Generalized Dynamic Cognitive Hierarchy Models for Strategic Driving Behavior

Atrisha Sarkar [1], **Kate Larson** [1], **Krzysztof Czarnecki** [2]

[1]David R Cheriton School of Computer Science [2]Department of Electrical and Computer Engineering
University of Waterloo, Ontario, Canada
atrisha.sarkar, kate.larson, k2czarne@uwaterloo.ca

Abstract

While there has been an increasing focus on the use of game theoretic models for autonomous driving, empirical evidence shows that there are still open questions around dealing with the challenges of common knowledge assumptions as well as modeling bounded rationality. To address some of these practical challenges, we develop a framework of generalized dynamic cognitive hierarchy for both modelling naturalistic human driving behavior as well as behavior planning for autonomous vehicles (AV). This framework is built upon a rich model of level-0 behavior through the use of automata strategies, an interpretable notion of bounded rationality through safety and maneuver satisficing, and a robust response for planning. Based on evaluation on two large naturalistic datasets as well as simulation of critical traffic scenarios, we show that i) automata strategies are well suited for level-0 behavior in a dynamic level-k framework, and ii) the proposed robust response to a heterogeneous population of strategic and non-strategic reasoners can be an effective approach for game theoretic planning in AV.

Introduction

One of many challenges of autonomous vehicles (AV) in an urban setting is the safe handling of other human road users who show complex and varied driving behaviors. As AVs integrate into human traffic, there has been a move from a *predict-and-plan* approach of behavior planning to a more strategic approach where the problem of behavior planning of an AV is set up as a non-zero sum game between road users and the AV; and such models have shown efficacy in simulation of different traffic scenarios (Fisac et al. 2019; Li et al. 2019; Sadigh et al. 2016; Tian et al. 2018). However, when game theoretic models are evaluated on naturalistic human driving data, studies have shown that human driving behavior is diverse, and therefore developing a unifying framework that can both model the diversity of human driving behavior as well as plan a response from the perspective of an AV is challenging (Sarkar and Czarnecki 2021; Sun et al. 2020).

The first challenge is dealing with *common knowledge* (Geanakoplos 1992) assumptions. Whether in the case of

Nash equilibiria based models (rational agents assuming everyone else is a rational agent) (Geiger and Straehle 2021; Schwarting et al. 2019), in Stackelberg equilibrium based models (common understanding of the leader-follower relationship) (Fisac et al. 2019), or level-k model (the level of reasoning of AVs and humans), there has to be a consensus between an AV planner and other road users on the type of reasoning everyone is engaging in. How to reconcile this assumption with the evidence that human drivers engage in different types of reasoning processes is a key question.

Whereas the challenge of common knowledge deals with questions around how agents model other agents, the second challenge is dealing with bounded rational agents, i.e. modelling sub-optimal behavior in their own response. With the general understanding that human drivers are bounded rational agents, the Quantal level-k model (QLk) has been widely proposed as a model of strategic planning for AV (Li et al. 2018, 2019; Tian et al. 2018). One way the QLk model deals with bounded rationality is through suboptimal reasoning, i.e., by mapping each agent into a hierarchy (k) of bounded cognitive reasoning. In this model, the choice of level-0 model becomes critical, since the behavior of every agent in the hierarchy depends on the assumption about the behavior of level-0 agents. The main models proposed for level-0 behavior include simple obstacle avoidance (Tian et al. 2021), maxmax, and maxmin models (Sarkar and Czarnecki 2021; Wright and Leyton-Brown 2020). Although such elementary models may be acceptable in other domains of application, it is not clear why human drivers, especially in a dynamic game setting, would cognitively bound themselves to such elementary models. Another way QLk model deals with bounded rationality is through suboptimal response, i.e., by the use of a precision parameter, where agents instead of selecting the utility maximizing response, make cost proportional errors modelled by the precision parameter (Wright and Leyton-Brown 2010). While the precision parameter provides a convenient way to fit the model to empirical data, it also runs the risk of being a catch-all for different etiologies of bounded rationality, including, people being indifferent to choices, uncertainty around multiple objectives, as well as the model not being a true reflection of the decision process. This impairs the explainability of the model since it is hard to encode and communicate such disparate origins of bounded rationality
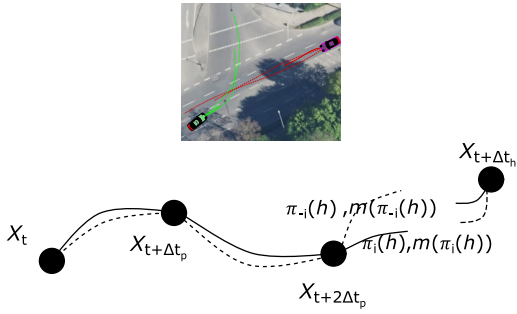
Figure 1: Schematic representation of the dynamic game. Each node is embedded in a spatio-temporal lattice and nodes are connected with a cubic spline trajectory.

through a single parameter.

