

Emergent Coordination through Game-Induced Nonlinear Opinion Dynamics

Haimin Hu¹, Kensuke Nakamura², Kai-Chieh Hsu¹, Naomi Ehrich Leonard², and Jaime Fernández Fisac¹

Abstract—We present a multi-agent decision-making framework for the emergent coordination of autonomous agents whose intents are initially undecided. Dynamic non-cooperative games have been used to encode multi-agent interaction, but ambiguity arising from factors such as goal preference or the presence of multiple equilibria may lead to coordination issues, ranging from the “freezing robot” problem to unsafe behavior in safety-critical events. The recently developed nonlinear opinion dynamics (NOD) [1] provide guarantees for breaking deadlocks. However, choosing the appropriate model parameters automatically in general multi-agent settings remains a challenge. In this paper, we first propose a novel and principled procedure for synthesizing NOD based on the value functions of dynamic games conditioned on agents’ intents. In particular, we provide for the two-player two-option case precise stability conditions for equilibria of the game-induced NOD based on the mismatch between agents’ opinions and their game values. We then propose an optimization-based trajectory optimization algorithm that computes agents’ policies guided by the evolution of opinions. The efficacy of our method is illustrated with a simulated toll station coordination example.

I. INTRODUCTION

As deployments of multi-agent autonomous systems, such as self-driving truck fleets and drone swarms, continue to scale up, there is a pressing need to coordinate efficient interaction between controllable agents and uncontrollable agents, including humans. While dynamic games (cf. [2]) capture rich interactive behaviors for multi-agent systems, existing game-theoretic formulations do not effectively coordinate agents when there is uncertainty in the game’s key parameters, such as goal preferences [3, 4], information structure [5], or equilibrium selection [6]. This can lead to dangerous behavior in which agents adopt policies leading to safety-critical deadlocks, sometimes known as the “freezing robot” problem [7]. The recently developed nonlinear opinion dynamics model [1] offers a principled way of describing multi-agent opinion formation over multiple discrete options. In particular, this model provides theoretical guarantees for breaking deadlocks even with no prior bias for a particular option. However, choosing appropriate parameters for the model in practical applications remains an open problem. In this work, we integrate dynamic games with nonlinear opinion dynamics to achieve efficient multi-agent coordination.

A. Related Work

Dynamic games have shown promise in addressing a wide range of multi-agent coordination scenarios, from au-

tonomous driving [8–10] and physical human-robot interaction [11] to smart-grid networks [12]. While computing equilibrium solutions of dynamic games is typically challenging, recently developed toolboxes solve linear-quadratic (LQ) approximations of intricate, non-convex games. Recent work [13] takes advantage of derivative information of system dynamics and planning objectives to iteratively optimize agent trajectories in a dynamic game. This has been shown to lead to efficient trajectories for multi-agent collision avoidance scenarios. However, in the presence of multiple suitable equilibria, this method does not offer a solution to the equilibrium selection problem [6]. Furthermore, the dynamic game becomes a partially observable stochastic game [14] when the intents of a player’s opponents are hidden. While intractable in general, such a game can be approximately solved using, e.g., scenario-based planning [15], Quasi-Newton optimization [9], and the QMDP approach [16, 17], where the Q-value of the underlying Markov decision process (MDP) is learned. In this work, we propose a novel trajectory planning framework that relies on opinion dynamics for handling ambiguities and deadlocks in partially observable stochastic games while remaining computationally tractable by leveraging the LQ and QMDP approximation techniques.

Cathcart et al. [18] use the nonlinear opinion dynamics model [1] for rapid and flexible breaking of social deadlock in robot social navigation. Park et al. [19] leverage the nonlinear opinion dynamics model [1] to investigate how cooperative behavior can emerge in *static* games that are played repeatedly. By jointly considering reciprocity and rationality, cooperation is elicited despite the Nash equilibrium solution being non-cooperative. The region of attraction to the mutually cooperative equilibrium is shown to increase as attention to social interaction increases. However, neither of these works provides an answer for *how* the model parameters of nonlinear opinion dynamics should be chosen. A central contribution of this work is a principled and automatic procedure for constructing nonlinear opinion dynamics based on the outcomes of dynamic games.

B. Contributions and Paper Organization

In this paper, we leverage non-cooperative differential games and nonlinear opinion dynamics to propose a novel trajectory planning framework for multi-agent emergent coordination in tasks about which agents are initially undecided. Our contributions are threefold:

- 1) We propose for the first time an automatic procedure for synthesizing NOD, in which the coupling parameters among opinions depend on the game value functions.

¹Department of Electrical and Computer Engineering, Princeton University, {haiminh, kaichieh, jfisac}@princeton.edu

²Department of Mechanical and Aerospace Engineering, Princeton University, {k.nakamura, naomi}@princeton.edu

We show how the NOD effectively captures opinion evolution driven by the physical state of the system.

- 2) We provide, for the two-player two-option case, precise stability conditions for NOD equilibria based on the mismatch between opinions and game values, which depend on the physical states.
- 3) We present a computationally efficient trajectory planning framework that computes agents' policies guided by their evolving opinions such that coordination on tasks emerges.

The paper is organized as follows. Sec. II provides a brief summary of general-sum differential games and nonlinear opinion dynamics. In Sec. III, we formulate the problem of interest as a differential game subject to parameter uncertainties and describe the construction of game-induced nonlinear opinion dynamics (GiNOD). In Sec. IV, we derive stability conditions for GiNOD. Sec. V presents our main algorithmic contributions for emergent coordination using GiNOD. We show that our proposed planning approach leads to deadlock-free interactions in a toll station coordination scenario in Sec. VI. We conclude our work in Sec. VII.

II. PRELIMINARIES

A. General-Sum Differential Games

We consider an N_a -player finite-horizon general-sum differential game governed by a nonlinear dynamical system:

$$\dot{x} = f(x(t), \mathbf{u}(t)), \quad (1)$$

where $t \in \mathbb{R}$ is the time, $x \in \mathbb{R}^{n_x}$ is the state of the system, $\mathbf{u}(t) := u^{[1:N_a]}(t)$ where $u^i \in \mathbb{R}^{n_{u^i}}$ is the control of player $i \in \mathcal{I}_a := \{1, 2, \dots, N_a\}$. We assume f is continuous in t and continuously differentiable in $\{x(t), \mathbf{u}(t)\}$ uniformly in t . Each player i seeks to minimize a cost functional:

$$J^i \left(\pi^{[1:N_a]} \right) := \int_0^T c^i(x(t), \mathbf{u}(t)) dt, \quad (2)$$

where π^i is player i 's control policy with $u^i(t) = \pi^i(t, x(t))$, $c^i(\cdot)$ is the stage cost of player i , and we assume $c^i(\cdot)$ is twice differentiable in $\{x(t), \mathbf{u}(t)\}$ for all $t \geq 0$.

Finding equilibrium solutions for a general-sum differential game with nonlinear dynamics can be computationally prohibitive in the general case. Fridovich-Keil et al. [13] approach this problem by finding a feedback Nash equilibrium to a *local* approximation of the original game following an iterative linear-quadratic (ILQ) scheme. This method linearizes the dynamics and quadratizes the stage cost function along the nominal trajectory in each iteration. A finite-horizon continuous-time LQ game can then be constructed, which admits a closed-form Riccati differential solution [2, Chapter 6]. The resulting approximate feedback Nash equilibrium solution consists of a tuple of linear strategies. Specifically, player i 's policy is $\pi^{i*}(t, x(t)) = \bar{u}^i(t) + K^i(t)\delta x(t) + \kappa^i(t)$ where $\bar{u}^i(t)$ is the nominal control, $K^i(t) \in \mathbb{R}^{n_{u^i} \times n_x}$ are the gains, $\kappa^i(t) \in \mathbb{R}^{n_{u^i}}$ are the affine terms, and $\delta x(t) := x(t) - \bar{x}(t)$ with nominal state $\bar{x}(t)$. Given all players' approximate Nash equilibrium strategies, the *game value function* of each

player $i \in \mathcal{I}_a$ can be locally approximated by a quadratic function: $V^i(x) \approx \frac{1}{2}\delta x^\top Z^i \delta x + \delta x^\top \zeta^i + v^i(\bar{x})$ with nominal value $v^i(\bar{x})$. Matrices $Z^i \in \mathbb{R}^{n_x \times n_x}$ and $\zeta^i \in \mathbb{R}^{n_x}$ represent the quadratic and linear changes in the nominal value due to small deviations from nominal state $\bar{x}(t)$.

