# MIT Open Access Articles

## Semi-Cooperative Control for Autonomous Emergency Vehicles

**Massachusetts Institute of Technology**

# Semi-Cooperative Control for Autonomous Emergency Vehicles

Noam Buckman[1], Wilko Schwarting[1], Sertac Karaman[2], and Daniela Rus[1]

*Abstract*— **Autonomous control of an emergency vehicle will save lives through faster transport and shorter response. Towards this goal, it must overcome the challenge of interacting with existing human drivers on the road. We present a game-theoretic approach for semi-cooperative control of an autonomous emergency vehicle that can interact efficiently with humans on the road. We model the interactions between autonomous and human driven cars with Social Value Orientation, a metric from social psychology, that allows the controller to leverage their influence on the trajectories of neighboring human drivers. In addition, by using a modified version of iterative best response, we direct the algorithm to converge to Nash equilibria that are cooperative. We demonstrate the efficacy of our algorithm in simulations of drivers in traffic, with a variety of traffic densities and driver personalities. In simulations of prosocial human drivers, our algorithm provides an 8% improvement in distance-traveled compared to egoistic human drivers.**

## I. INTRODUCTION

Autonomous vehicles show the potential to improve the efficiency and safety of transportation. The benefits are even more promising for emergency vehicles. Consider an autonomous ambulance that can arrive at the hospital faster, with additional medical personnel to help the patient, and a reduction in accidents with other vehicles [1], [2]. However, to achieve these performance gains in the near term, the autonomous ambulance must be able to interact with human drivers on the road and leverage the impact of their own actions on the actions of other drivers. This creates a challenging control problem for the planner: it must simultaneously find safe and efficient control inputs to avoid collisions while anticipating various levels of cooperation with humans.

This work focuses on designing controllers that allow the autonomous vehicle to seamlessly cooperate with other agents on the road, without the need for strict traffic rules or full control of the surrounding vehicles. Current approaches focus on either predicting human trajectories using learning-based approaches [3] or assuming simple dynamics for obstacle-avoidance [4]. Other approaches that consider the system-wide optimization are either restricted to full team control, as in vehicle platooning [5], or game theory [6], [7], where humans are modeled as competitive. In contrast,

[1]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA [nbuckman, wilkos, rus] at mit.edu

[2]Laboratory of Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA sertac@mit.edu
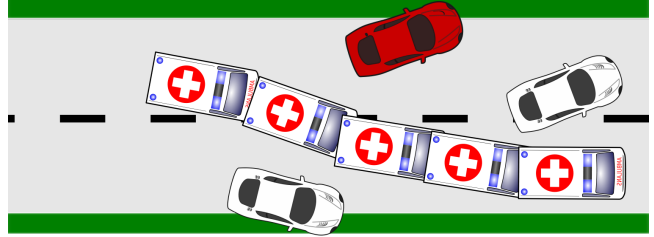
Fig. 1: Humans cooperate with the emergency vehicle by modifying their own control inputs to accommodate the ambulance's traversal through traffic.

we take inspiration from social psychology and naturalistic driving data to model human drivers as semi-cooperative [8], enabling an autonomous ambulance to work with human drivers on the road.

Our approach formulates the control problem for each vehicle on the road as a non-linear optimization that includes both efficiency costs and safety constraints that can be solved by a nonlinear model predictive control (MPC). However, rather than modeling humans as simply self-interested rational agents who consider only their own performance, we incorporate a metric from social psychology behavioral decision theory, Social Value Orientation, which models each driver's willingness to cooperate with a neighboring agent. This pairwise metric leads to a *semi-cooperative* utility function for each agent that linearly combines its own reward with the reward of other agents, including the ambulance. Finally, we solve for control inputs that satisfy Nash Equilibrium using a modified version of Iterative Best Response, where vehicles can imagine shared control with other agents. This dynamic game yields a controller for the ambulance that can plan for semi-cooperative drivers, leading to highly interactive emergent behavior where the ambulance and human drivers work together to allow the ambulance to pass quickly and safely.

### A. Contributions

In summary, the contributions of this paper are:

1) A semi-cooperative optimal control formulation for autonomous control that models human drivers using Social Value Orientation
2) An iterative best response algorithm that considers shared control of neighboring vehicles to obtain trajectories that are cooperative and Nash Equilibrium
3) Validation of our approach in a simulated multi-lane highway under a variety of human personalities, traffic

densities, and number of vehicle (up to 100 drivers), with baseline comparisons and ablation study.

## II. RELATED WORKS

### A. Control Around Humans

Recently, learning-based control has shown promise in utilizing naturalistic data to generate control policies for autonomous vehicles [9], [10], [11] and predicting human trajectories [12]. While learning-based approaches allow robots to navigate around humans, they do not provide insight on how to cooperatively work with humans. For predicting human pedestrian and driver trajectories, Social-LSTM [12], [13], [14] uses recurrent neural networks to learn from previous trajectories and predict future motion. [15] uses inverse reinforcement learning to learn a hierarchical model of learned rules from driving demonstrations. For a probabilistic approach, [16], [17] incorporate agent prediction with planning for autonomous vehicles. In Socially Aware CADRL [3], a reinforcement learning algorithm simulates multiple agents to generate collision avoidance policies while complying with social norms by biasing the reward to a predetermined set of social rules. Alternatively, explicit models for human driving, such as the Intelligent Driver Model (IDM) [18] for acceleration and MOBIL [19] for lane changes, can be used to predict the high-level maneuvers of humans, however, they struggle with complex multi-agent interactions since they are designed primarily for "normal" speed-following settings. In all these approaches, while robots can tolerate humans, they do not take advantage of the potential to cooperate with humans drivers. In addition, they are fragile to scenarios where fixed rules do not apply and limited training data is available.

### B. Distributed MPC

The most efficient and cooperative approach would be to control all the vehicles on the road, something that is only possible if all the vehicles are autonomous and choose to form an explicit team. In such scenarios, a platoon of vehicles or team of robots share a joint cost function and can implement a distributed model predictive control across the team. In [20], they propose a Decentralized Model Predictive Control (MPC) framework for a team of robots to jointly optimize a shared cost function, while sharing plans with neighboring agents. Similarly, [21], [22], [23] proposed variants of distributed MPC that include shared collision avoidance constraints to achieve complex formations and maneuvers. In all of these, each individual agent is assumed to share a cost function and constraint. Human drivers, however, can not join explicit teams and rarely share a single joint utility function across vehicles.

