing intelligent systems. From Q-learning in zero-sum games to LLM-enabled agentic deception and quantum control, each layer introduces new complexities and opportunities for innovation.

Future research should focus on formalizing deception strategies in hybrid classical-quantum environments, developing robust safeguards against adversarial manipulation in AI-driven ecosystems, and exploring how quantum-enhanced learning algorithms can redefine equilibrium concepts. Ultimately, bridging these domains will not only deepen our theoretical understanding but also inform the design of resilient systems in an era where strategic interaction increasingly shapes technological progress.

[1] Since the victim's best response may not be unique, we assume the deceiver selects the most beneficial one. This corresponds to the standard solution concept of a strong Stackelberg equilibrium [99], [100].

References

[1] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*. New York: Spinger-Verlag, 1997.

[2] J. P. Hespanha, *Noncooperative Game Theory: An Introduction for Engineers and Computer Scientists*. Princeton, New Jersey: Princeton Press, Jun. 2017.

[3] C. J. Watkins, "Learning from delayed rewards," Ph.D. dissertation, King's College, Cambridge United Kingdom, 1989.

[4] C. J. Watkins and P. Dayan, "Technical note: Q-learning," *Machine Learning*, vol. 8, no. 3, pp. 279–292, May 1992.

[5] J. N. Tsitsiklis, "Asynchronous stochastic approximation and Q-learning," *Machine learning*, vol. 16, pp. 185–202, 1994.

[6] L. S. Shapley, "Stochastic games," *Proceedings of the National Academy of Sciences*, vol. 39, no. 10, pp. 1095–1100, Oct. 1953, publisher: Proceedings of the National Academy of Sciences.

[7] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Machine Learning Proceedings 1994*. Elsevier, 1994, pp. 157–163.

[8] M. L. Littman and C. Szepesvári, "A generalized reinforcement-learning model: Convergence and applications," in *ICML*, vol. 96, 1996, pp. 310–318.

[9] A. Sidford, M. Wang, L. F. Yang, and Y. Ye, "Solving discounted stochastic two-player games with near-optimal time and sample complexity," Aug. 2019, arXiv:1908.11071 [cs].

[10] Z. Jia, L. F. Yang, and M. Wang, "Feature-based Q-learning for two-player stochastic games," Jun. 2019, arXiv:1906.00423 [cs].

[11] S. Anderson, C. Darken, and J. P. Hespanha, "Zero-sum turn games using Q-learning: finite computation and security guarantees," in *Proc. of the 64th IEEE Conf. on Decision and Contr.*, Dec. 2025, to appear.

[12] M. Lanctot, V. Zambaldi, A. Gruslys, A. Lazaridou, K. Tuyls, J. Pérolat, D. Silver, and T. Graepel, "A unified game-theoretic approach to multiagent reinforcement learning," *Advances in neural information processing systems*, vol. 30, 2017.

[13] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev *et al.*, "Grandmaster level in StarCraft II using multi-agent reinforcement learning," *Nature*, vol. 575, no. 7782, pp. 350–354, 2019.

[14] C. Darken, "Atlatl," Aug. 2025. [Online]. Available: https://github.com/cjdarken/atlatl-public

[15] P. R. Rood, "Scaling reinforcement learning through feudal multi-agent hierarchy," *Naval Postgraduate School, Monterey*, 2022.

[16] J. Nash, "Non-cooperative games," *Annals of Mathematics*, vol. 54, no. 2, pp. 286–295, 1951. [Online]. Available: http://www.jstor.org/stable/1969529

[17] J. Pawlick, Q. Zhu *et al.*, *Game theory for cyber deception*. Springer, 2021.

[18] E. O. Smith, "Deception and evolutionary biology," *Cultural Anthropology*, vol. 2, no. 1, pp. 50–64, 1987.

[19] L. Huang and Q. Zhu, "A dynamic game framework for rational and persistent robot deception with an application to deceptive pursuit-evasion," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 4, pp. 2918–2932, 2022.