The primary contribution of our work is a framework that addresses the aforementioned challenges by unifying modeling of heterogeneous human driving behavior with strategic planning for AV. In this framework, behavior models are mapped into three layers of increasing capacity to reason about other agents' behavior – *non-strategic*, *strategic*, and *robust*. Within each layer, the possibility of different types of behavior models lends support for a population of heterogeneous behavior, with a robust layer on top addressing the problem of behavior planning with a relaxed common knowledge assumptions. Standard level-k type and equilibrium models are nested within this framework, and in the context of those models, secondary contributions of our work are a) the use of automata strategies as a model of level-0 behavior in dynamic games, resulting in behavior that is rich enough to capture naturalistic human driving (dLk($\mathcal{A}$) model), and b) an interpretable support for bounded rationality based on different modalities of satisficing — *safety* and *maneuver*. Finally, the efficacy of the approach is demonstrated with evaluation on two large naturalistic driving datasets as well as simulation of critical traffic scenarios.

Game Tree, Utilities, and Agent Types

The dynamic game is constructed as a sequence of simultaneous move games played starting at time $t = 0$ at a period of Δt_p secs. over a horizon of Δt_h secs. Each vehicle $i \in \{1, 2, .., N\}$'s state at time t is a vector $X_{i,t} = [x, y, v_x, v_y, \dot{v}_x, \dot{v}_y, \theta]$ representing positional co-ordinates (x, y) on R^2, lateral and longitudinal velocity (v_x, v_y) in the body frame, acceleration $(\dot{v}_x, \dot{v}_y)$, and yaw (θ). The nodes of the game tree $X_t = X_{i,t}^N$ are the joint system states embedded in a spatio-temporal lattice (Ziegler and Stiller 2009), and the actions are cubic spline trajectories (Al Jawahiri 2018; Kelly and Nagy 2003) generated based on kinematic limits of vehicles with respect to bounds on lateral and longitudinal velocity, acceleration, and jerk (Bae, Moon, and Seo 2019). A history h_t of the game consists of sequence of nodes $X_0..X_t$ traversed by both agents along the game tree until time t. We also use a hierarchical approach in the trajectory generation process (Fisac et al. 2019; Michon 1985; Sarkar and Czarnecki 2021), where at each node, the trajectories are generated with respect to high-level maneu-

vers, namely, *wait* and *proceed* maneuvers. For wait maneuver trajectories, a moving vehicle decelerates (or remain stopped if it is already stopped), and for proceed maneuvers, a vehicle maintains its moving velocity or accelerates to a range of target speeds. Strategies are presented in the behavior strategy form, where $\pi_i(h_t) \in T_i(X_t)$ is a pure strategy response (a trajectory) of an agent i that maps a history h_t to a trajectory in the set $T_i(X_t)$, which is the set of valid trajectories that can be generated at the node X_t corresponding to both maneuvers. The associated maneuver for a trajectory is represented as $m(\pi_i(h)) \in \{wait, proceed\}$. Depending on the context where the response depends on only the current node instead of the entire history, we use the notation $\pi_i(X_t)$; we also drop the time subscript t on the history when a formulation holds true for all t. The overall strategy of the dynamic game is the cross product of behavior strategies along all possible histories of the game $\sigma : \pi_1(h) \times \pi_2(h) \times ..\pi_N(h); \forall h$. We use the standard game-theoretic notation of i and $-i$ to refer to an agent and other agents respectively in a game.

The utilities in the game are formulated as multi-objective utilities consisting of two components — **safety** $u_{s,i}(\pi_i(h), \pi_{-i}(h)) \in [-1, 1]$ (modelled as a sigmoidal function that maps the minimum distance gap between trajectories to a utility interval [-1,1]) and **progress** $u_{p,i}(\pi_i(h), \pi_{-i}(h)) \in [0, 1]; \forall i, -i$ (a function that maps the trajectory length in meters to a utility interval [0,1]). In general, these two are the main utilities (often referred to as inhibitory and excitatory utilities, respectively) upon which different driving styles are built (Sagberg et al. 2015). Agent types $\gamma_i \in \Gamma$ are numeric in the range [-1,1] representing each agent's *safety aspiration level* — a level of safety that an agent is comfortable operating. This construction is motivated by traffic behavior models such as Risk Monitoring Model (Vaa 2011), Task Difficulty Homeostasis theory (Fuller 2008), where based on traffic situations, drivers continually compare each possible action to their own risk tolerance threshold and make decisions accordingly. To avoid dealing with complexities that arise out of agent types being continuous, for the scope of this paper, we discretize the types by increments of 0.5 for our experiments. Based on their type (i.e. their safety aspiration level) how each agent selects specific actions at each node depends on the particular behavior model, and is elaborated in the next section.