More recently, MPC approaches have been applied specifically to self-driving cars in traffic where an agent's own utility is considered. [24], [25], [26], [27] applied distributed MPC to controlling fully autonomous vehicles at intersections and highway-merging. Additionally, MPC has been used to control platoons of autonomous vehicles driving on highways with the goal of maximizing road efficiency and

safety [28], [5]. In these approaches, the vehicles maintain their own agent-specific cost, however, must also maintain vehicle ordering or priority. This extra level of coordination limits our ability to model more competitive behaviors and requires full autonomy and centralized coordination.

### C. Game-Theoretic Planners

For more competitive driving scenarios, game theory provides a framework for designing controllers that can consider each agent's competing utility function. In addition, it does not assume an explicit team or centralized planner in generating control. [29] implemented a best response algorithm for agile interactions between two autonomous cars, where each vehicle optimizes their own distance traveled and an inter-vehicle cost to maintain a desired distance between the two vehicles. [6] extends this to a more competitive scenario, where the autonomous vehicles are directly competing against each other for distance traveled. They use a modified iterative best response with an additional sensitivity cost that accounts for the potential negative effect of the adversarial agent. In contrast to [29] and [6], we motivate our cost functions based on human preferences, assuming more cooperative cost functions that are derived by considering each agent's own performance and effort cost.

In [30], an autonomous vehicle interacts with humans by modeling the interaction as a Stackelberg game, which is then used to learn a utility function of the human using inverse reinforcement learning. In [7], a Nash game is used instead to remove the assumption that an autonomous vehicle has a strategic advantage in selecting actions. The controller is broken down into an offline high-level strategic planner and a low-level tactical MPC for control. Our method combines these two levels so that we can explore more interactive trajectories that are dynamically feasible. In addition, we do not assume that the agents are inherently competitive and focus on cooperating with multiple humans at a time. [8] first introduced the concept of modeling human drivers using Social Value Orientation. They demonstrate the usefulness of semi-cooperative rewards for prediction of vehicles and validate using highway data. In this work, we take advantage of SVO to enhance control of the ego vehicle in highly interactive scenarios where agents may work together in a semi-cooperative manner to help the ambulance. In addition, we introduce a pairwise SVO to account for heterogeneous cooperation between different agents.

## III. PROBLEM STATEMENT

An autonomous emergency vehicle $i$ is driving in a traffic environment that contains surrounding human-driven vehicles $j = 1 \ldots N$ where $N$ is the number of vehicles in the agents planning horizon. The goal is to design controls $\boldsymbol{u}_i$ for the autonomous vehicle that is safe (i.e. collision free) and semi-cooperative so that it can quickly traverse through traffic (i.e. travel the greatest distance). We denote the set of vehicles that excludes an ego vehicle $i$, simply as $\neg i$ (or ado vehicles) and the ambulance as $i = 0$. We denote a trajectory, a sequence of control inputs and states over time, as $\boldsymbol{u}_i, \boldsymbol{x}_i$.
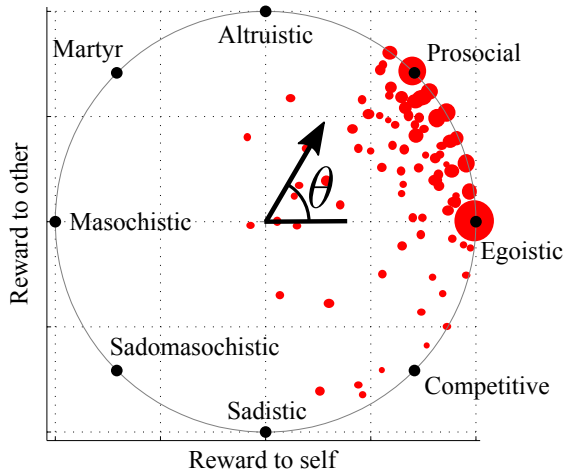
Fig. 2: Social Value Ring. The Social Value Orientation of human subjects are plotted (red), where size of circle corresponds to proportion of subjects. Adapted from [32]

We assume that a model of the non-linear dynamics of each vehicle, $\dot{\boldsymbol{x}}_i = f_i(\boldsymbol{x}_i, \boldsymbol{u}_i)$ is known among all vehicles on the road.

### A. Agent-Specific Reward

We first begin with a typical, non-social reward model for human drivers, where the reward is primarily a function of the agent's control and state. We denote this agent-specific reward as

$$R_i(\boldsymbol{u}_i, \boldsymbol{x}_i, \boldsymbol{x}_{\neg i}) = P_i(\boldsymbol{x}_i) + E_i(\boldsymbol{u}_i) + C_i(\boldsymbol{x}_i, \boldsymbol{x}_{\neg i}) \quad (1)$$

where $P_i(\boldsymbol{x}_i)$ is the performance reward as they travel down the lane (e.g. how fast they progress), $E_i(\boldsymbol{u}_i)$ is a control effort cost related to how much they accelerate and steer. $C_i(\boldsymbol{x}_i, \boldsymbol{x}_{\neg i})$ is an (optional) interagent cost between the ego vehicle ($i$) and all other ado vehicles ($\neg i$). Examples of interagent costs include collision costs based on the distance between vehicles or risk costs that are a function of both velocity and position [31].

To account for path tracking along an arbitrary lane, we augment the state of the vehicle with an additional state $s$ that can parameterize the desired path $\boldsymbol{\xi}_D$ such that $\boldsymbol{\xi}_D = \begin{bmatrix} X_D(s), Y_D(s), \Phi_D(s) \end{bmatrix}^T$ where $X_D$, $Y_D$, $\Phi_D$ are the 2D components of the path at point $s$ on the path. For simplicity, we include $s$ in the state $\boldsymbol{x}_i$ and note that the performance reward $P_i(\boldsymbol{x}_i)$ will primarily be a function of $s$ and $\boldsymbol{\xi}_D$.