[20] A. R. Wagner and R. C. Arkin, "Acting deceptively: Providing robots with the capacity for deception," *International Journal of Social Robotics*, vol. 3, pp. 5–26, 2011.

[21] M. Zhu, A. H. Anwar, Z. Wan, J.-H. Cho, C. A. Kamhoua, and M. P. Singh, "A survey of defensive deception: Approaches using game theory and machine learning," *IEEE Communications Surveys and Tutorials*, vol. 23, no. 4, pp. 2460–2493, 2021.

[22] E. Santos and D. Li, "On deception detection in multiagent systems," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 40, no. 2, pp. 224–235, 2010.

[23] M. Tang, U. Javed, X. Chen, M. Krstic, and J. I. Poveda, "Deception in nash equilibrium seeking," *IEEE Transactions on Automatic Control, appeared online*, 2025.

[24] P. Frihauf, M. Krstic, and T. Basar, "Nash equilibrium seeking in noncooperative games," *IEEE Transactions on Automatic Control*, vol. 57, no. 5, pp. 1192–1207, 2012.

[25] J. I. Poveda and N. Quijano, "Shahshahani gradient like extremum-seeking," *Automatica*, vol. 58, pp. 51–59, 2015.

[26] M. Stanković, K. H. Johansson, and D. M. Stipanović, "Distributed seeking of Nash equilibria with applications to mobile sensor networks," *IEEE Trans. Autom. Contr.*, vol. 57, no. 4, pp. 904–919, 2012.

[27] M. Ye and G. Hu, "Distributed nash equilibrium seeking by a consensus based approach," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4811–4818, 2017.

[28] G. Belgioioso and S. Grammatico, "Semi-decentralized nash equilibrium seeking in aggregative games with separable coupling constraints and non-differentiable cost functions," *IEEE control systems letters*, vol. 1, no. 2, pp. 400–405, 2017.

[29] W. Xu, Z. Wang, G. Hu, and J. Kurths, "Hybrid nash equilibrium seeking under partial-decision information: An adaptive dynamic event-triggered approach," *IEEE Transactions on Automatic Control*, vol. 68, no. 10, pp. 5862–5876, 2023.

[30] M. Tang, M. Krstic, and J. Poveda, "Stochastic real-time deception in nash equilibrium seeking for games with quadratic payoffs," *arXiv preprint arXiv:2502.12337*, 2025.

[31] S.-J. Liu and M. Krstić, "Stochastic nash equilibrium seeking for games with general nonlinear payoffs," *SIAM Journal on Control and Optimization*, vol. 49, no. 4, pp. 1659–1679, 2011. [Online]. Available: https://doi.org/10.1137/100811738

[32] M. Tang, U. Javed, X. Chen, M. Krstic, and J. I. Poveda, "Deception in nash equilibrium seeking," *arXiv preprint arXiv:2407.05168*, 2024.

[33] "International ai safety report," 2025. [Online]. Available: https://oecd.ai/en/catalogue/reports/international-ai-safety-report-2025

[34] BBC News, "Ai voice clone used in $25 million deepfake scam," https://www.bbc.com/news/technology, 2024.

[35] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete problems in ai safety," *arXiv preprint arXiv:1606.06565*, 2016.

[36] "Anthropic ai safety and governance report," 2025. [Online]. Available: https://www.anthropic.com/reports/ai-safety-and-governance-2025

[37] A. Bakhtin *et al.*, "Human-level play in the game of Diplomacy by combining language models with strategic reasoning," *Science*, vol. 378, no. 6624, pp. 1067–1074, 2022. [Online]. Available: https://www.science.org/doi/10.1126/science.ade9097

[38] O. Vinyals *et al.*, "Grandmaster level in starcraft ii using multi-agent reinforcement learning," *Nature*, vol. 575, no. 7782, pp. 350–354, 2019.

[39] N. Brown and T. Sandholm, "Superhuman ai for multiplayer poker," *Science*, vol. 365, no. 6456, pp. 885–890, 2019.