Unless mentioned otherwise, the utilities (both safety and progress) at a node with associated history h_t are calculated as discounted sum of utilities over the horizon of the game conditioned on the strategy σ, type γ_i and discount factor δ as follows

$$\sum_{k=1}^{\lfloor \Delta t_h - t \backslash \Delta t_p \rfloor} \delta^k u_i(\pi_i(h_{t+k-1}), \pi_{-i}(h_{t+k-1}); \sigma, \gamma_i) + \mathcal{N}u_{i,C}$$

where $u_{i,C}$ are the continuation utilities beyond the horizon of the game and are estimated based on agents continuing on with the same chosen trajectory as undertaken in the last decision node of the game tree for another Δt_h seconds, and $\mathcal{N}$ is a normalization constant to keep the sum of utilities in the same range as the step utilites.

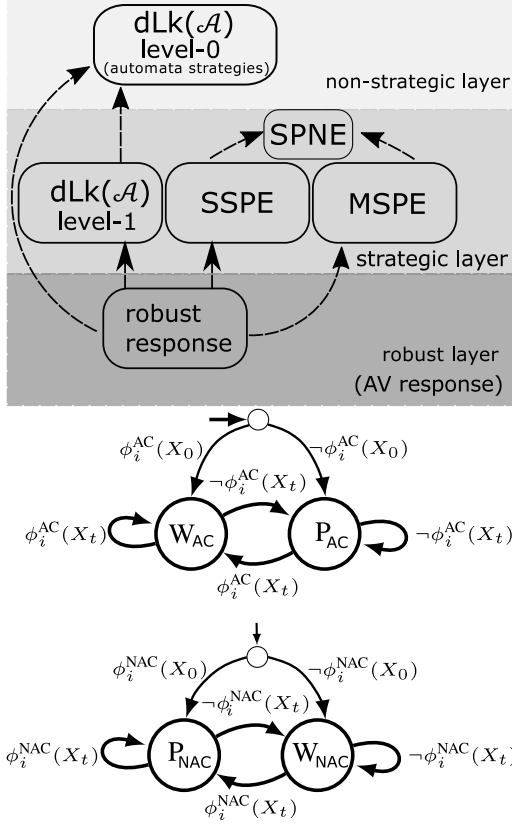# Generalized Dynamic Cognitive Hierarchy Model

Figure 2: (a) Organization of models in the generalized dynamic cognitive hierarchy framework. Dashed arrows indicate agents' belief about the population. (b) Automata $\mathcal{A}^{\text{AC}}$ (*accomodating*) and $\mathcal{A}^{\text{NAC}}$ (*non-accomodating*)

In order to support heterogeneous behavior models, the generalized dynamic cognitive hierarchy model consists of three layers of increasing sophistication of strategic reasoning, each of which can hold multiple behavior models. The three layers include: a) *non-strategic*, where agents do not reason about the other agents' strategies, b) *strategic*, where agents can reason about the strategies of other agents, and c) *robust*, where agents not only reason about the strategies of other agents but also the behavior model (Wright and Leyton-Brown 2020) that other agents may be following (Fig. 2a). All of these layers operate in a setting of a dynamic game and we present the models and the solution concepts used in each layer in order.

Non-strategic Layer

Similar to level-0 models in the standard level-k reasoning (Camerer, Ho, and Chong 2004), non-strategic agents form the base of the cognitive hierarchy in our model. However, in our case, we extend the behavior for the dynamic game. The main challenge of constructing the right level-0 model for a dynamic game is that it has to adhere to the formal con-

straints of non-strategic behavior, i.e. not reason over other agents' utilities (Wright and Leyton-Brown 2020), while at the same time cannot be too elementary for the purpose of modeling human driving, which implies allowing for agents to change their behavior over time in the game based on the game state.

We propose that automata strategies, which were introduced as a way to address bounded rational behavior that results from agents having limited memory capacity to reason over the entire strategy space in a dynamic game (Marks 1992; Rubinstein 1986), address the above problem by striking a balance between adequate sophistication and non-strategic behavior. To this end, we extend the standard level-k model for a dynamic game setting with level-0 behavior mediated by automata strategies (referred to as dLk($\mathcal{A}$) henceforth in the paper). In this framework, a level-0 agent has two reactive modes of operation, each reflecting a specific driving style modeled as automata strategies; *accomodating* ($\mathcal{A}^{\text{AC}}$) and *non-accomodating* ($\mathcal{A}^{\text{NAC}}$) (Fig. 2b). An agent i playing $\mathcal{A}^{\text{AC}}$ as type $\gamma_i^{\text{AC}} \in \Gamma$ in state W_{AC} randomize uniformly between trajectories belonging to *wait* maneuver that have step safety utility at least γ_i^{AC}. If no such *wait* trajectories are available to them, they move to state P_{AC} and randomizes uniformly between trajectories belonging to *proceed* maneuver. $\mathcal{A}^{\text{NAC}}$ is similar, but the states are reversed thereby resulting in a predisposition that prefers the proceed state. The switching between the states of the automata is mediated by the preference conditions ϕ_i^{AC} and ϕ_i^{NAC} shown in the equations below.