### B. Modeling Human Cooperation

For the ambulance to efficiently navigate through traffic, it must be able to model and anticipate the planning and cooperation of human drivers on the road. Our approach takes inspiration from social psychology which observes that humans are not purely egoistic, as typically assumed, but rather show certain traits of cooperation and even altruism. Specifically, this human personality or Social Value Orientation, $\theta_i$ [33] encodes a human's willingness to consider the other person's reward. The value function of an agent $i$

becomes a combination of their own reward $R_i$ and the reward of others $R_j$. This can alternatively be thought as a joint utility function that *each individual human* considers when optimizing their actions $V_i = \cos\theta_i R_i + \sin\theta_i R_j$. Figure 2 shows the Social Value Ring, along with real-world data from human subjects playing a monetary trade-off game [32].

In [8], SVO is successfully used to predict driver behavior by first calculating a driver's SVO from their past trajectories and then using SVO to predict the future actions of the agents, observing a range of cooperative and competitive SVOs in highway data. While other behavior models exist [34], [35], SVO provides a compact representation of human behavior that works well in practice [8] and is a standard metric for cooperation in the social psychology community. Emergency vehicles in particular have the potential to improve their own performance by cooperating with human drivers since humans are observed to increase their SVO in emergency scenarios [36]. For that reason, in this work we focus on incorporating SVO in a longer horizon planner so that the autonomous vehicle can achieve more cooperative trajectories.

In addition, in contrast to the psychology games of [33] which include only two participants, vehicles on the road interact with many other vehicles at the same time and may have different personalities for each vehicle. For example, a human driver may act more cooperatively towards an ambulance than a neighboring driver. Similarly, drivers may feel differently towards autonomous drivers than human drivers. To account for the pairwise nature of social preferences, we represent the SVO of agent $i$ as a pairwise property $\theta_{ij}$ which captures the agent's personality with respect to agent $j$. This allows a human $i$ to be cooperative with an ambulance $\theta_{i,j=0} = \pi/4$ while egoistic towards other humans $\theta_{i,j\neq 0} = 0$. We augment the social value function in [8] to account for multiple vehicles and SVOs

$$V_i = \frac{1}{N} \sum_{j \neq i} \cos\theta_{ij} R_i + \sin\theta_{ij} R_j \quad (2)$$

where $V_i$ is agent $i$'s utility, $\theta_{ij}$ is the Social Value Orientation between agent $i$ and $j$, $R_i$ and $R_j$ are the respective agent-specific rewards. For brevity, (2) does not include the control and state inputs of each agent, however, by expanding $R_i$ and $R_j$ we can see that a single agent's utility $V_i$ will be a function of the surrounding control inputs and states.

### C. Nash Equilibrium Conditions

We assume that every driver on the road is rational and thus chooses their actions to maximize their own value function

$$\boldsymbol{u}_i^* = \arg\max_{\boldsymbol{u}_i} V_i(\cdot) \quad (3)$$

where $V_i$ is the social value function and $\boldsymbol{u}_i$ are all possible control inputs of the ego vehicle. Note that we do not assume a single, explicit joint reward function across all vehicles, but rather a value function for each agent that may include the reward of other drivers. During planning, we assume that

the ambulance has learned the reward functions of the other vehicles, $V_{\neg i}$ which includes the agent-specific reward $R_i$ and each agent's SVO $\theta_{ij}$, similar to [8], [30]

We desire control policies that are stable with respect to the other agents. More specifically, we assume that controls executed by agents meet the Nash equilibrium constraint

$$V_i(u_i^*, u_{\neg i}^\star) \geq V_i(u_i, u_{\neg i}^\star) \quad \forall u_i \neq u_i^* \tag{4}$$

where $u_{\neg i}^\star$ are the optimal control policies of the other agents. Intuitively, if the Nash Equilibrium condition is met, agents will not have an incentive to deviate from their chosen control actions. In general, it is very difficult to solve for a (4) and the subsequent section, we present our approach for obtaining a local Nash Equilibrium.

## IV. ITERATIVE BEST RESPONSE WITH SHARED CONTROL

### A. Obtaining a Nash Equilibrium Controller

The non-linear optimization for each agent $i$ is solved using model predictive control (MPC), where at step $m$ of MPC an optimization for a subsequence of control input $\boldsymbol{u}_i^m$ is formulated

$$\boldsymbol{u}_i^m = \arg\max_{\boldsymbol{u}_i} V_i(\boldsymbol{u}_i, \boldsymbol{u}_{\neg i}^m) \tag{5}$$

$$\text{s.t.} \quad \dot{\boldsymbol{x}}_i = f_i(\boldsymbol{x}_i, \boldsymbol{u}_i) \tag{6}$$

$$G_i(\boldsymbol{x}_i^m, \boldsymbol{x}_i^m, \boldsymbol{x}_{\neg i}^m, \boldsymbol{x}_{\neg i}^m) \geq 0 \tag{7}$$

where $\boldsymbol{u}_i^m$ and $\boldsymbol{u}_{\neg i}^m$ are the control sequences for the ego agent and ado agents at MPC iteration $m$, respectively, and $G_i(\boldsymbol{x}_i, \boldsymbol{u}_i, \boldsymbol{x}_{\neg i}, \boldsymbol{u}_{\neg i})$ contains all agent-specific constraints such as actuation limits and inter-agent constraints (such as collision avoidance). For brevity, we exclude $\boldsymbol{x}_i, \boldsymbol{x}_{\neg i}$ from utility $V_i$ since they can be inferred by the control inputs $\boldsymbol{u}_i$, $\boldsymbol{u}_{\neg i}$ and dynamics $f_i$. As is typical in MPC, at each round $m$, the agent solves for a subtrajectory $\boldsymbol{u}_i^m$ over time horizon $T$, then executes a subset of control inputs $n_{\text{exec}}$ and re-initializes the optimization. This receding horizon optimization makes the optimization more computationally efficient and allows for replanning in case of uncertainty in dynamics.