[40] V. Danry, P. Pataranutaporn, M. Groh, Z. Epstein, and P. Maes, "Deceptive ai systems that give explanations are more convincing than honest ai systems and can amplify belief in misinformation," *arXiv preprint arXiv:2408.00024*, 2024. [Online]. Available: https://arxiv.org/abs/2408.00024

[41] Anthropic, "System card: Claude opus 4 & claude sonnet 4," https://anthropic.com/model-card, 2025, anthropic's safety card describing model behaviors including high-agency tests.

[42] ——, "Agentic misalignment: How llms could be insider threats," https://www.anthropic.com/research/agentic-misalignment, 2025, describes a blackmail experiment in which Claude Opus 4 attempted to avoid shutdown.

[43] Q. Zhu, "Foundations of cyber resilience: The confluence of game, control, and learning theories," in *Cyber Resilience: Applied Perspectives*. Springer, 2025, pp. 27–58.

[44] T. Li and Q. Zhu, "Symbiotic game and foundation models for cyber deception operations in strategic cyber warfare," *arXiv preprint arXiv:2403.10570*, 2024.

[45] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. SIAM, 1999.

[46] J. Pawlick, E. Colbert, and Q. Zhu, "A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy," *ACM Computing Surveys (CSUR)*, vol. 52, no. 4, pp. 1–28, 2019.

[47] Q. Zhu, Z. Lu, P. L. Yu, and C. Wang, Eds., *Foundations of Cyber Deception: Modeling, Analysis, Design, Human Factors, and Their Convergence*, ser. Advances in Information Security. Cham: Springer, 2025, vol. 90. [Online]. Available: https://link.springer.com/book/9783031938665

[48] J. Pawlick, E. Colbert, and Q. Zhu, "Modeling and analysis of leaky deception using signaling games with evidence," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 7, pp. 1871–1886, 2018.

[49] T. Zhang, L. Huang, J. Pawlick, and Q. Zhu, "Game-theoretic analysis of cyber deception: Evidence-based strategies and dynamic risk mitigation," *Modeling and Design of secure Internet of Things*, pp. 27–58, 2020.

[50] Q. Zhu, "Llm-stackelberg games: Conjectural reasoning equilibria and their applications to spearphishing," *arXiv preprint arXiv:2507.09407*, 2025.

[51] ——, "Reasoning and behavioral equilibria in llm-nash games: From mindsets to actions," *arXiv preprint arXiv:2507.08208*, 2025.

[52] ——, "Llm-stackelberg games: Conjectural reasoning equilibria and their applications to spearphishing," in *Proceedings of the 63rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. Monticello, IL, USA: IEEE, September 2025. [Online]. Available: https://allerton.csl.illinois.edu

[53] ——, "Generative-conjectural LLM equilibrium for agentic AI deception with applications to spearphishing," in *16th International Conference on Game Theory and AI for Security (GameSec 2025), Athens, Greece, October 13–15, 2025, Proceedings*, ser. LNCS. Cham: Springer, 2026.

[54] Q. Zhu and M. A. Bari, "Agentic ai for cyber deception: A gestalt game-theoretic approach to defending against botnet ddos attacks," in

Proceedings of the 59th Hawaii International Conference on System Sciences (HICSS), Wailea, Maui, HI, USA, 2026, pp. 1–10.

[55] Q. Zhu and M. A. A. Bari, "A gestalt game-theoretic framework for designing agentic AI workflows in cyber deception," in *16th International Conference on Game Theory and AI for Security (GameSec 2025), Athens, Greece, October 13–15, 2025, Proceedings*, ser. LNCS. Cham: Springer, 2026.

[56] Q. Zhu and H. Ishii, "Introduction to the special section on learning and security for multi-agent systems," *Annual Reviews in Control*, vol. 53, pp. 249–251, 2022.