B. Nonlinear Opinion Dynamics

The nonlinear opinion dynamics model [1] captures complex opinion-forming behaviors among multiple agents. For a multi-agent system of N_a agents, each having N_{θ^i} options, the NOD can be expressed as

$$\begin{aligned} \dot{z}^i &= -d^i z^i + b^i + \lambda^i S_z^i(z^i) \\ \dot{\lambda}^i &= -m\lambda^i + S_\lambda^i(z^i) \end{aligned} \quad (3)$$

where

$$\begin{aligned} S_z^{i,\ell}(z^i) &= S_1 \left(\alpha^i z_\ell^i + \sum_{j \in \mathcal{I}_a \setminus \{i\}} \gamma^{ij} z_\ell^j \right) \\ &+ \sum_{p \in \mathcal{I}_{\theta^i} \setminus \{\ell\}} S_2 \left(\beta^i z_p^i + \sum_{j \in \mathcal{I}_a \setminus \{i\}} \delta^{ij} z_p^j \right). \end{aligned}$$

Here, $\mathcal{I}_{\theta^i} := \{1, 2, \dots, N_{\theta^i}\}$, $z^i \in \mathbb{R}^{N_{\theta^i}}$ is agent i 's *opinion vector* in which an element $z_\ell^i > 0$ ($z_\ell^i < 0$) if agent i favors (disfavors) option $\ell \in \mathcal{I}_{\theta^i}$, $d^i > 0$ is the damping term, b^i represents agent i 's own bias, $\lambda^i > 0$ is the attention weight on nonlinear opinion-exchange, $\alpha^i \geq 0$ is the self-reinforcement gain, $\beta^i \geq 0$ is the same-agent inter-option coupling gain, γ^{ij} is the gain of the same-option inter-agent coupling with another agent j , δ^{ij} is the gain of the inter-option inter-agent coupling with another agent j . S_r , $r \in \{1, 2\}$, is a nonlinear saturation function satisfying $S_r(0) = 0$, $S_r'(0) = 1$, $S_r''(0) \neq 0$, $S_r'''(0) \neq 0$, e.g., a sigmoid function or the hyperbolic tangent function \tanh .

The NOD capture and enable a wide range of dynamic interactions and behaviors, notably, fast and flexible multi-agent decision-making. However, it remains an open problem to synthesize parameters that capitalize on these features. One key contribution of this paper is a novel and principled algorithmic approach for automatically synthesizing NOD from a set of differential game value functions.

III. MODELING INDECISION IN DIFFERENTIAL GAMES USING NONLINEAR OPINION DYNAMICS

In this section, we derive a class of nonlinear opinion dynamics, which can be automatically synthesized based on dynamic game solutions, and are later used in the emergent coordination planning framework to be presented in Sec. V.

A. Differential Games with Stochastic Parameters

In this paper, we are interested in differential games where each player's game value functions are dependent on a set of stochastic parameters, i.e., $V^i(x) = V^i(x; \theta^1, \dots, \theta^{N_a})$. We assume that the parameter value θ^i of each player i is supported on a discrete set, i.e. $\theta^i \in \Theta^i := \{\theta_1^i, \dots, \theta_{N_{\theta^i}}^i\}$, which is known to all other players. We further allow Θ^i to be *heterogeneous* for different players. Consequently, parameter $\theta^i \sim p(\theta^i) := (P(\theta_1^i), \dots, P(\theta_{N_{\theta^i}}^i))$ where $p(\theta^i)$ is a categorical distribution over the $(N_{\theta^i} - 1)$ -simplex. The parametrized game value function can capture a broad class

of differential games with *categorically different* outcomes. We provide three examples:

- 1) *Tracking objectives.* We can encode different tracking objectives in players' stage costs. Specifically, we assume player i 's stage cost $c^i(\cdot)$ in (2) can be decomposed into two parts: $c^i(\cdot, \cdot; \theta^i) := c_I^i(\cdot, \cdot) + c_D^i(\cdot, \cdot; \theta^i)$ where $c_I^i(\cdot, \cdot)$ is the parameter-independent part that captures, e.g., regular control objectives and safety specifications, and $c_D^i(\cdot, \cdot; \theta^i)$ depends on the parameter θ^i , which encodes the player's tracking objectives. We present a running example below for illustration.
- 2) *Multiple distinct equilibria.* It is common for multiple equilibria to arise in a differential game [6]. In this case, let $\Theta^i \equiv \Theta$ be homogeneous for all players $i \in \mathcal{I}_a$ and each $\theta_\ell \in \Theta$ represent a particular equilibrium solution of the game. For example, θ_1 represents the equilibrium solution in which player 1 yields to player 2, and θ_2 in which player 2 yields to player 1.
- 3) *Information structure.* Similar to the multiple equilibria case, different information structures in a leader-follower (Stackelberg) game [8] can also define the parameter set Θ and affect the topology of players' trajectories. For example, θ_1 represents a game where player 1 is the leader and θ_2 where player 2 is the leader.

Running example: Consider two autonomous vehicles proceeding towards a toll station as depicted in Fig. 1. The toll booths are modeled as static obstacles that the cars shall avoid. Additional safety-critical specifications include avoiding collisions with other vehicles and driving off the road. For each vehicle, $i \in \{1, 2\}$, let $\Theta^i = \{\theta_1^i, \theta_2^i\}$, where θ_1^i (θ_2^i) represents the option that vehicle i goes through toll booth 1 (toll booth 2). The parameter-independent stage cost $c_I^i(\cdot, \cdot)$ penalizes velocity tracking error, safety violations, and fuel consumption. The parameter-dependent stage cost $c_D^i(x^i; \theta^i) := -w_{\theta^i}^i \mathbf{1}[x^i \in \mathcal{T}_{\theta^i}]$, where $w_{\theta^i}^i > 0$ is the weight and $\mathbf{1}[\cdot]$ is the indicator function, produces a reward (negative cost) when vehicle i is inside the light green *target region* \mathcal{T}_{θ^i} , and zero reward otherwise.

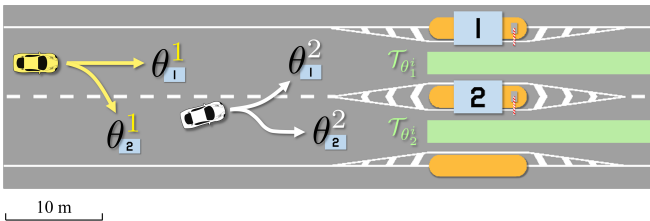


Fig. 1: Emergent coordination involving two autonomous cars at a toll station. Superscripts of parameters θ denote the agent number and subscripts denote the agent's preferred toll booth.

Remark 1: The above game formulation is related to the partially observable stochastic game [14] in that opponent agents' parameters are uncertain. Our formulation *additionally* models the ego agent's parameter as a random variable, which represents the ego agent's *indecision*.