However, just solving this optimization does not ensure that $\boldsymbol{u}_i^m$ satisfies the Nash Equilibrium. In general, it is difficult to obtain a Nash Equilibrium controller for an arbitrary problem. A popular algorithm for obtaining a local Nash Equilibrium is Iterative Best Response (IBR) (Algorithm 1) where each agents solves a relaxed open-loop Nash game. At each iteration $k$ of IBR, the agent solves their own best response $\boldsymbol{u}_i^{m,k}$ while fixing the controllers of the other agents $\bar{\boldsymbol{u}}_{\neg i}^{m,k}$

$$\boldsymbol{u}_i^{m,k} = \arg\max_{\boldsymbol{u}_i} V_i(\boldsymbol{u}_i, \bar{\boldsymbol{u}}_{\neg i}^{m,k-1}) \tag{8}$$

where $k$ is the round of IBR and $m$ is the current step in the MPC. Note that at a given IBR iteration $k$, an ego vehicle's best response is solved with respect to potentially sub-optimal ado vehicle controls, $\boldsymbol{u}_{m,\neg i}^k \neq \boldsymbol{u}_{m,\neg i}^\star$. However, as multiple iteration of IBR proceeds, the controls of each agent improves, approaching a locally optimal solution.

---

**Algorithm 1** Iterative Best Response

$T$: planning horizon
$\boldsymbol{u}_{i,m} := [u_{i,m\Delta T}, u_{i,(m+1)\Delta T}, \ldots, u_{i,(m+T)\Delta T}]$
$\boldsymbol{u}_{\neg i,m}^0 = \text{Extend}(\boldsymbol{u}_{\neg i,m-1})$ // Extend previous MPC
**for** $k = 1 \ldots n_{\text{IBR}}$ **do**
    **for** $i = 0 \ldots N_{agents}$ **do**
        $\boldsymbol{u}_{i,m}^k = \arg\max_{\boldsymbol{u}_i} V_i(\boldsymbol{u}_i, \boldsymbol{u}_{\neg i,m}^{k-1})$
    **end for**
**end for**
**return** $\boldsymbol{u}_{i,m}^{n_{\text{IBR}}}$

---

**Algorithm 2** Iterative Best Response with Shared Control

$T$: planning horizon
$m$: MPC step
$n$: Maximum number of agents in Shared Control
$\boldsymbol{u}_i^m := [u_{i,m\Delta T}, u_{i,(m+1)\Delta T}, \ldots, u_{i,(m+T)\Delta T}]$
$\boldsymbol{u}_{\neg i}^{m,0} = \text{Extend}(\boldsymbol{u}_{\neg i}^{m-1})$ // Extend previous MPC
**for** $k = 1 \ldots n_{\text{IBR}}$ **do**
    **for** $i = 0 \ldots N_{agents}$ **do**
        $\mathcal{N}_{sc}^i \leftarrow \text{ClosestVehiclesBehind}(i, n, m)$
        $\boldsymbol{u}_i^{m,k} = \arg\max_{\boldsymbol{u}_i, \boldsymbol{u}_{j \in \mathcal{N}_{sc}^i}} V_i(\boldsymbol{u}_i, \boldsymbol{u}_j, \boldsymbol{u}_{\neg(i \cup \mathcal{N}_{sc}^i)}^{m,k})$
    **end for**
**end for**
$\boldsymbol{u}_i^m \leftarrow \boldsymbol{u}_i^{m,n_{\text{IBR}}}$
**return** $\boldsymbol{u}_i^m$

---

In contrast to [29], we run multiple rounds of IBR at each step $m$ of the MPC to ensure that the ado vehicle controls are "up-to-date" with respect to the ego vehicle's controls. This allows the ego vehicle to plan for more interactive maneuvers and not just reacting to the ado vehicle's past actions. In the case of an ambulance, we can further add structure and assume the ambulance takes the first action in the IBR, since it initiates the cooperative maneuvers by either explicitly signalling an emergency or implicitly, by simply entering the field of view of the other vehicles. While IBR does not guarantee convergence to a solution, we show in the results in Sec. V that it converges to a fixed point.

The benefits of Iterative Best Response are two-fold: it reduces the optimization variables in the MPC (5) by only solving a single vehicle's controls at each round of IBR, and it provides a level of confidence to the ambulance, by ensuring that human drivers do not have an incentive to deviate from the their predicted trajectories. This is critically important because if ado vehicles deviate from their plan $\boldsymbol{u}_{\neg i}^m$ then the ego vehicle's trajectory $\boldsymbol{x}_i^m$ may no longer be collision free.

### B. Imagining Shared Control

One limitation of the open-loop relaxation in IBR is its limited ability to anticipate the response of other vehicles to one's own actions, since IBR fixes the controls of ado vehicles at each round of the optimization. This can lead IBR to converge to a Nash Equilibrium that includes little cooperation or consideration of the other agent's action. The

following lemma demonstrates such a scenario, where IBR converges to a Nash Equilibrium that is agnostic to the neighboring vehicle's utility.

*Lemma 1:* Consider only two agents, and $C_i(\boldsymbol{x}_i, \boldsymbol{x}_j) = C_j(\boldsymbol{x}_i, \boldsymbol{x}_j) = 0$, then iterative best response converges to a solution that ignores $V_j$ (and thus $\boldsymbol{u}_j, \boldsymbol{x}_j$) for all $\theta_{ij}$

*Proof:* We first substitute the agent-specific reward function (1) with the social reward function (2)

$$V_i = \big(P_i(\boldsymbol{x}_i) + E_i(\boldsymbol{u}_i)\big)\cos\theta_{ij} + \big(P_j(\boldsymbol{x}_j) + E_j(\boldsymbol{u}_j)\big)\sin\theta_{ij}. \tag{9}$$