$$\phi_i^{\text{AC}} := \max_{\text{T}_i(X_t)|_{\text{W}}} u_{s,i}^{\Delta t_p}(\pi(X_t)) \leqslant \gamma_i^{\text{AC}} \qquad (1)$$

$$\phi_i^{\text{NAC}} := \max_{\text{T}_i(X_t)|_{\text{P}}} u_{s,i}^{\Delta t_p}(\pi(X_t)) > \gamma_i^{\text{NAC}} \qquad (2)$$

ϕ_i^{AC} is true when there is at least one wait trajectory (and conversely, at least one proceed trajectory in the case of ϕ_i^{NAC}) available whose step safety utility ($u_{s,i}^{\Delta t_p}$) at the current node X_t is above γ_i^{AC} (and γ_i^{NAC} for ϕ_i^{NAC}). Recall that agent types are in the range [-1,1] and are reflective of safety aspiration.

Along with switching between states, agents are also free to switch between the two automata; however, we leave open the question of modeling the underlying process and the causal factors. We envision them to be non-deterministic and a function of the agent's affective states, such as impatience, attitude, etc., which are an indispensable component of modeling driving behavior (Sagberg et al. 2015). For example, one can imagine a left turning vehicle approaching the intersection starting with an accommodating strategy, but along the game play, changes its strategy to non-accommodating on account of impatience or other endogenous factors. Although leaving open the choice of mode switching leads to the level-0 model in this paradigm to be partly descriptive rather than predictive, as we will see in the next section, such a choice does not compromise the ability of higher level agents' to form consistent beliefs about the level-0 agent and respond accordingly to their strategies. This richer model of level-0 behavior not only imparts more realism in the context of human driving, but also allows for level-0 agent to adapt their behavior based on the game situation in a dy-

namic game.

Strategic Layer

The difference between strategic and non-strategic models is that the latter adhere to the two properties of *other responsiveness*, i.e. reasoning over the utilities of the other agents, and *dominance responsiveness*, i.e. optimizing over their own utilities (Wright and Leyton-Brown 2020). Agents in the strategic layer in our framework adhere to the above two properties and we include three models of behavior, namely, level-k ($k \geqslant 1$) behavior based on the dLk($\mathcal{A}$) model and two types of bounded rational equilibrium behavior, i.e. SSPE (Safety satisfied perfect equilibria) and MSPE (Maneuver satisfied perfect equilibria). We note that the choice of models in the strategic layer is not exhaustive, however, we select popular game-theoretic models (level-k and equilibrium based) that have been used in the context of autonomous driving and address some of the gaps within the use of those models as secondary contributions.

For the strategic models to reason over the utilities of other agents and respond accordingly, the multi-objective utilities presented in section titled Game Tree, Utilities, and Agent Types, need to be aggregated into a scalar utility. To that end, we use a lexicographic thresholding approach to aggregate the two multi-objective utilities (Li and Czarnecki 2019), and furthermore, we connect the lexicographic threshold to an agent's type γ_i. Specifically, the combined utility $u_i(\pi_i(h), \pi_{-i}(h))$ is equal to $u_{s,i}(\pi_i(h), \pi_{-i}(h))$, i.e., the safety utility when $u_{s,i}(\pi_i(h), \pi_{-i}(h)) \leqslant \gamma_i$, and otherwise $u_i(\pi_i(h), \pi_{-i}(h)) = u_{p,i}(\pi_i(h), \pi_{-i}(h))$, i.e., the progress utility. Based on the aggregation method, an agent would consider the progress utility of an action only when it's safety utility is greater than γ_i. This way of aggregation provides a natural connection between an agent's type and their risk tolerance. In the following sections, when we use the agent types in different behavior models, we also index the agent types with the specific models for clarity.

dLk($\mathcal{A}$) Model

A level-1 agent [1] believes that the population consists solely of level-0 agents, and generates a best response to those (level-0) strategies. In a dynamic setting, however, a level-1 agent has to update it's belief about level-0 agent based on observation of the game play. This means that in order to best respond to level-0 strategy, a level-1 agent should form a *consistent* belief based on the observed history h_t of the game play about the type $(\gamma_{-i}^{AC}, \gamma_{-i}^{NAC})$ a level-0 agent plays in each automaton.