B. QMDP Approximation and Subgames

In this paper, we adopt the QMDP approximation technique [16] in the game-theoretic setting for tractable computation of players' strategies. Each player i 's game value function under parameter uncertainty is computed by solving a QMDP planning problem:

$$\begin{aligned} \tilde{V}^i(x; p(\theta^1), \dots, p(\theta^{N_a})) &:= \min_{u^i \in \mathcal{U}^i} c_I^i(x, u^{[1:N_a]}) + \\ &\mathbb{E}_{\theta^i \sim p(\theta^i), \forall i \in \mathcal{I}_a} [V^i(x^+; \theta^1, \dots, \theta^{N_a})] \\ x^+ &= \bar{f}(x, u^i, \{\pi^j(x; \theta^1, \dots, \theta^{N_a})\}_{j \in \mathcal{I}_a \setminus \{i\}}), \end{aligned} \quad (4)$$

where set $\mathcal{U}^i \subseteq \mathbb{R}^{n_{u_i}}$ encodes the control limit of agent i , $c_I^i(\cdot, \cdot)$ is the parameter-independent stage cost, game value function $V^i(x; \theta^1, \dots, \theta^{N_a})$ results from equilibrium policies under a set of *fully revealed* parameters, $\bar{f}(\cdot)$ is the time-discretized dynamics (1), and $\pi^j(x; \theta^1, \dots, \theta^{N_a})$ is player j 's equilibrium policy of the game parametrized by $(\theta^1, \dots, \theta^{N_a})$. We call the game governed by a particular set of parameters $(\theta^1, \dots, \theta^{N_a})$ a *subgame*. In this paper, we solve for approximate subgame value functions using the ILQ Game method [13]. QMDP (4) optimistically assumes that the parameter uncertainties disappear in one step after *the ego agent* takes an action, and that all opponents are *clairvoyant* playing their corresponding subgame policies.

Assumption 1: We assume the availability of Subroutine \mathcal{S} that computes (approximate) subgame value functions $V^i(x; \theta^1, \dots, \theta^{N_a})$ for all players $i \in \mathcal{I}_a$ and all possible parameter combinations $\theta^1 \in \Theta^1, \dots, \theta^{N_a} \in \Theta^{N_a}$.

QMDP (4) is now tractable as a single-agent trajectory optimization problem, and can be solved repeatedly in a receding horizon fashion. Nonetheless, propagating parameter uncertainties $p(\theta^i)$ can still be challenging. Existing works [6, 15] use Bayesian inference [20] to propagate $p(\theta^i)$ based on observations of opponents' behaviors. Drawbacks of Bayesian approaches include intractability for general (multi-modal) distributions and difficulty in defining an appropriate likelihood model. In this paper, we take an alternative approach by interpreting $p(\theta^i)$ as the agent's *degree of indecision*, and model its time evolution using nonlinear opinion dynamics.

Running example: There are four subgames resulting from the parameter combinations (θ_1^1, θ_1^2) , (θ_1^1, θ_2^2) , (θ_2^1, θ_1^2) , and (θ_2^1, θ_2^2) , which encode agents' choices to go through a particular toll booth. We solve each subgame using the ILQ method [13] for agents' state trajectories that correspond to an approximate local Nash equilibrium solution.

C. From Probabilities to Opinions

We let opinion state vector $z^i \in \mathbb{R}^{N_{\theta^i}}$ model distribution $p(\theta^i)$, leveraging the softmax operation:

$$P(\theta_\ell^i) \leftarrow \sigma_\ell(z^i) := \frac{e^{z_\ell^i}}{\sum_{q=1}^{N_{\theta^i}} e^{z_q^i}}, \quad (5)$$

and define $\sigma(z^i) := (\sigma_1(z^i), \dots, \sigma_{N_{\theta^i}}(z^i))$.

By delegating uncertainties to opinions, given a physical state x and all players' opinions $\mathbf{z} := (z^1, \dots, z^{N_a})$, we can

define player i 's *opinion-weighted* game value function:

$$\hat{V}^i(\mathbf{z}, x) := \sum_{\ell_1=1}^{N_{\theta^1}} \cdots \sum_{\ell_{N_a}=1}^{N_{\theta^{N_a}}} \left(\prod_{i=1}^{N_a} \sigma_{\ell_i}(z^i) \right) V^i(x; \theta_{\ell_1}^1, \dots, \theta_{\ell_{N_a}}^{N_a}), \quad (6)$$

as a proxy for $\mathbb{E}_{\theta^1, \dots, \theta^{N_a}} [V^i(x; \theta^1, \dots, \theta^{N_a})]$, the expected game value in QMDP (4).

Remark 2: The neutral opinion $z = (0, \dots, 0)$ corresponds to an (uninformative) uniform distribution $p(\theta) \leftarrow \sigma(z) = \frac{1}{N_{\theta}} \mathbf{1}_{N_{\theta}}$, where $\mathbf{1}_{N_{\theta}} \in \mathbb{R}^{N_{\theta}}$ is a vector of all ones.

D. Synthesizing Opinion Dynamics from Subgames

Opinion evolution as gradient flow. In this section, we propose a *constructive way* to synthesize opinion dynamics from subgames. We model the agents as selfish players who seek to drive their opinions in a direction that minimizes their own expected game value. To this end, we let all agents implement the gradient flow [21, Ch. 14] dynamics (the continuous-time counterpart of gradient descent) that describe the evolution of their opinion states:

$$\dot{\mathbf{z}} = [-\nabla_{z^i} \hat{V}^i(\mathbf{z}, x)]_{i \in \mathcal{I}_a} \quad (7)$$

where $[\cdot]_{i \in \mathcal{I}_a}$ denotes vector concatenation by rows.

Linear opinion dynamics. The local behavior of gradient flow (7) around a given $\bar{\mathbf{z}}$ induces a linear opinion dynamics model, originally introduced as a weighted-averaging process [22]. This is given by linearizing (7) at $\bar{\mathbf{z}}$:

$$\delta \dot{\mathbf{z}} = \begin{bmatrix} -H_{11}^1(x) & \cdots & -H_{1N_a}^1(x) \\ \vdots & \ddots & \vdots \\ -H_{N_a 1}^{N_a}(x) & \cdots & -H_{N_a N_a}^{N_a}(x) \end{bmatrix} \begin{bmatrix} \delta z^1 \\ \vdots \\ \delta z^{N_a} \end{bmatrix}, \quad (8)$$

where $\delta \mathbf{z} := \mathbf{z} - \bar{\mathbf{z}}$, $\delta z^i := z^i - \bar{z}^i$, and

$$H_{ij}^i(x) := \nabla_{z^i z^j} \hat{V}^i(\mathbf{z}, x) \Big|_{\bar{\mathbf{z}}}$$

is the Hessian matrix of $\hat{V}^i(\cdot)$ with respect to z^i and z^j , evaluated at $\bar{\mathbf{z}}$. We can rewrite linear opinion dynamics (8) equivalently as

$$\begin{aligned} \delta \dot{z}_{\ell}^i &= \alpha_{\ell}^i(x) \delta z_{\ell}^i + \sum_{j=1}^{N_a} \gamma_{\ell}^{ij}(x) \delta z_{\ell}^j \\ &+ \sum_{p=1}^{N_{\theta^i}} \beta_{\ell p}^i(x) \delta z_p^i + \sum_{p=1}^{N_{\theta^i}} \sum_{j=1}^{N_a} \eta_{\ell p}^{ij}(x) \delta z_p^j, \end{aligned} \quad (9)$$

for player i 's parameter θ_{ℓ}^i , where the *state-dependent* dynamics parameters are defined as

$$\alpha_{\ell}^i(x) := -[H_{ii}^i(x)]_{\ell \ell} \quad (10a)$$

$$\gamma_{\ell}^{ij}(x) := -[H_{ij}^i(x)]_{\ell \ell} \quad (10b)$$

$$\beta_{\ell p}^i(x) := -[H_{ii}^i(x)]_{\ell p} \quad (10c)$$

$$\eta_{\ell p}^{ij}(x) := -[H_{ij}^i(x)]_{\ell p} \quad (10d)$$

Here, $[H(\cdot)]_{\ell p}$ denotes the entry located at the ℓ -th row and p -th column of $H(\cdot)$. The reasoning behind those dynamics parameters is the same as (3) with additional flexibility to model inter-option dependency (see [23] and Appendix A in

[1]). The linearized gradient flow dynamics (8) synthesized from subgames fall into the category of linear opinion dynamics [22], also known, in discrete time, as the DeGroot model [24].