For IBR, the ado vehicle's control is fixed as $\bar{\boldsymbol{u}}_j, \bar{\boldsymbol{x}}_j$ and agent $i$'s optimization (8) becomes

$$
\begin{aligned}
\boldsymbol{u}_i^* &= \arg\max_{\boldsymbol{u}_i} \big(P_i(\boldsymbol{x}_i) + E_i(\boldsymbol{u}_i)\big)\cos\theta_{ij} \\
&\quad + \big(P_j(\bar{\boldsymbol{x}}_j) + E_j(\bar{\boldsymbol{u}}_j)\big)\sin\theta_{ij} \\
&= \arg\max_{\boldsymbol{u}_i} \big(P_i(\boldsymbol{x}_i) + E_i(\boldsymbol{u}_i)\big)\cos\theta_{ij} \\
&= \arg\max_{\boldsymbol{u}_i} \big(P_i(\boldsymbol{x}_i) + E_i(\boldsymbol{u}_i)\big) \\
&\quad \text{s.t. } G(\boldsymbol{x}_i, \boldsymbol{u}_i, \boldsymbol{x}_j, \boldsymbol{u}_j) \geq 0.
\end{aligned}
$$

where $\bar{\boldsymbol{x}}_j, \bar{\boldsymbol{u}}_j$ and $\theta_{ij}$ become constants and can be ignored when optimizing over $\boldsymbol{u}_i$. Note that while $\boldsymbol{x}_j$ and $\boldsymbol{u}_j$ are included in the final constraint $G(\cdot)$, they do not enter the value function. Which means that while the agent $i$'s optimization is **aware** of the ado vehicle trajectories, it will not **value** their trajectories since value function is independent of $\theta_{ij}$ and $\boldsymbol{x}_j, \boldsymbol{u}_j$. ∎

To counter this effect and encourage a cooperative Nash equilibrium, we allow vehicles to "imagine" sharing control with the ambulance during the first $n_{sc} < n_{\text{IBR}}$ rounds of iterative best response to encourage considering a more cooperative Nash Equilibrium. Specifically, each agent selects a neighborhood of vehicles $\mathcal{N}_{sc}^i$ to control during iterative best response. The modified IBR is now

$$\boldsymbol{u}_i^* = \arg\max_{\boldsymbol{u}_i, \ \boldsymbol{u}_j \forall j \in \mathcal{N}_{sc}^i} V_i\big(\boldsymbol{u}_i, \boldsymbol{u}_j, \bar{\boldsymbol{u}}_{\neg(i \cup \mathcal{N}_{sc}^i)}\big) \tag{10}$$

where $\boldsymbol{u}_i^*$ is the new control trajectory for the planning agent $i$, $\boldsymbol{u}_j$ is the "imagined" control of the other agents, and $\boldsymbol{u}_{\neg(i \cup \mathcal{N}_{sc}^i)}$ are the fixed control inputs of the remaining ado vehicles. Algorithm 2 describes the full Iterative Best Response with Shared Control algorithm that is executed at each round of MPC. The control trajectories of each vehicle are initialized with the solution from the previous time-step and extended assuming a line-following controller. Each vehicle selects $n_{sc}$ vehicles behind it, including the ambulance, and chooses the optimal best response control. We limit $\mathcal{N}_{sc}$ to vehicles *behind* the best response vehicle so to not optimistically control a vehicle in front $i$. Finally, the "best response" vehicle only executes its own control $\boldsymbol{u}_i^*$ and does not store $\boldsymbol{u}_j \ \forall j \in \mathcal{N}_{sc}^i$, since it is only used for guiding their own control.

### C. Limitations

Our approach assumes that each agent is playing the same Nash game with the autonomous vehicle. If some humans do
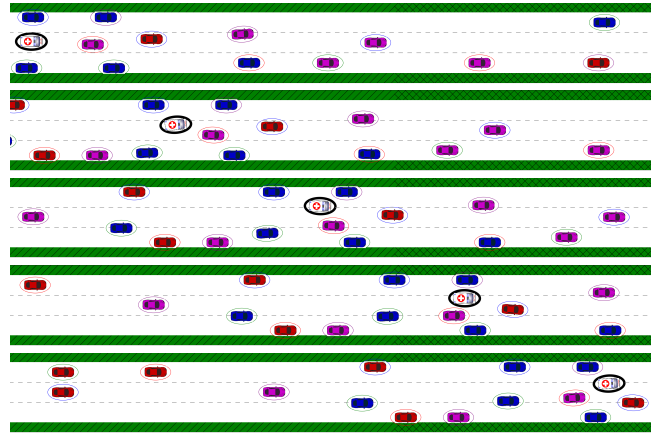


Fig. 3: Autonomous ambulance travels between vehicles while anticipating that prosocial vehicles (magenta, blue) vehicles will let them through. Egoistic vehicles (red) are less likely to move out of the way for the autonomous ambulance. Ellipses shown are used for collision avoidance

not act rationally or the prediction of each agent's SVO is incorrect, then the agents may be playing incorrect games, leading to diverging controls. In the absence of communication, our approach does not guarantee safety, however, our approach does ensure that the controls of the ambulance are dynamically feasible, collision free, and rational assuming consistency in game. Furthermore, by adding additional risk or collision costs, the autonomous ambulance can bias controls towards trajectories that provide a safety buffer.

Another limitation is that our framework requires multiple rounds of iterative best response for each agent, for each round of MPC. This limits the planning horizon and number of agents that can be included in the optimization while running quickly. While runtime is not the primary focus of this work, we can return solutions more quickly by placing some computations offline or parallelizing.

## V. RESULTS

### A. MPC Details

We implement our algorithm for an emergency vehicle traversing a highway with $N = 10, 30, 50, 100$ vehicles on roads with 2 or 3 lanes, while varying the density traffic and SVO of human drivers. The ambulance is egoistic towards all other vehicles $\theta_{ij} = 0$ whereas the surrounding vehicles have various SVOs $\theta_{i,j=0} \in \{0, \pi/6, \pi/4, \pi/2\}$ and $\theta_{i,j} = 0$ for $j \neq 0$ (i.e. humans are egoistic towards other humans). Each vehicle plans for a horizon time $T_{mpc} = 5s$ discretized by $dt = 0.2s$ and $40\%$ of the MPC is executed at a time (i.e, $2s$). For larger scale simulations, we limit the rounds of iterative best response to $n_{\text{IBR}} = 3$, shared control ends after $k = 2$ rounds of IBR, and agents can imagine up to $n_{sc} = 2$ agents in shared control. Fig. 3 shows an example maneuver executed by the ambulance, anticipating cooperative maneuvers from the altruistic agents ($\theta_{ij} = \pi/2$).