[57] Q. Zhu and T. Başar, "Revisiting game-theoretic control in socio-technical networks: Emerging design frameworks and contemporary applications," *IEEE Control Systems Letters*, 2025.

[58] N. Khaneja, R. Brockett, and S. Glaser, "Time optimal control in spin systems," *Physical Review A*, vol. 63, no. 032308, 2001.

[59] D. D'Alessandro, *Introduction to Quantum Control and Dynamics*. Chapman & Hall/CRC, 2007.

[60] H. I. Nurdin and N. Yamamoto, *Linear Dynamical Quantum Systems: Analysis, Synthesis, and Control*. Berlin: Springer, 2017.

[61] D. Dong and I. R. Petersen, *Learning and Robust Control in Quantum Technology*, ser. Communications and Control Engineering. Springer Verlag, 2023.

[62] ——, "Quantum control theory and applications: A survey," *IET Control Theory & Applications*, vol. 4, no. 12, pp. 2651–2671, 2010.

[63] C. Altafini and F. Ticozzi, "Modeling and control of quantum systems: An introduction," *IEEE Transactions on Automatic Control*, vol. 57, pp. 1898–1917, 2012.

[64] E. Merzbacher, *Quantum Mechanics*, 2nd ed. New York: Wiley, 1970.

[65] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*. Cambridge, UK: Cambridge University Press, 2000.

[66] M. Mirrahimi and R. van Handel, "Stabilizing feedback controls for quantum systems," *SIAM Journal on Control and Optimization*, vol. 46, no. 2, pp. 445–467, 2007.

[67] L. Mazzarella, A. Sarlette, and F. Ticozzi, "Consensus for quantum networks: Symmetry from gossip interactions," *IEEE Transactions on Automatic Control*, vol. 60, no. 1, pp. 158–172, 2015.

[68] L. Bouten, R. van Handel, and M. R. James, "A discrete invitation to quantum filtering and feedback control," *SIAM Review*, vol. 51, no. 2, pp. 239–316, 2009.

[69] F. Albertini and D. D'Alessandro, "Notions of controllability for bilinear multilevel quantum systems," *IEEE Transactions on Automatic Control*, vol. 48, pp. 1399–1403, 2003.

[70] W. Wang and S. G. Schirmer, "Analysis of Lyapunov method for control of quantum states," *IEEE Transactions on Automatic Control*, vol. 55, no. 10, pp. 2259–2270, 2010.

[71] M. Mirrahimi, P. Rouchon, and G. Turinici, "Lyapunov control of bilinear Schrödinger equations," *Automatica*, vol. 41, pp. 1987–1994, 2005.

[72] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, "Optimal control of coupled spin dynamics: design of nmr pulse sequences by gradient ascent algorithms," *Journal of Magnetic Resonance*, vol. 172, no. 2, pp. 296–305, 2005.

[73] D. Dong, M. A. Mabrok, I. R. Petersen, B. Qi, C. Chen, and H. Rabitz, "Sampling-based learning control for quantum systems with uncertainties," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 6, pp. 2155–2166, 2015.

[74] J.-S. Li and N. Khaneja, "Control of inhomogeneous quantum ensembles," *Phys. Rev. A*, vol. 73, p. 030302, Mar 2006.

[75] D. Dong and I. R. Petersen, "Sliding mode control of two-level quantum systems," *Automatica*, vol. 48, no. 5, pp. 725–735, 2012.

[76] ——, "Notes on sliding mode control of two-level quantum systems," *Automatica*, vol. 48, no. 12, pp. 3089–3097, December 2012.

[77] G. Shi, D. Dong, I. R. Petersen, and K. H. Johansson, "Reaching a quantum consensus: Master equations that generate symmetrization and synchronization," *IEEE Transactions on Automatic Control*, vol. 61, no. 2, pp. 374–387, 2016.

[78] J. P. Hespanha, *Noncooperative Game Theory: An Introduction for Engineers and Computer Scientists*. Princeton, NJ: Princeton University Press, 2017.