Game-induced nonlinear opinion dynamics. Motivated by the discovery made in recent work [1] that nonlinearity addresses lack of structural stability of agreement and disagreement equilibria and other issues in linear opinion dynamics, we modify linear opinion dynamics (8) and propose the Game-induced Nonlinear Opinion Dynamics (GiNOD):

$$\begin{aligned} \delta \dot{z}_{\ell}^i &= g_{\ell}^i(\delta \mathbf{z}, \lambda^i, x) := -D^i \delta z_{\ell}^i + \lambda^i [g_{\ell}^i(\delta \mathbf{z}, x)]_{\ell \in \mathcal{I}_{\theta^i}} \\ g_{\ell}^i(\delta \mathbf{z}, x) &:= S_1 \left(\alpha_{\ell}^i(x) \delta z_{\ell}^i + \sum_{j \in \mathcal{I}_a \setminus \{i\}} \gamma_{\ell}^{ij}(x) \delta z_{\ell}^j \right) + \\ &\sum_{p \in \mathcal{I}_{\theta^i} \setminus \{\ell\}} S_2 \left(\beta_{\ell p}^i(x) \delta z_p^i + \sum_{j \in \mathcal{I}_a \setminus \{i\}} \eta_{\ell p}^{ij}(x) \delta z_p^j \right), \end{aligned} \quad (11)$$

for all players $i \in \mathcal{I}_a$ and parameters θ_{ℓ}^i , $\ell \in \mathcal{I}_{\theta^i}$, where $D^i = \text{diag}(d_1^i, \dots, d_{N_{\theta^i}}^i) \in \mathbb{R}^{N_{\theta^i} \times N_{\theta^i}}$ is a symmetric positive definite matrix that describes damping of the opinion states, $\lambda^i \in \mathbb{R}$ is the attention. The attention λ^i can be interpreted as a gain on the saturated gradient flow dynamics $g_{\ell}^i(\cdot)$. A greater λ^i promotes opinion formation (z^i departing the origin), and a smaller λ^i discourages opinion formation (z^i approaching the origin). Note that the full opinion states are recovered by $\mathbf{z} = \delta \mathbf{z} + \bar{\mathbf{z}}$, where the nominal opinion states $\bar{\mathbf{z}}$ are recursively updated using an iterative algorithm (Alg. 1) to be introduced in the next section.

E. When to Pay Attention in Games: the Price of Indecision

We now introduce for each player i a measure called the *Price of Indecision* (PoI) defined as

$$\text{PoI}^i(\mathbf{z}, x) := \max_{\ell_j \in \mathcal{I}_{\theta^j}, \forall j \neq i} \frac{\sum_{\ell_i \in \mathcal{I}_{\theta^i}} \sigma_{\ell_i}(z^i) V^i(x; \theta_{\ell_1}^1, \dots, \theta_{\ell_{N_a}}^{N_a})}{\min_{p_i \in \mathcal{I}_{\theta^i}} V^i(x; \theta_{\ell_1}^1, \dots, \theta_{p_i}^i, \dots, \theta_{\ell_{N_a}}^{N_a})}. \quad (12)$$

The PoI, inspired by the Price of Anarchy [25], is a ratio lower bounded by 1 that measures how player i 's efficiency degrades due to indecision. For a set of worst-case parameters selected by opponent players $j \neq i$, PoI will be large if the (expected) game value when player i chooses not to form an opinion (the numerator in (12)) outweighs the game value when player i declares an optimal opinion (the denominator in (12)). Given the player's PoI, we introduce a state-and-opinion-dependent dynamic equation for evolving player i 's attention [1]:

$$\dot{\lambda}^i = g_{\lambda}^i(\lambda^i, \mathbf{z}, x) := -m^i \lambda^i + \rho^i (\text{PoI}^i(\mathbf{z}, x) - 1), \quad (13)$$

where $m^i > 0$ and $\rho^i > 0$ are damping and scaling parameters. The PoI-based dynamics (13) increase the attention λ^i when PoIⁱ is large, which promotes opinion formation. In other words, the agents are more inclined to form their opinions *only if* doing so increases their efficiency.

IV. STABILITY ANALYSIS OF GAME-INDUCED NOD

In this section, we derive precise stability conditions for GiNOD equilibria (11) in a two-player, two-option setting,

i.e. $\mathcal{I}_a = \{1, 2\}$, $\Theta_1 = \{\theta_1^1, \theta_2^1\}$, and $\Theta_2 = \{\theta_1^2, \theta_2^2\}$. Note again that the elements in Θ_1 and Θ_2 need not coincide with each other. Key properties¹ of GiNOD discovered from our analysis are summarized below:

- 1) The neutral opinion for both players is unstable for an *arbitrarily* small damping parameter. That is, when $\bar{\mathbf{z}} = 0$, $\delta\mathbf{z} = 0$ is an unstable equilibrium of GiNOD and thus indecision is easily broken (Theorem 1),
- 2) When both agents have formed their non-neutral opinions, $\delta\mathbf{z} = 0$ is a locally exponentially stable equilibrium for GiNOD for an *arbitrarily* small damping parameter, if both agents' opinions correspond to a lower (i.e. better) game value from the current physical state (Theorem 2),
- 3) If for both agents there is no difference between game values of distinct options, then the opinions are driven purely by the damping terms (Corollary 1).

We start by reviewing two useful linear algebra lemmas.

Lemma 1 (Theorem 4.2.12 [26]): Let $A \in \mathbb{R}^{m \times m}$ and $B \in \mathbb{R}^{n \times n}$ with $\lambda \in \text{spec}(A)$ and $\mu \in \text{spec}(B)$, where $\text{spec}(\cdot)$ denotes spectrum. Then $\lambda\mu$ is an eigenvalue of $A \otimes B$, where \otimes is the Kronecker product. Any eigenvalue of $A \otimes B$ arises as such a product of eigenvalues of A and B .

Lemma 2: Let $D = dI_n$ where $d \in \mathbb{R}$ and I_n is the identity matrix in $\mathbb{R}^{n \times n}$. If λ is an eigenvalue of $H \in \mathbb{R}^{n \times n}$, then $d + c\lambda$ is an eigenvalue of $D + cH$ where $c \in \mathbb{R}$.

Proof: Since $\lambda \in \text{spec}(H)$, we have that $Hv = \lambda v$ where v is the eigenvector associated with λ . It follows that $(D + cH)v = dIv + cHv = (d + c\lambda)v$. ■

In the two-player, two-option case, it is possible to derive and analyze each entry of the system matrix in opinion dynamics (8), as shown in Lemma 3. To ease the notation we denote value function $V_{\ell p}^i := V^i(x; \theta_\ell^1, \theta_p^2)$.

Lemma 3: Let value function $\hat{V}^i(\mathbf{z}, x)$ be defined in (6). Matrix $\mathbf{H}(x) := - \left[J_{\mathbf{z}} \nabla_{\mathbf{z}^i} \hat{V}^i(\mathbf{z}, x) \right]_{i \in \mathcal{I}_a} \Big|_{\bar{\mathbf{z}}}$ = $\Gamma \otimes H$ with $H := \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ and $\Gamma := \begin{bmatrix} a_1 & b_1 \\ b_2 & a_2 \end{bmatrix}$ where

$$\begin{aligned} a_1 &:= \phi_a(\bar{z}^1) [\sigma_1(\bar{z}^2) (V_{11}^1 - V_{21}^1) + \sigma_2(\bar{z}^2) (V_{12}^1 - V_{22}^1)] \\ a_2 &:= \phi_a(\bar{z}^2) [\sigma_1(\bar{z}^1) (V_{11}^2 - V_{12}^2) + \sigma_2(\bar{z}^1) (V_{21}^2 - V_{22}^2)] \\ b_i &:= \phi_b(\bar{z}^1) \phi_b(\bar{z}^2) (-V_{11}^i - V_{22}^i + V_{12}^i + V_{21}^i) \\ \phi_a(\bar{z}^i) &:= (\sigma_1(\bar{z}^i) - \sigma_2(\bar{z}^i)) \phi_b(\bar{z}^i) \\ \phi_b(\bar{z}^i) &:= \sigma_1(\bar{z}^i) \sigma_2(\bar{z}^i) \end{aligned}$$

for $i \in \{1, 2\}$, and $J_{\mathbf{z}}(\cdot)$ denotes the Jacobian matrix with respect to \mathbf{z} . Furthermore, $\text{spec}(\mathbf{H}(x)) = \{0, 0, a_1 + a_2 \pm ((a_1 - a_2)^2 + 4b_1b_2)^{1/2}\}$.