We use a contour controlling model predictive controller (CC-MPC) [4] to control the autonomous vehicle's non-
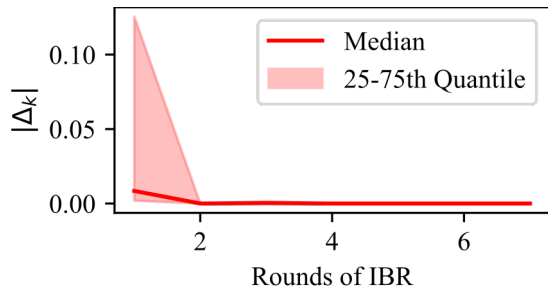
Fig. 4: Convergence of ambulance steering control. The magnitude of steering control converges to a Nash Equilibrium after two rounds of IBR.

linear dynamics while optimizing the value function of the vehicle as it traverses a lane. We utilize a kinematic bicycle model [37] to model the dynamics of each vehicle, where the state is its 2D pose $(X, Y, \Phi)$, speed $(V)$, and front wheel angle $(\delta_f)$ and the control inputs are acceleration $(V^u)$ and steering rate $(\delta_f^u)$. For collision avoidance, we approximate each car with an ellipse and compute the Minkowski sum using the minimum trace ellipse method in [38] to obtain a collision ellipse $Q_{ij}(\beta)$ for constraining the ego vehicle's position. In addition, we add a collision cost that is inversely proportional to the squared distance from the collision constraint and a cost related to lateral deviations from the lane centerline. The entire optimization is implemented in CASADI [39] with an interior point solver (IPOPT) for solving the non-convex optimization.

### B. Iterative Best Response Convergence to Nash

To measure the convergence of IBR, we increase the rounds of IBR and measure the difference in control inputs between rounds of IBR. Specifically, we inspect the first steering control command $\delta_f^{u_k}$ at each round $k$ of IBR and compute the difference between subsequent rounds of IBR

$$\Delta_k = \delta_f^{u_{k+1}} - \delta_f^{u_k} \tag{11}$$

where $\delta_f^{u_{k+1}}$ is the front wheel steering control at the $k$th round of IBR.

Fig. 4 shows the convergence of IBR for the ambulance and Fig. 5 plots the same convergence for the six closest vehicles to the ambulance for 9 random experiments (with 50 rounds of MPC each). We see the ambulance converges after 2 rounds of IBR and ado vehicles in 3 rounds.

### C. Ablation Studies

We vary the size of shared control neighborhood $|\mathcal{N}_{sc}^i|$ and rounds of shared control $n_{sc}$ to understand the effect on the level of cooperation between ambulance and surrounding vehicles. In Fig. 7a, as number of vehicles in the shared control neighborhood increases, the cooperation increases and the ambulance traveled additional distance. Similarly, Fig. 7b shows that as more rounds of iterative best response include shared control, the performance of the ambulance improves. Both ablation studies show that the ambulance's IBR performs better with larger shared control teams by
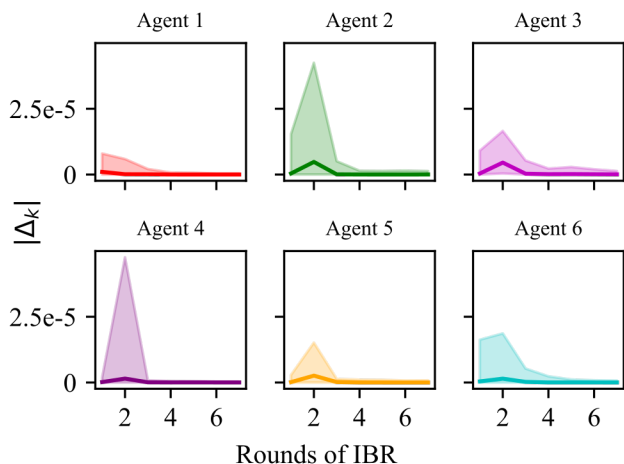


Fig. 5: Convergence of ado vehicles' steering control inputs. Human drivers converge to a Nash Equilibrium after 3 rounds of best response. Six closest vehicles shown for multiple rounds of MPC and experiments.
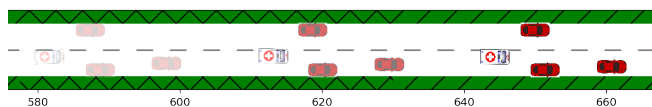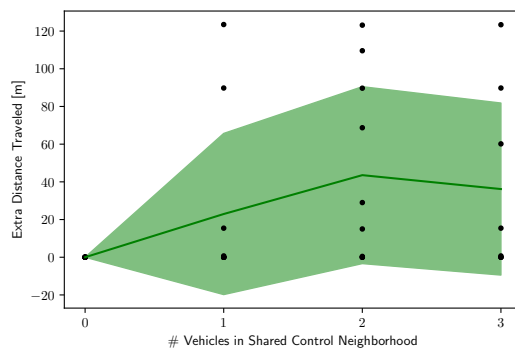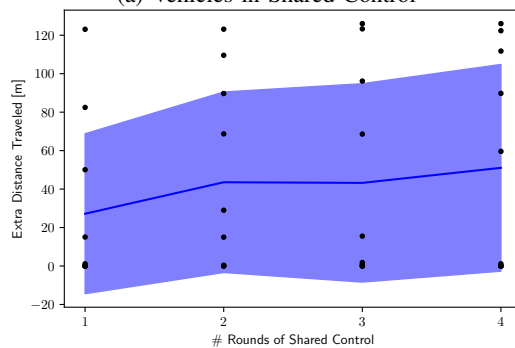


Fig. 6: Ambulance blocked behind two egoistic agents who do not consider the benefit of allowing the ambulance pass.



(a) Vehicles in Shared Control



(b) Rounds of Shared Control

Fig. 7: Ablation Study of Shared-IBR. Improvement in distance traveled compared to common baseline strategy of no shared control for random SVOs. Shaded region corresponds to values within one standard deviation of the mean.
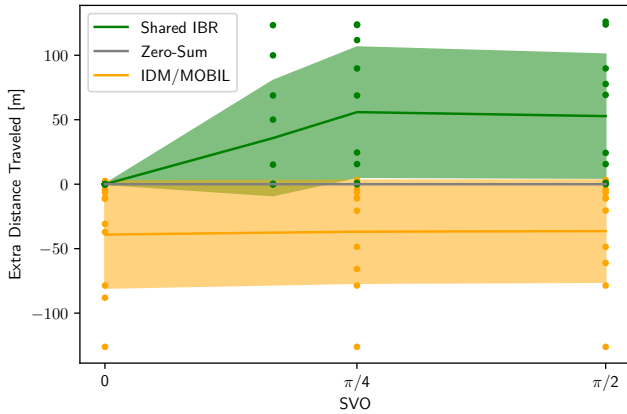
Fig. 8: Baseline comparisons. Performance of our approach (green) compared to zero-sum IBR (gray) and IDM/MOBIL (yellow) for various SVO populations. Shaded region corresponds to values within one standard deviation of the mean.

allowing ado vehicles to considering cooperative maneuvers such as proactive lane changes to allow the ambulance to pass.