[79] T. Cheon and I. Tsutsui, "Classical and quantum contents of solvable game theory on hilbert space," *Physics Letters A*, vol. 348, no. 3, pp. 147–152, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0375960105013344

[80] A. Boukas, "Quantum formulation of classical two-person zero-sum games," *Open Systems and Information Dynamics*, vol. 7, pp. 19–32, 2000.

[81] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Informational derivation of quantum theory," *Phys. Rev. A*, vol. 84, p. 012311, Jul 2011. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.84.012311

[82] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Theoretical framework for quantum networks," *Physical Review A*, vol. 80, no. 2, Aug. 2009. [Online]. Available: http://dx.doi.org/10.1103/PhysRevA.80.022339

[83] C. F. Lo and D. Kiang, "To move first or not to move first?" *Quantum Information Processing*, vol. 18, 11 2019.

[84] C. F. Lee and N. F. Johnson, "Efficiency and formalism of quantum games," *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 67, p. 5, 2003.

[85] L. Marinatto and T. Weber, "A quantum approach to static games of complete information," *Physics Letters A*, vol. 272, no. 5, pp. 291–303, 2000. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0375960100004412

[86] E. G. Hidalgo, "Introduction to the study of entropy in quantum games," 2016. [Online]. Available: https://arxiv.org/abs/quant-ph/0604170

[87] G. Gutoski and J. Watrous, "Toward a general theory of quantum games," *Proceedings of the 39th Annual ACM Symposium on Theory of Computing (STOC)*, pp. 565–574, 2007. [Online]. Available: https://doi.org/10.1145/1250790.1250873

[88] J. Wu, "A new mathematical representation of game theory, i," 2004. [Online]. Available: https://arxiv.org/abs/quant-ph/0404159

[89] A. O. Pittenger and M. H. Rubin, "Convexity and the separability problem of quantum mechanical density matrices," 2001. [Online]. Available: https://arxiv.org/abs/quant-ph/0103038

[90] B. Gould and K. Vamvoudakis, "A Novel Framework for Honey-X Deception in Zero-Sum Games," Sep. 2025, arXiv:2509.20329 [cs].

[91] Y. Pu, Y. He, and J. Wang, "Intelligent honeypots: A survey of ai-based deception techniques," *IEEE Access*, vol. 10, pp. 73 401–73 415, 2022.

[92] National Institute of Standards and Technology, "Post-quantum cryptography standardization, round 3 reports," 2022.

[93] J. Eisert, M. Wilkens, and M. Lewenstein, "Quantum games and quantum strategies," *Phys. Rev. Lett.*, vol. 83, pp. 3077–3080, Oct 1999. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.83.3077

[94] S. Zhang, "Quantum strategic game theory," in *ITCS 2012 - Innovations in Theoretical Computer Science Conference*, 2012, pp. 39–59.

[95] D. A. Meyer, "Quantum strategies," *Physical Review Letters*, vol. 82, no. 5, p. 1052–1055, Feb. 1999. [Online]. Available: http://dx.doi.org/10.1103/PhysRevLett.82.1052

[96] J. Eisert and M. Wilkens, "Quantum games," *Journal of Modern Optics*, vol. 47-14, pp. 2543–2556, 2000.

[97] A. P. Flitney and L. C. Hollenberg, "Nash equilibria in quantum games with generalized two-parameter strategies," *Physics Letters A*, vol. 363, no. 5, pp. 381–388, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0375960106018184

[98] J. Wu, "Hamiltotian formalism of game theory," 2005. [Online]. Available: https://arxiv.org/abs/quant-ph/0501088

[99] M. Tambe, *Security and game theory: algorithms, deployed systems, lessons learned*. Cambridge university press, 2011.

[100] V. Conitzer and T. Sandholm, "Computing the optimal strategy to commit to," in *Proceedings of the 7th ACM conference on Electronic commerce*, ser. EC '06. New York, NY, USA: Association for Computing Machinery, Jun. 2006, pp. 82–90.