Proof: Entries of $\mathbf{H}(x)$ are computed from the Jacobian matrices of $\nabla_{\mathbf{z}^i} \hat{V}^i(\mathbf{z}, x)$ for $i \in \{1, 2\}$. Spectrum of $\mathbf{H}(x)$ follows by Lemma 1 since $\text{spec}(\Gamma) = \{\frac{1}{2}[a_1 + a_2 \pm ((a_1 - a_2)^2 + 4b_1b_2)^{1/2}]\}$ and $\text{spec}(H) = \{0, 2\}$. ■

It is easily observed that $\delta\mathbf{z} = 0$ is an equilibrium of (11). To facilitate the analysis of stability of $\delta\mathbf{z} = 0$, we consider

¹All informal conclusions listed here are subject to additional technical assumptions, which can be found in the Theorems and Corollary.

GiNOD (11) with a *steady-state* attention $\lambda^1 = \lambda^2 \equiv \bar{\lambda} := \lambda_\infty(\text{PoI}) > 0$ under attention dynamics (13) with a fixed PoI, i.e. $\text{PoI}^1(\cdot) = \text{PoI}^2(\cdot) \equiv \text{PoI} \geq 1$. The steady-state attention $\bar{\lambda}$ is guaranteed to exist in practice since $m^i > 0$ and thus $\lim_{t \rightarrow \infty} \lambda^i(t) < \infty$ for any $\text{PoI} < \infty$. We also assume the same damping term for both players' options, i.e., $D^1 = D^2 = D = \text{diag}(d, d)$. Define block diagonal matrix $\mathbf{D} = \text{blkdiag}(D, D)$.

Theorem 1 (Instability at neutral opinion): Let $\bar{\mathbf{z}} = 0$, i.e., all agents hold a neutral opinion. Then, $\delta\mathbf{z} = 0$ is an unstable equilibrium of GiNOD (11) if $d < 2\bar{\lambda} \text{Re}(\sqrt{b_1b_2})$.

Proof: When $\bar{\mathbf{z}} = 0$, linearization of GiNOD at $\delta\mathbf{z} = 0$ gives linear system $\delta\dot{\mathbf{z}} = (-\mathbf{D} + \bar{\lambda}\mathbf{H})\delta\mathbf{z}$, where $a_1 = a_2 = 0$ in \mathbf{H} since $\phi_a(\bar{z}^1) = \phi_a(\bar{z}^2) = 0$. By Lemma 2 and 3, we have $\text{spec}(-\mathbf{D} + \bar{\lambda}\mathbf{H}) = \{-d, -d, -d \pm 2\bar{\lambda}\sqrt{b_1b_2}\}$. Thus, $\delta\mathbf{z} = 0$ is unstable if $d < 2\bar{\lambda} \text{Re}(\sqrt{b_1b_2})$. ■

Remark 3 (Instability and (dis)agreement): It is desirable that, at the neutral opinion, GiNOD can be made unstable even if damping d is not close to zero so that agents can quickly and reliably form a non-neutral opinion to break any deadlock. This requires that $\sqrt{b_1b_2}$ has a non-zero real part, or $b_1b_2 > 0$. Since $\phi_b(\bar{z}^1)\phi_b(\bar{z}^2) > 0$, $b_1b_2 > 0$ if and only if $V_b := \prod_{i \in \{1, 2\}} (-V_{11}^i - V_{22}^i + V_{12}^i + V_{21}^i) > 0$. This is true in two cases. First, if both players find that agreeing to the same option would be costly (e.g. squeezing into the same toll station in the Running Example), i.e. $V_{\ell\ell}^1$ and V_{pp}^2 are large for some $\ell, p \in \{1, 2\}$, then $V_b > 0$ as a result of multiplying two negative numbers. Second, if the cost of disagreement is high for both players (typical case in cooperative settings), i.e. $V_{\ell-\ell}^1$ and V_{p-p}^2 are large for some $\ell, p \in \{1, 2\}$, where $-\ell$ denotes the alternative option to ℓ , then $V_b > 0$ since it is the product of two positive numbers.

Theorem 2 (Opinion reflects game value): If for $\ell_1, \ell_2 \in \{1, 2\}$ it holds that $\sigma_{\ell_1}(\bar{z}^1) > \sigma_{-\ell_1}(\bar{z}^1)$, $\sigma_{\ell_2}(\bar{z}^2) > \sigma_{-\ell_2}(\bar{z}^2)$, $V_{\ell_1\ell_2}^1 < V_{-\ell_1\ell_2}^1$, $V_{\ell_1\ell_2}^2 < V_{-\ell_1\ell_2}^2$, and $a_1a_2 > b_1b_2$, then $\delta\mathbf{z} = 0$ is a locally exponentially stable equilibrium of GiNOD with an arbitrarily small damping $d > 0$.

Proof: We follow the proof of Theorem 1 by examining the spectrum of the linearized system matrix $-\mathbf{D} + \bar{\lambda}\mathbf{H}$. From $\sigma_{\ell_1}(\bar{z}^1) > \sigma_{-\ell_1}(\bar{z}^1)$, $\sigma_{\ell_2}(\bar{z}^2) > \sigma_{-\ell_2}(\bar{z}^2)$, $V_{\ell_1\ell_2}^1 < V_{-\ell_1\ell_2}^1$, and $V_{\ell_1\ell_2}^2 < V_{-\ell_1\ell_2}^2$, we have that $a_1, a_2 < 0$. $\delta\mathbf{z} = 0$ is locally exponentially stable with any $d > 0$ if and only if $\max(\text{spec}(-\mathbf{D} + \bar{\lambda}\mathbf{H})) = \max\{-d, -d + \bar{\lambda}(a_1 + a_2 + ((a_1 - a_2)^2 + 4b_1b_2)^{1/2})\} < 0$. Since the steady-state attention $\bar{\lambda} > 0$, if $a_1 + a_2 + ((a_1 - a_2)^2 + 4b_1b_2)^{1/2} < 0$, then $\max(\text{spec}(-\mathbf{D} + \bar{\lambda}\mathbf{H})) < 0$ with an arbitrarily small $d > 0$. Solving the inequality gives $a_1a_2 > b_1b_2$. ■

Remark 4 (Interpreting inequality $a_1a_2 > b_1b_2$): We observe that $a_1a_2 > b_1b_2$ holds with mild assumptions on game values when both players *have formed their opinions*, that is, $\sigma_{\ell_1}(\bar{z}^1) \rightarrow 1$ and $\sigma_{\ell_2}(\bar{z}^2) \rightarrow 1$ for some $\ell_1, \ell_2 \in \{1, 2\}$, and players' opinions reflect their game values (conditions $V_{\ell_1\ell_2}^1 < V_{-\ell_1\ell_2}^1$ and $V_{\ell_1\ell_2}^2 < V_{-\ell_1\ell_2}^2$ in Theorem 2). Since $a_1, a_2 < 0$, $a_1a_2 > b_1b_2$ trivially holds if $b_1b_2 < 0$.