### D. Comparison to Baselines

We compare our semi-cooperative SVO controller to two baselines: IDM/MOBIL [18], [19] and fully cooperative zero-sum game. In the first, human drivers' acceleration is assumed to follow the Intelligent Driver Model's (IDM) velocity-following assumption and lane changing is determined using the MOBIL criteria which considers the required accelerations of surrounding vehicles. For MOBIL, we compare against different levels of "politeness", from $p = 0.0$ to $p = 1.0$ corresponding to egoistic and altruistic SVO personalities, and solve the low-level steering control using MPC. In the zero-sum setting, we repeat our own simulations with $\theta_{ij} = -\pi/4$ which corresponds to a zero-sum reward function. Fig. 8 plots the ambulance improved performance of various approaches compared to IBR with all egoistic agents ($\theta_{ij} = 0$). IBR with Shared Control performs better than IDM/MOBIL even in environment with only egoistic agents and improves as human drivers become pro-social ($\pi/2$). Overall, we observe that IDM/MOBIL struggles to allow cooperative lane changes necessary for the ambulance to pass multiple vehicles simultaneously.

### E. Effect of Human SVO

| Ambulance | Mean (m) | Median (m) | Std. Dev. (m) |
|---|---|---|---|
| 0 (Egoistic): | 473.19 | 466.94 | 27.37 |
| $\pi/6$ (Prosocial): | 500.85 | 505.69 | 12.57 |
| $\pi/4$ (Prosocial): | 505.69 | 505.69 | 0.19 |
| $\pi/2$ (Altruistic): | 504.51 | 505.69 | 3.53 |

TABLE I: Distance traveled in 38s for different 30 human SVO over 10 simulations with varied initial positions.

We measure the performance of our algorithm under different SVOs by first generating 10 different random placement of vehicles (using a Poisson distribution) and then for

each scenario, varying the SVO of an entire population of vehicles, $\theta_i = 0, \pi/6, \pi/4, \pi/2$, for a total of 40 simulations. We limit the experiments to two lane roads with a high density of traffic to make the ambulance control problem more difficult. Table I reports the distance traveled by the ambulance in a fixed time span. We see that the ambulance travels further with higher SVOs, traveling 8% further than scenarios with all egoistic agents. One reason for this improvement is that the ambulance can get stuck behind groups of egoistic vehicles who do not have an incentive to move out of the way, as seen in Fig. 6. This also leads to high variation in performance with egoistic populations. In all these scenarios, the human drivers are not required to brake and stop, rather they are able to continue driving while cooperating with the ambulance.

### F. Effect of Vehicle Density

To test the effect of traffic density, we generate scenarios under different vehicle arrival rates and re-initialize each scenario with a different SVO ($\theta_i = 0, \pi/4$). In Table II, we report the mean distance traveled by the ambulance over a fixed run time for 18 experiments. In lower density traffic, there is little performance difference between SVO types. However, as traffic density increases, the performance gap between prosocial and egoistic populations increases.

| | Vehicle Arrival Rates | | |
|---|---|---|---|
| SVO | Low | Medium | High |
| 0 (Egoistic): | 1283m | 1223m | 1182m |
| $\pi/4$ (Prosocial): | 1283m | 1280m | 1239m |

TABLE II: Mean distance traveled by ambulance for varying traffic densities. Largest performance improvements achieved with high density traffic in prosocial populations.

## VI. CONCLUSION

We show that modeling the semi-cooperative nature of humans enables autonomous vehicles to plan along-side human drivers on the road. Central to this approach is a semi-cooperative value function for human drivers grounded in psychology and a game-theoretic algorithm that explicitly explores cooperative maneuvers, while ensuring stability. This yields a result where prosocial human drivers help the autonomous ambulance even without explicitly forming a team. This suggests there is a system-wide benefit to autonomous control of vehicles even in the absence of their full adoption by drivers.

## ACKNOWLEDGEMENTS

## REFERENCES

[1] P. T. Savolainen, K. C. Dey, I. Ghosh, T. L. N. Karra, and A. Lamb, "Investigation of Emergency Vehicle Crashes in the State of Michigan," USDOT Region V Regional University Transportation Center, Tech. Rep., 2009.

[2] H. Hsiao, J. Chang, and P. Simeonov, "Preventing emergency vehicle crashes: Status and challenges of human factors issues," *Human Factors*, vol. 60, no. 7, pp. 1048–1072, 2018, pMID: 29965790.

[3] Y. F. Chen, M. Everett, M. Liu, and J. P. How, "Socially aware motion planning with deep reinforcement learning," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 1343–1350.

[4] B. Brito, B. Floor, L. Ferranti, and J. Alonso-Mora, "Model predictive contouring control for collision avoidance in unstructured dynamic environments," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4459–4466, 2019.

[5] Y. Zheng, S. E. Li, K. Li, F. Borrelli, and J. K. Hedrick, "Distributed model predictive control for heterogeneous vehicle platoons under unidirectional topologies," *IEEE Transactions on Control Systems Technology*, vol. 25, no. 3, pp. 899–910, 2017.

[6] M. Wang, Z. Wang, J. Talbot, J. C. Gerdes, and M. Schwager, "Game theoretic planning for self-driving cars in competitive scenarios," in *Proceedings of Robotics: Science and Systems*, FreiburgimBreisgau, Germany, June 2019.

[7] J. F. Fisac, E. Bronstein, E. Stefansson, D. Sadigh, S. S. Sastry, and A. D. Dragan, "Hierarchical game-theoretic planning for autonomous vehicles," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 9590–9596.