Definition 1. *A belief $\mathcal{B}_{l1} = \{\hat{\Upsilon}_{l0}^{AC}, \hat{\Upsilon}_{l0}^{NAC}\}$ of a level-1 agent at history h_t is consistent iff $\forall \gamma_{-i}^{AC} \in \hat{\Upsilon}_{l0}^{AC}$, $\pi_{-i}(X_0)...\pi_{-i}(X_t) \in \mathcal{T}(\mathcal{A}^{AC}; \gamma_{-i}^{AC}, X_0..X_t)$ and $\forall \gamma_{-i}^{NAC} \in \hat{\Upsilon}_{l0}^{NAC}$, $\pi_{-i}(X_0)...\pi_{-i}(X_t) \in \mathcal{T}(\mathcal{A}^{NAC}; \gamma_{-i}^{NAC}, X_0..X_t)$ where $\hat{\Upsilon}_{l0}^{AC} \subseteq \Gamma$, $\hat{\Upsilon}_{l0}^{NAC} \subseteq \Gamma$, and (dropping the superscripts) $\mathcal{T}(\mathcal{A}; \gamma, X_0..X_t)$ is the trace of the automaton, defined as*

---

[1] We focus on $k = 1$ behavior in this section as well as later in the experiments, but the best response behavior can be extended to $k > 1$ similar to a standard level-k model.

the set of all valid sequence of actions generated by an level-0 agent playing an automaton $\mathcal{A}$ as type γ and encountering the nodes $X_0..X_t$.

Before the level-1 agent has observed any action by the level-0 agent, their estimates for $\hat{\Upsilon}_{l0}^{AC}$ and $\hat{\Upsilon}_{l0}^{NAC}$ is the default set of types, i.e. $\hat{\Upsilon}_{l0}^{AC} = \hat{\Upsilon}_{l0}^{NAC} = \Gamma$ with range 2 (recall $\Gamma = [-1, 1]$). However, over time with more observations of level-0 actions, level-1 agent forms a tighter estimate, i.e. $|\max(\hat{\Upsilon}_{l0}^{AC}) - \min(\hat{\Upsilon}_{l0}^{AC})| \leqslant 2$, of level-0's type. The following theorem formulates the set of consistent beliefs $\mathcal{B}_{l1}$ based on observed level-0 actions with history h more formally.

Theorem 1. *For any consistent belief γ_{-i}^{AC} and γ_{-i}^{NAC}, where $\gamma_{-i}^{AC} \in \hat{\Upsilon}_{l0}^{AC}$ and $\gamma_{-i}^{NAC} \in \hat{\Upsilon}_{l0}^{AC}$ the following inequalities hold true for any history h of the game,*

$$\gamma_{-i}^{AC} \quad \begin{aligned} &< \min_{h[P]} \max_{T_{-i}(X)|_W} u_{s,l0}^{\Delta_{t_p}}(\pi(X)) \\ &\geqslant \max_{h[W]} \max_{T_{-i}(X)|_W} u_{s,l0}^{\Delta_{t_p}}(\pi(X)) \end{aligned}$$

$$\gamma_{-i}^{NAC} \quad \begin{aligned} &\geqslant \max_{h[W]} \max_{T_{-i}(X)|_P} u_{s,l0}^{\Delta_{t_p}}(\pi(X)) \\ &< \min_{h[P]} \max_{T_{-i}(X)|_P} u_{s,l0}^{\Delta_{t_p}}(\pi(X)) \end{aligned}$$

where $h[P]$ and $h[W]$ are set of game nodes where level-0 agent chose a proceed and wait maneuver respectively, and $T(X)|_P, T(X)|_W$ are the available trajectories at node X belonging to the two maneuvers, proceed and wait, respectively.

Proof. There are two parts to the equation, one for γ_{-i}^{AC} and another γ_{-i}^{NAC}. We prove the bounds of γ_{-i}^{AC} corresponding to automata $\mathcal{A}^{AC}$, and the proof for γ_{-i}^{NAC} corresponding to automata $\mathcal{A}^{NAC}$ follows in an identical manner.

By construction of automata $\mathcal{A}^{AC}$, proceed trajectories are only generated in state P_{AC}, which follows one of the three transitions $\{inp.|P_{AC}|W_{AC}\} \xrightarrow{\neg \phi_i^{AC}(X_t)} P_{AC}$. Therefore, $\forall X_t \in h[P]$, $\phi_i^{AC} := \bot$ for the transition to happen. Based on eqn. 1, this means that $\forall X_t \in h[P]$,

$$\gamma_i^{AC} < \max_{T_{-i}(X_t)|_W} u_{s,i}^{\Delta_{t_p}}(\pi(X_t)) \qquad (A.1)$$

Since γ_i^{AC} stays constant throughout the play of the game, for γ_i^{AC} to be consistent for the set of all nodes $X_t \in h[P]$, the lower bound (i.e., at least as high to be true for all nodes $X_t \in h[P]$) of γ_i^{AC} based on eqn A.1 is

$$\gamma_i^{AC} < \min_{h[P]} \max_{T_{-i}(X)|_W} u_{s,l0}(\pi(X)) \qquad (A.2)$$