When $b_1b_2 > 0$, $a_1a_2 > b_1b_2$ holds if $a_1a_2/b_1b_2 > 1$. Without loss of generality, assume $\sigma_1(\bar{z}^i) \rightarrow 1$ and $\sigma_2(\bar{z}^i) \rightarrow 0$ for both players. Since $\sigma_1(\bar{z}^i) \rightarrow 1$, the ratio $a_1a_2/b_1b_2 \rightarrow$

$c_1 c_2 V'$ where $c_1 := (\sigma_1(\bar{z}^1) - \sigma_2(\bar{z}^1))/\sigma_2(\bar{z}^2) \gg 1$, $c_2 := (\sigma_1(\bar{z}^2) - \sigma_2(\bar{z}^2))/\sigma_2(\bar{z}^1) \gg 1$ and $V' := V_a/V_b$ where both the numerator $V_a := (V_{11}^1 - V_{21}^1)(V_{11}^2 - V_{12}^2)$ and the denominator $V_b := \prod_{i \in \{1,2\}} (-V_{11}^i - V_{22}^i + V_{12}^i + V_{21}^i)$ are positive. Therefore, as long as the game values $V_{\ell_1 \ell_2}^i$ are such that V' is not too small (e.g., when those game values are roughly of the same magnitude), $a_1 a_2 > b_1 b_2$ holds.

Remark 5 (Guidelines for choosing damping): Based on Theorem 1 and 2, it is recommended to pick a damping term that satisfies $0 < d < 2\lambda\sqrt{b_1 b_2}$ when $b_1 b_2 > 0$. In this way the equilibrium $\delta \mathbf{z} = 0$ of GiNOD is unstable at the neutral opinion, while remaining locally exponentially stable when agents' opinions match their game value differences.

Corollary 1 (Identical game values cannot form opinion): If both agents have the same game values for different options, i.e. $V_{11}^1 = V_{21}^1$, $V_{12}^1 = V_{22}^1$, $V_{11}^2 = V_{12}^2$, $V_{21}^2 = V_{22}^2$, then the opinions are purely driven by the damping term d , i.e., $\delta \mathbf{z} = 0$ is always locally exponentially stable.

Proof: By Lemma 3 we have in this case $\mathbf{H} = 0$. GiNOD (11) becomes $\delta \dot{z}^i = -D^i \delta z^i$ for $i \in \{1, 2\}$. ■

V. EMERGENT COORDINATION PLANNING USING GAME-INDUCED NONLINEAR OPINION DYNAMICS

In this paper, we seek to combine the best attributes from differential games and game-induced nonlinear opinion dynamics towards efficient, deadlock-free multi-agent emergent coordination, in which all players are initially undecided about their parameters of the game. To this end, we first formulate an *opinion-weighted* QMDP based on the up-to-date opinions evolved with GiNOD in Sec. V-A. Then we modify and extend the QMDP formulation using ideas from cognitive hierarchy to enable *active* opinion manipulation in Sec. V-B. The overall framework of our approach applied in receding horizon fashion is summarized in Algorithm 1.

A. Opinion-Weighted QMDP

Given a physical state x and players' opinions \mathbf{z} , we can formulate the opinion-weighted QMDP by combining QMDP (4) and opinion-weighted game value function (6). Player i 's strategy is given by:

$$\begin{aligned} \pi_{L0}^i(x, \mathbf{z}) &:= \arg \min_{u^i \in \mathcal{U}^i} c_I^i(x, u^i) + \hat{V}^i(\mathbf{z}, x^+) \\ x^+ &= \bar{f}(x, \pi_{L0}^i(x, \mathbf{z}), \{\pi^j(x; \theta^1, \dots, \theta^{N_a})\}_{j \in \mathcal{I}_a \setminus \{i\}}), \end{aligned} \quad (14)$$

where $c_I^i(\cdot, \cdot)$ is the parameter-independent stage cost, $\hat{V}^i(\cdot)$ is the opinion-weighted game value function defined in (6), and $\pi^j(x; \theta^1, \dots, \theta^{N_a})$ is player j 's equilibrium policy of the subgame parametrized by $(\theta^1, \dots, \theta^{N_a})$. We refer to policy (14) as the *Level-0* opinion-weighted QMDP (L0-QMDP) policy, whose namesake will become clearer in the next section as we introduce the Level-1 QMDP policy. Aligned with the QMDP principle, an agent using the L0-QMDP policy first declares an action, then commits to an option and assumes that the parameter uncertainties of other agents disappear. The following proposition shows that if the subgames are LQ games, then the L0-QMDP, under mild assumptions, can be cast as a convex quadratic program, which can be solved efficiently via off-the-shelf solvers.

Proposition 1: If all subgames are LQ games, the (physical) state evolves under a control-affine dynamic model, i.e. $x^+ = \bar{f}(x) + \sum_{i \in \mathcal{I}_a} \bar{g}^i(x) u^i$, stage cost $c_I^i(\cdot, \cdot)$ is a convex quadratic function in u^i , and control set \mathcal{U}^i is convex, then QMDP (14) is a convex quadratic program.

Proof: Each term $V^i(x^+; \theta_{\ell_1}^1, \dots, \theta_{\ell_{N_a}}^{N_a})$ in $\hat{V}^i(\cdot)$ can be expanded by plugging in the dynamic model as $V^i(\cdot) = [(u^i - \bar{u}^i)^\top \bar{g}^i(x)^\top \bar{Z}^i + \bar{\zeta}^{i,\top}] \bar{g}^i(x) (u^i - \bar{u}^i) + \bar{C}$, which is a quadratic function in u^i , where $(\bar{u}^i, \bar{Z}^i, \bar{\zeta}^i)$ are shorthand notations for the value function parameters of the subgame defined by $(\theta_{\ell_1}^1, \dots, \theta_{\ell_{N_a}}^{N_a})$, and $\bar{C} \in \mathbb{R}$ is a constant. ■

Despite its simplicity and efficient computation, we note that the L0-QMDP policy does not take into account opinion evolution as a result of agents' actions, thereby unable to actively steer the opinions. In the next section, we modify the L0-QMDP to enable active manipulation of agents' opinions.

Remark 6: As an alternative to QMDP (14) where opponents are assumed as clairvoyant players, we may solve a QMDP Game:

$$\begin{aligned} \pi^i(x, \mathbf{z}) &:= \arg \min_{u^i \in \mathcal{U}^i} c_I^i(x, u^i) + \hat{V}^i(\mathbf{z}, x^+) \\ x^+ &= \bar{f}(x, \pi^1(x, \mathbf{z}), \dots, \pi^{N_a}(x, \mathbf{z})), \end{aligned}$$

in which the agents' QMDP problems are coupled. If the conditions in Prop. 1 are satisfied and all agents' inputs are unconstrained, then it can be shown that the QMDP Game is an LQ Game, whose *global* feedback Nash equilibrium can be computed efficiently via coupled Riccati equations [2].

B. Actively Manipulating Opinions

We now introduce the *Level-1* opinion-weighted QMDP (L1-QMDP) policy, which is inspired by the established work on cognitive hierarchy (K -level reasoning) [27]. The ego player using the L1-QMDP policy assumes that all opponents apply the L0-QMDP policy. This way, the ego player can declare two actions sequentially in time—the first one evolves the current opinions forward in time through GiNOD, and all uncertainties are assumed to disappear after the ego's second action is determined. Given players' physical state x , opinions \mathbf{z} , and attentions $\Lambda := (\lambda^1, \dots, \lambda^{N_a})$, we can formulate the L1-QMDP planning problem as:

$$\begin{aligned} \pi_{L1}^i(x, \mathbf{z}, \Lambda) &= u_0^i(x, \mathbf{z}, \Lambda) \\ \min_{u_0^i, u_1^i \in \mathcal{U}^i} & c_I^i(x_0, u_0^i) + c_I^i(x_1, u_1^i) + \hat{V}^i(\mathbf{z}_1, x_2) \\ \text{s.t. } & x_0 = x, \mathbf{z}_0 = \mathbf{z}, \Lambda_0 = \Lambda \\ & x_1 = \bar{f}(x_0, u_0^i, \{\pi_{L0}^j(x_0, \mathbf{z}_0)\}_{j \in \mathcal{I}_a \setminus \{i\}}) \\ & x_2 = \bar{f}(x_1, u_1^i, \{\pi^j(x_1; \theta^1, \dots, \theta^{N_a})\}_{j \in \mathcal{I}_a \setminus \{i\}}) \\ & \mathbf{z}_1 = \bar{g}_z(\mathbf{z}_0, \Lambda_0, x_1(u_0^i)), \end{aligned} \quad (15)$$

where $\bar{g}_z(\cdot) := [\bar{g}_z^i(\cdot)]_{i \in \mathcal{I}_a}$ is the discrete-time joint opinion dynamics given by concatenating players' time-discretized GiNOD $\bar{g}_z^i(\cdot)$ defined in (11). From (15) we can see that the control action of the ego agent i who plays the L1-QMDP policy is optimized given the knowledge that it is able to affect opinion \mathbf{z}_1 through GiNOD $\bar{g}_z(\cdot)$. Problem (15) is in general a non-convex optimization problem as it involves the nonlinear physical dynamics $\bar{f}(\cdot)$ and GiNOD $\bar{g}_z(\cdot)$.