[8] W. Schwarting, A. Pierson, J. Alonso-Mora, S. Karaman, and D. Rus, "Social behavior for autonomous vehicles," *Proceedings of the National Academy of Sciences*, vol. 116, no. 50, pp. 24 972–24 978, 2019.

[9] A. Amini, I. Gilitschenski, J. Phillips, J. Moseyko, R. Banerjee, S. Karaman, and D. Rus, "Learning robust control policies for end-to-end autonomous driving from data-driven simulation," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1143–1150, 2020.

[10] G. Bacchiani, D. Molinari, and M. Patander, "Microscopic traffic simulation by cooperative multi-agent deep reinforcement learning," in *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, ser. AAMAS '19. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2019, p. 1547–1555.

[11] C. Wu, A. Kreidieh, E. Vinitsky, and A. M. Bayen, "Emergent behaviors in mixed-autonomy traffic," ser. Proceedings of Machine Learning Research, S. Levine, V. Vanhoucke, and K. Goldberg, Eds., vol. 78. PMLR, 13–15 Nov 2017, pp. 398–407.

[12] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 961–971.

[13] A. Vemula, K. Muelling, and J. Oh, "Modeling cooperative navigation in dense human crowds," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 1685–1692.

[14] W. Ding, J. Chen, and S. Shen, "Predicting vehicle behaviors over an extended horizon using behavior interaction network," in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 8634–8640.

[15] B. Araki, K. Vodrahalli, T. Leech, C.-i. Vasile, M. Donahue, and D. Rus, "Deep Bayesian Nonparametric Learning of Rules and Plans from Demonstrations with a Learned Automaton Prior – Supplement," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020.

[16] S. Albrecht, C. Brewitt, J. Wilhelm, B. Gyevnar, F. Eiras, M. Dobre, and S. Ramamoorthy, "Interpretable goal-based prediction and planning for autonomous driving," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, United States, Feb. 2021.

[17] J. Hardy and M. Campbell, "Contingency planning over probabilistic obstacle predictions for autonomous road vehicles," *IEEE Transactions on Robotics*, vol. 29, no. 4, pp. 913–929, 2013.

[18] A. Kesting, M. Treiber, and D. Helbing, "General Lane-Changing Model MOBIL for Car-Following Models," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1999, no. 1, pp. 86–94, jan 2007.

[19] ——, "Enhanced intelligent driver model to access the impact of driving strategies on traffic capacity," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1928, pp. 4585–4605, 2010.

[20] A. Richards and J. How, "Decentralized model predictive control of cooperating uavs," in *2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601)*, vol. 4, 2004, pp. 4286–4291 Vol.4.

[21] X. Wang, V. Yadav, and S. N. Balakrishnan, "Cooperative uav formation flying with obstacle/collision avoidance," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 4, pp. 672–679, 2007.

[22] L. Dai, Q. Cao, Y. Xia, and Y. Gao, "Distributed mpc for formation of multi-agent systems with collision avoidance and obstacle avoidance," *Journal of the Franklin Institute*, vol. 354, no. 4, pp. 2068 – 2085, 2017.

[23] F. Xie and R. Fierro, "On motion coordination of multiple vehicles with nonholonomic constraints," in *2007 American Control Conference*, 2007, pp. 1888–1893.

[24] X. Qian, J. Gregoire, A. de La Fortelle, and F. Moutarde, "Decentralized model predictive control for smooth coordination of automated vehicles at intersection," in *2015 European Control Conference (ECC)*, 2015, pp. 3452–3458.

[25] Y. J. Zhang, A. A. Malikopoulos, and C. G. Cassandras, "Optimal control and coordination of connected and automated vehicles at urban traffic intersections," in *2016 American Control Conference (ACC)*, 2016, pp. 6227–6232.

[26] J. Rios-Torres and A. A. Malikopoulos, "Automated and cooperative vehicle merging at highway on-ramps," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 780–789, 2017.

[27] Y. Zhang, A. A. Malikopoulos, and C. G. Cassandras, "Decentralized optimal control for connected automated vehicles at intersections including left and right turns," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 4428–4433.

[28] K. Kim and P. R. Kumar, "An mpc-based approach to provable system-wide safety and liveness of autonomous ground traffic," *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3341–3356, 2014.

[29] G. Williams, B. Goldfain, P. Drews, J. M. Rehg, and E. A. Theodorou, "Best response model predictive control for agile interactions between autonomous ground vehicles," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 2403–2410.

[30] 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*, vol. 42, no. 7, pp. 1405–1426, 2018.

[31] A. Pierson, W. Schwarting, S. Karaman, and D. Rus, "Navigating congested environments with risk level sets," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 5712–5719.

[32] A. Garapin, L. Muller, and B. Rahali, "Does trust mean giving and not risking? experimental evidence from the trust game," *Revue d'économie politique*, vol. 125, no. 5, pp. 701–716, 2015.

[33] C. G. McClintock and S. T. Allison, "Social value orientation and helping behavior1," *Journal of Applied Social Psychology*, vol. 19, no. 4, pp. 353–362, 1989.

[34] C. E. Rusbult and P. A. Van Lange, "Interdependence, Interaction, and Relationships," *Annual Review of Psychology*, vol. 54, pp. 351–375, 2003.

[35] L. Wang, L. Sun, M. Tomizuka, and W. Zhan, "Socially-Compatible Behavior Design of Autonomous Vehicles with Verification on Real Human Data," *IEEE Robotics and Automation Letters*, vol. 3766, no. Feb, pp. 1–1, 2021.

[36] M. Moussaïd and M. Trauernicht, "Patterns of cooperation during collective emergencies in the help-or-escape social dilemma," *Scientific Reports*, vol. 6, no. June, pp. 1–9, 2016.

[37] J. Kong, M. Pfeiffer, G. Schildbach, and F. Borrelli, "Kinematic and dynamic vehicle models for autonomous driving control design," in *2015 IEEE Intelligent Vehicles Symposium (IV)*, 2015, pp. 1094–1099.

[38] A. Halder, "On the parameterized computation of minimum volume outer ellipsoid of minkowski sum of ellipsoids," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 4040–4045.

[39] 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*, vol. 11, no. 1, pp. 1–36, 2019.