Similarly, by construction of automata, wait trajectories are only generated in state W_{AC}, which follows one of the three transitions $\{inp.|P_{AC}|W_{AC}\} \xrightarrow{\phi_i^{AC}(X_t)} W_{AC}$. Therefore, $\forall X_t \in h[W]$, $\phi_i^{AC} := \top$ for the transition to happen. Based on eqn. 1, this means that $\forall X_t \in h[W]$,

$$\gamma_i^{AC} \geqslant \max_{T_{-i}(X_t)|_P} u_{s,i}^{\Delta_{t_p}}(\pi(X_t)) \qquad (A.3)$$

and the upper bound (i.e., at least as low to be true for all nodes $X_t \in h[\mathrm{W}]$) of γ_i^{AC} based on eqn A.3 $\forall X_t \in h[\mathrm{W}]$

$$\gamma_i^{\mathrm{AC}} \geqslant \max_{h[\mathrm{W}]} \max_{\mathrm{T}(X_t)|_W} u_{s,l0}(\pi(X)) \qquad (\text{A.4})$$

Since $h[\mathrm{P}] \bigcap h[\mathrm{W}] = \varnothing$ and $h = h[\mathrm{P}] \bigcup h[\mathrm{W}]$, equations A.2 and A.4 in conjunction proves the case for γ_i^{AC} bounds.

The proof for γ_i^{NAC} follows in the identical manner as γ_i^{AC}, but with the condition reversed based on the P and W states. $\square$

The above theorem formalizes the idea that looking at maneuver choices made by a level-0 agent at each node in the history, as well as the range of step safety utility at that node for both maneuvers (recall that preference conditions of the automata are based on step safety utility), a level-1 agent can calculate ranges for $\gamma_{-i}^{\mathrm{AC/NAC}}$ from Eqns 1 and 2, for which the observed actions were consistent with each automata. With respect to the set of consistent belief $\mathcal{B}_{l1}$ about level-0 agent's strategy, level-1 agent now needs to generate a best response that is consistent with $\mathcal{B}_{l1}$. Dropping the AC/NAC superscripts Let $\pi(X_t; \mathcal{A}, \hat{\Upsilon}_{l0}) = \{\pi(X_t; \mathcal{A}, \gamma_{-i}); \forall \gamma_{-i} \in \hat{\Upsilon}_{l0}\}$ be the union of all actions when the automata $\mathcal{A}$ is played by the types in $\hat{\Upsilon}_{l0}$, then $\Pi_{l0}(X_t, \mathcal{B}_{l1}) = \pi(X_t; \mathcal{A}^{\mathrm{AC}}, \hat{\Upsilon}_{l0}^{\mathrm{AC}}) \bigcup \pi(X_t; \mathcal{A}^{\mathrm{NAC}}, \hat{\Upsilon}_{l0}^{\mathrm{NAC}})$ is the set of all actions that level-0 agent can play based on level-1's consistent belief $\mathcal{B}_{l1}$. The response to those actions by level-1 agent (indexed as i) is as follows.

$$\pi_i(h_t; \mathcal{B}_{l1}) = \underset{\substack{\mathrm{T}_i(X_t) \\ \Pi_{l0}(X_t, \mathcal{B}_{l1})}}{\arg\max} \ u_i(\pi_i(h), \pi_{-i}(X_t)|\gamma_i^{l1}) \qquad (3)$$

where $\gamma_i^{l1} \in \Gamma$ is the type of the level-1 agent, $\pi_i(h) \in \mathrm{T}_i(X_t)$ is a valid trajectory that can be generated at the node with history h, and $\pi_{-i}(X_t) \in \Pi_{l0}(X_t, \mathcal{B}_{l1})$ is the other agent's trajectory. Note that the strategy of the level-1 agent, unlike level-0 agent depends on the history h_t instead of just the state of the node X_t; since the history influences the belief $\mathcal{B}_{l1}$ which in turn influences the response.

The above model of best response can be extended similarly for $k > 1$ just like in standard level-k models. However, we restrict our focus for k up to 1 mainly due to following reasons: i) aspects of higher level thinking are already incorporated in the equilibrium based models; ii) in other domains, we often see diminishing returns with higher k values with respect to empirical data (Wright and Leyton-Brown 2010).

Equilibrium Models

Along with the level-k (k$\geqslant$ 1) behavior, another notion of strategic behavior that have been proposed as a model of behavior planning in AV (Pruekprasert et al. 2019; Schwarting et al. 2019) are based on an equilibrium. However, when it comes to modelling human driving behavior, an equilibrium model needs to accommodate bounded rational agents in a principled manner, and ideally should provide a reasonable explanation of the origin of the bounded rationality. Based on the idea that drivers are indifferent as long as an action achieves their own subjective safety aspiration threshold (Lewis-Evans 2012), we use the idea of *satisficing* as the

main framework for bounded rationality in our equilibrium models (Stirling 2003). Specifically, we develop two notions of satisficing; one based on safety satisficing (SSPE), where agents choose actions close to the Nash equilibria as long as the actions are above their own safety aspiration threshold, and another based on maneuver satisficing (MSPE), where agents chose actions close to the Nash equilibria as long as the actions are of the same high-level maneuver as the optimal action.