Algorithm 1 Receding Horizon QMDP using GiNOD

Input: Initial state $x(0)$, opinions $\mathbf{z}(0)$, nominal opinions $\bar{\mathbf{z}}(0)$, attentions $\Lambda(0)$, horizon step T

- 1: Initialize time step $t \leftarrow 0$
 - 2: **for** $t = 0, 1, \dots, T - 1$ **do**
 - // Solve subgames**
 - 3: $V^i(x(t); \theta^1, \dots, \theta^{N_a}) \leftarrow$ Solve subgames using Subroutine \mathcal{S} for all players $i \in \mathcal{I}_a$ and all parameter combinations $\theta^1 \in \Theta^1, \dots, \theta^{N_a} \in \Theta^{N_a}$
 - // Construct opinion dynamics**
 - 4: Set nominal opinions $\bar{\mathbf{z}}(t) \leftarrow \mathbf{z}(t)$
 - 5: Construct GiNOD $g_z^i(\delta \mathbf{z}, \lambda^i, x)$ in (11) and attention dynamics $g_\lambda^i(\lambda^i, \mathbf{z}, x)$ in (13) for all players $i \in \mathcal{I}_a$
 - // Compute QMDP policies**
 - 6: $u^i(t) \leftarrow$ Compute control action using the L0-QMDP policy $\pi_{L0}^i(x(t), \mathbf{z}(t))$ in (14) or the L1-QMDP policy $\pi_{L1}^i(x(t), \mathbf{z}(t), \Lambda(t))$ in (15) for each player $i \in \mathcal{I}_a$
 - // Update states, opinions, and attentions**
 - 7: $x(t+1) \leftarrow$ Integrate $f(x(t), \mathbf{u}(t))$
 - 8: **if** $t \geq 1$ **then**
 - 9: $\delta \mathbf{z}^i(t) \leftarrow$ Integrate GiNOD $g_z^i(\delta \mathbf{z}(t-1), \lambda^i(t-1), x(t))$ for all players $i \in \mathcal{I}_a$
 - 10: $\lambda^i(t) \leftarrow$ Integrate attention dynamics $g_\lambda^i(\lambda^i(t-1), \mathbf{z}(t), x(t))$ for all players $i \in \mathcal{I}_a$
 - 11: $\mathbf{z}(t) \leftarrow \bar{\mathbf{z}}(t-1) + \delta \mathbf{z}(t)$
 - 12: **end if**
 - 13: **end for**
-

VI. SIMULATION RESULTS

We apply the receding-horizon opinion-weighted QMDP planning framework (Alg. 1) to the toll station coordination task described in the Running Example. Both vehicles $i \in \{1, 2\}$ are described by a kinematic bicycle model [28], whose state is defined as $x^i = (p_x^i, p_y^i, \varphi^i, v^i)$. Here, p_x^i and p_y^i are the center position of car i 's rear axes, φ^i is the yaw angle with respect to the x -axis, and v^i is the velocity with respect to the rear axes. The joint state vector is $x := (x^1, x^2) \in \mathbb{R}^8$. All continuous-time dynamics were discretized with a time step of $\Delta t = 0.2$ s using the forward Euler method. We used a JAX [29]-based implementation of the ILQ Game method [13] as the Subroutine \mathcal{S} for solving the subgames. The QMDP optimization problems were modeled and solved using CasADi [30]. The open-source code is available online.²

The physical states of the vehicles were initialized to be $x(0) = (0 \text{ m}, 5 \text{ m}, 0 \text{ rad}, 3 \text{ m/s}, 5 \text{ m}, 2 \text{ m}, 0 \text{ rad}, 3 \text{ m/s})$. We modeled the *initially undecided* agents by using an almost-neutral initial opinion $z_{\ell_i}^i = \epsilon$ for $i \in \{1, 2\}$ and $\ell_i \in \{1, 2\}$ in all simulations, where $\epsilon > 0$ is a small number that prevents opinions from staying at equilibrium $\delta \mathbf{z} = 0$ forever. Recall that the parameter-dependent stage cost is defined as $c_D^i(x^i; \theta^i) = -w_{\theta^i}^i \mathbf{1}[x^i \in \mathcal{T}_{\theta^i}]$ where $w_{\theta^i}^i > 0$ is the cost weight that encodes agent i 's degree of preference to go through a toll booth.

²https://github.com/SafeRoboticsLab/opinion_game

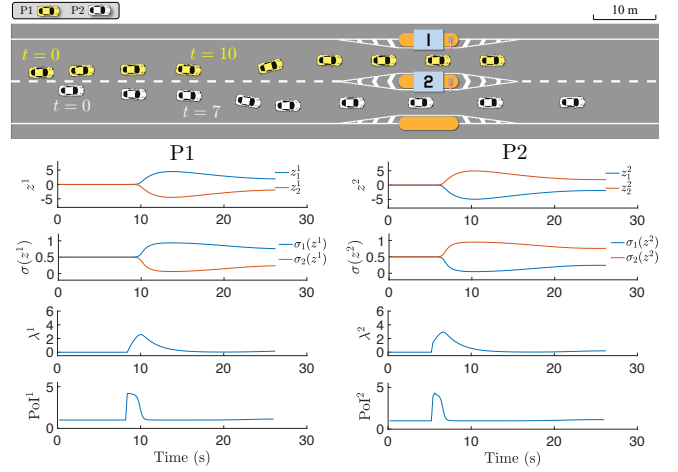


Fig. 2: Agents' state, opinion (original and softmax), attention, and PoI trajectories using the L0-QMDP policy with homogeneous cost weights. Vehicle snapshots are plotted every 3 seconds.

Homogeneous cost weights. We first examine a case in which both vehicles had the same cost weight $w_1^1 = w_2^1 = w_1^2 = w_2^2 = 15$. The resulting closed-loop state, opinion, attention, and PoI trajectories using the L0-QMDP policy are plotted in Fig. 2. At the beginning of the simulation, agents' opinions were constantly neutral since the vehicles were farther away from the toll stations and the parameter-dependent cost $c_D^i(x^i; \theta^i)$ evaluated to 0 for both options, hence the subgame values were identical. This validated Corollary 1. As $c_D^i(\cdot)$ started to produce nonzero rewards for car 1 at around $t = 10$ s and car 2 at around $t = 7$ s, the attentions driven by the PoI spiked up, agents rapidly formed an opinion, and both cars safely passed through a toll station, which empirically verified Theorem 1.

Heterogeneous cost weights. Next, we consider a set of heterogeneous cost weights $w_1^1 = 40$, $w_2^1 = 50$, $w_1^2 = 50$, $w_2^2 = 40$, encoding that car 1 prefers to go through toll booth 2, and car 2 prefers to go through toll booth 1. The state and opinion trajectories when both cars are using the L0-QMDP policy are plotted in Fig. 3. Due to the interference from car 2 (cutting in front of car 1), car 1 formed an opinion to stay in the left lane and went through the less preferred toll station. In another trial under the same initial condition, we applied the L1-QMDP policy to car 1 while keeping the L0-QMDP policy for car 2. The resulting trajectories are shown in Fig. 4. By leveraging the active opinion manipulation feature of the L1-QMDP policy, car 1 was able to plan a more efficient trajectory toward its preferred toll booth. We also simulated the above scenario with both cars applying the L1-QMDP policy. Interestingly, we observed similar opinion formation and physical trajectories to Fig. 4 in this case, despite the agents' incorrect assumption of their opponent's cognitive level. In general, however, if the opponent agent's behavior deviates significantly from the L0-QMDP or L1-QMDP assumption, unexpected opinion evolution or optimized trajectories could appear as a result of applying Alg. 1.

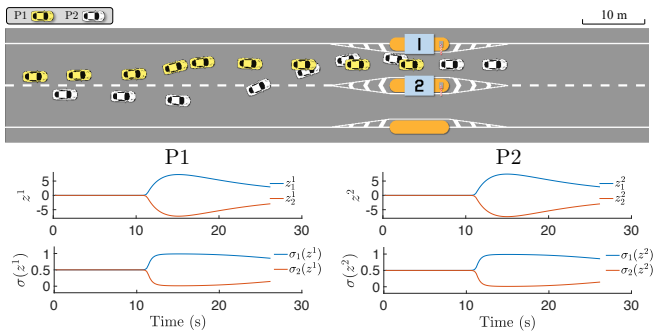


Fig. 3: Agents' state and opinion (original and softmax) trajectories using the L0-QMDP policy with heterogeneous cost weights.

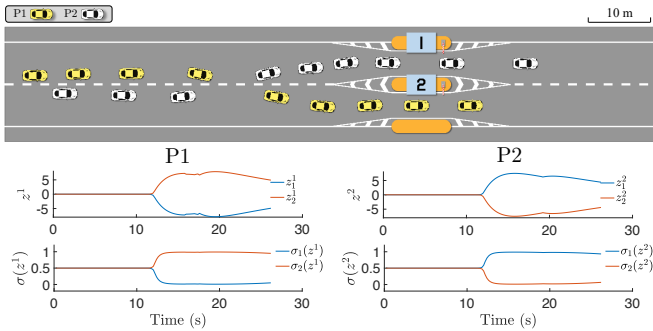


Fig. 4: Agents' state and opinion (original and softmax) trajectories with heterogeneous cost weights. Car 1 uses the L1-QMDP policy and car 2 uses the L0-QMDP policy.

VII. CONCLUSIONS

We proposed a principled theoretical and algorithmic framework for synthesizing game-induced nonlinear opinion dynamics (GiNOD) based on the value functions of dynamic games under different agent intent parameters. We provided a detailed stability analysis for GiNOD in the two-player two-option case. And we developed a trajectory optimization algorithm that uses opinions evolved through GiNOD as guidance. Future works include generalizing the stability analysis to the multi-player multi-option case and demonstrating our approach with robotic hardware.

REFERENCES

- [1] A. Bizyaeva, A. Franci, and N. E. Leonard. Nonlinear opinion dynamics with tunable sensitivity. *IEEE Transactions on Automatic Control* 68.3, 2023.
- [2] T. Başar and G. J. Olsder. *Dynamic Noncooperative Game Theory*. London: SIAM, 1998.
- [3] D. Fridovich-Keil et al. Confidence-aware motion prediction for real-time collision avoidance. *The International Journal of Robotics Research* 39.2-3, 2020.
- [4] X. Liu, L. Peters, and J. Alonso-Mora. Learning to play trajectory games against opponents with unknown objectives. *IEEE Robotics and Automation Letters* 8.7, 2023.
- [5] T. Zrnic, E. Mazumdar, S. Sastry, and M. Jordan. Who leads and who follows in strategic classification? *Advances in Neural Information Processing Systems*. Vol. 34. 2021.
- [6] L. Peters, D. Fridovich-Keil, C. J. Tomlin, and Z. N. Sunberg. Inference-based strategy alignment for general-sum differential games. *Proc. 19th International Conference on Autonomous Agents and MultiAgent Systems*. Auckland, New Zealand, 2020.
- [7] P. Trautman and A. Krause. Unfreezing the robot: Navigation in dense, interacting crowds. *IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2010.

- [8] J. F. Fisac, E. Bronstein, E. Stefansson, D. Sadigh, S. S. Sastry, and A. D. Dragan. Hierarchical game-theoretic planning for autonomous vehicles. *IEEE International Conference on Robotics and Automation (ICRA)*. 2019.
- [9] D. Sadigh, N. Landolfi, S. S. Sastry, S. A. Seshia, and A. D. Dragan. Planning for cars that coordinate with people: leveraging effects on human actions for planning and active information gathering over human internal state. *Autonomous Robots* 42.7, 2018.
- [10] A. Zanardi, E. Mion, M. Bruschetta, S. Bolognani, A. Censi, and E. Frazzoli. Urban driving games with lexicographic preferences and socially efficient Nash equilibria. *IEEE Robotics and Automation Letters* 6.3, 2021.
- [11] Y. Li, G. Carboni, F. Gonzalez, D. Campolo, and E. Burdet. Differential game theory for versatile physical human-robot interaction. *Nature Machine Intelligence* 1.1, 2019.
- [12] S. Zazo, S. V. Macua, M. Sánchez-Fernández, and J. Zazo. Dynamic potential games with constraints: Fundamentals and applications in communications. *IEEE Transactions on Signal Processing* 64.14, 2016.
- [13] D. Fridovich-Keil, E. Ratner, L. Peters, A. D. Dragan, and C. J. Tomlin. Efficient iterative linear-quadratic approximations for nonlinear multi-player general-sum differential games. *IEEE International Conference on Robotics and Automation (ICRA)*. 2020.
- [14] E. A. Hansen, D. S. Bernstein, and S. Zilberstein. Dynamic programming for partially observable stochastic games. *Proc. 19th National Conference on Artificial Intelligence*. San Jose, CA, 2004.
- [15] H. Hu and J. F. Fisac. Active uncertainty reduction for human-robot interaction: an implicit dual control approach. *Algorithmic Foundations of Robotics XV*. Springer, 2022.
- [16] M. L. Littman, A. R. Cassandra, and L. P. Kaelbling. Learning policies for partially observable environments: Scaling up. *Machine Learning Proceedings 1995*. San Francisco (CA): Morgan Kaufmann, 1995.
- [17] H. Hu, K. Nakamura, and J. F. Fisac. SHARP: Shielding-aware robust planning for safe and efficient human-robot interaction. *IEEE Robotics and Automation Letters* 7.2, 2022.
- [18] C. Cathcart, M. Santos, S. Park, and N. E. Leonard. Proactive opinion-driven robot navigation around human movers. *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2023.
- [19] S. Park, A. Bizyaeva, M. Kawakatsu, A. Franci, and N. E. Leonard. Tuning cooperative behavior in games with nonlinear opinion dynamics. *IEEE Control Systems Letters* 6, 2022.
- [20] Z. Chen. Bayesian filtering: From Kalman filters to particle filters, and beyond. *Statistics* 182.1, 2003.
- [21] J. F. Kolen and S. C. Kremer. "Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-Term Dependencies". *A Field Guide to Dynamical Recurrent Networks*. 2001.
- [22] R. P. Abelson. Mathematical models of the distribution of attitudes under controversy. *Contributions to Mathematical Psychology*. Ed. by N. Frederiksen and H. Gulliksen. Vol. 14. Holt, Rinehart, & Winston, 1964.
- [23] A. Bizyaeva, A. Franci, and N. E. Leonard. Multi-topic belief formation through bifurcations over signed social networks. 2023.
- [24] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association* 69.345, 1974.
- [25] E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. *Computer Science Review* 3.2, 2009.
- [26] R. A. Horn, R. A. Horn, and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1994.
- [27] D. O. Stahl II and P. W. Wilson. Experimental evidence on players' models of other players. *Journal of Economic Behavior & Organization* 25.3, 1994.
- [28] X. Zhang, A. Liniger, and F. Borrelli. Optimization-based collision avoidance. *IEEE Transactions on Control Systems Technology* 29.3, 2020.
- [29] J. Bradbury et al. JAX: composable transformations of Python+NumPy programs, 2020.
- [30] J. A. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl. CasADi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation* 11.1, 2019.