Safety-satisfied Perfect Equilibrium (SSPE). The main idea behind satisficing is that a bounded rational agent, instead of always selecting the best response action, selects a response that is *good enough*, where *good enough* is defined as an aspiration level where the agent is indifferent between the optimal response and the response in question. In the case of Safety-satisfied perfect equilibrium (SSPE), we define a response good enough for agent i if the response is above their own safety aspiration threshold as determined by their type. The goal of this model is to let agents chose their action based on the subgame perfect Nash equilibrium of the game, however, also allow for selection of actions close to the equilibrium that are safe enough based on their own safety aspiration level. Therefore, in response to other agents' equilibrium action, any action that is safer than either their own equilibrium action or their own safety aspiration level γ_i is part of the solution. A more formal definition is as follows

Definition 2. *A strategy* $\sigma; (\gamma_i^{SS}, \gamma_{-i}^{SS}) : \prod_{\forall i \in N} \pi_i(h)$, *is in safety satisfied perfect equilibria for a combination of agent types* $(\gamma_i^{SS}, \gamma_{-i}^{SS}) \in \Gamma^N$ *if for every history h of the game and* $\forall i \in N$

$$u_{s,i}(\pi_i(h), \pi_{-i}^*(h)) \geqslant \min\{u_{s,i}^*(h), \gamma_i^{SS}\}$$

where $\sigma^*; (\gamma_i^{SS}, \gamma_{-i}^{SS}) : \prod_{\forall i \in N} \pi_i^*(h)$, *is a subgame perfect Nash equilibrium in pure strategies of the game for agents with type* $(\gamma_i^{SS}, \gamma_{-i}^{SS})$ *and* $u_{s,i}^*(h) = u_{s,i}(\pi_i^*(h), \pi_{-i}^*(h))$. [2]

Based on the above definition, if the safety utility of the best response of agent i to agent $-i$'s subgame perfect Nash equilibrium (SPNE) strategies at history h is less than agent i's own safety threshold as expressed by their type γ_i^{SS}, then the SSPE response is any trajectory that matches the safety utility of the SPNE response. However, if the SPNE response is higher than their safety threshold, then any suboptimal response that has safety utility higher than γ_i^{SS} is a *satisfied* response, and thus in SSPE.

Maneuver-satisfied Perfect Equilibrium (MSPE). This model of satisficing is based on the idea that agents may be indifferent between actions that belong to the maneuver corresponding to the equilibrium trajectory. For example, at any node, if the equilibrium trajectory belongs to *wait* maneuver, then all trajectories belonging to *wait* maneuver will be in MSPE. However, there are additional constraints that need

---

[2] We calculate the SPNE using backward induction with the combined utilities (lexicographic thresholding) in a complete information setting where agent types are known to each other.

to be imposed on this manner of action selection. In order to avoid selection of trajectories that belong to equilibrium maneuver but have lower utility than a non-equilibrium maneuver, we add the constraint that the utility of a selected trajectory has to be strictly higher than that of any non-equilibrium maneuver's trajectory. A typical example for the need of this constraint include situations where the equilibrium trajectory corresponds to *wait* maneuver, but selecting a trajectory that is akin to a rolling stop, which although falls under wait maneuver, is worse than executing a proceed trajectory, i.e., a non-equilibrium maneuver. Therefore, such trajectories should be excluded from the MSPE solution set. A formal definition follows.

Definition 3. *A strategy* $\sigma; (\gamma_i^{MS}, \gamma_{-i}^{MS}) : \prod_{\forall i \in N} \pi_i(h)$*, is in maneuver satisfied perfect equilibria for a combination of agent types* $(\gamma_i^{MS}, \gamma_{-i}^{MS}) \in \Gamma^N$ *if for every history* h *of the game and* $\forall i \in N$*,* $m(\pi_i(h)) = m(\pi_i^*(h))$ *and*

$$u_{s,i}(\pi_i(h), \pi_{-i}^*(h)) > \max_{\pi_i'(h) \in \mathrm{T}_i(X) \setminus m^*} u_i(\pi_i'(h), \pi_{-i}^*(h))$$

where $\sigma^*; (\gamma_i^{MS}, \gamma_{-i}^{MS}) : \prod_{\forall i \in N} \pi_i^*(h)$*, is a subgame perfect equilibrium in pure strategies of the game with agent types* $(\gamma_i^{MS}, \gamma_{-i}^{MS})$*, and* $\mathrm{T}_i(X) \setminus m^* = \{\pi_i(h) : m(\pi_i(h)) \neq m(\pi_i^*(X))\}$ *or in other words, the set of available trajectories at node* X *that do not belong to the maneuver corresponding to the equilibrium trajectory* $m(\pi_i^*(h))$ *and* X *is the last node in the history* h*.*

Robust Layer

While the presence of multiple models in the strategic layers allow for a population of heterogeneous reasoners, an agent following one of those models still has specific assumptions about the reasoning process of other agents, e.g. level-1 agents believing that the population consists of level-0 agents and equilibrium responders believing that other agents adhere to a common knowledge of rationality. Based on the position that a planner for AV should not hold such strict assumptions, we develop the robust layer. What differentiates the robust layer from the strategic layer is that along with the two properties of other responsiveness and dominance responsiveness for the strategic layer, agents in the robust layer also adhere to the property of *model responsiveness*, i.e., the ability to reason over the behavior models of other agents. This gives them the ability to reason about (forming beliefs about and responding to) a population of different types of reasoners including strategic, non-strategic, as well as agents following different models within each layer. The overall behavior of a robust agent can be broken down into three sequential steps as follows.

i. Type expansion: Since the robust agent not only has to reason over the types of other agents, but also the possible behavior models, we augment the initial set of agent types Γ that were based on agents' safety aspiration with the corresponding agent models. Let $\Gamma^+ : \mathcal{M} \times \Gamma$ be the augmented type of an agent, where $\mathcal{M}$ is the set of models presented earlier, i.e., {accomodating, non-accomodating, level-1, SSPE,

MSPE} and Γ is the (non-augmented) agent type agent type $(\gamma_i^{AC}, \gamma_i^{NAC}, \gamma_i^{l1}, \gamma_i^{SS}, \gamma_i^{MS})$ within each model.

ii. Consistent beliefs: Similar to strategic agents, based on the observed history h of the game, a robust agent forms a belief $\beta_h \subseteq \Gamma^+$ such that the observed actions of the other agent in the history h is consistent with the augmented types (i.e., model as well as the agent type) in β_h. The process of checking whether a history is consistent with a combination of a model and agent type was already developed earlier for two non-strategic models (Def. 1). For level-1 models, the history is consistent if at each node in history, the response of the other agent adheres to equation 3; and for the equilibrium models, a history h is consistent if based on definitions 2 and 3 for SSPE and MSPE respectively, the actions are along the equilibrium paths of the game tree. Assuming that in driving situations agents behave truly according to their types, β_h is then constructed as an union of all the consistent beliefs for each model.

iii. Robust response: The idea of a robust response to heterogeneous models is along the lines of the robust game theory approach of (Aghassi and Bertsimas 2006). The belief set β_h represents the uncertainty over the possible models the other agents' may be using along with the corresponding agent types within those models. A robust response to that is the optimization over the worst possible outcome that can happen with respect to that set. Eqn. 4 formulates this response of a robust agent playing as agent i.

$$\pi_i(h; \beta_h) = \arg\max_{\mathrm{T}_i(X)} \inf_{\forall \beta \in \beta_h} \max_{\forall \pi_{-i}(h;\beta)} u_i(\pi(h), \pi_{-i}(h;\beta); \gamma_i^R) \tag{4}$$

where $\pi_{-i}(h; \beta)$ are the possible actions of the other agent based on the augmented type $\beta \in \beta_h$ and $\gamma_i^R \in \Gamma$ is the robust agent's own type. In this response, the minimization happens over the agent types (inner $\inf$ operator), rather than over all the actions $\pi_{-i}(h)$ as is common in a maxmin response. Since driving situations are not, in most cases, purely adversarial, this is a less conservative, yet robust, response compared to a maxmin response.
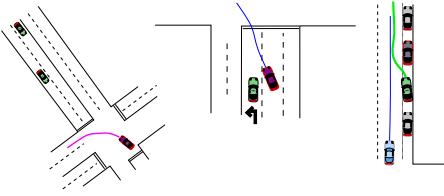
Experiments and Evaluation

In this section we present the evaluation of the models under two different experiment setups. First, we compare the models with respect to large naturalistic observational driving data using a) the Intersection dataset from the Waterloo multi-agent traffic dataset (WMA) recorded at a busy Canadian intersection (Sarkar and Czarnecki 2021) available at http://wiselab.uwaterloo.ca/waterloo-multi-agent-traffic-dataset/, and b) the inD dataset recorded at intersections in Germany (Bock et al. 2020) (Fig. 3a). From both datasets, which include around 10k vehicles in total, we extract the long duration unprotected left turn (LT) and right turn (RT) scenarios, and instantiate games between left (and right) turning vehicles and oncoming vehicles with $\Delta t_h = 6s$ and $\Delta t_p = 2s$, resulting in a total of 1678 games. The second part of the evaluation is based on simulation of three critical traffic scenarios derived from the NHTSA precrash database (Najm et al. 2007), where we instantiate agents with a range of safety aspirations as well as initial game states, and evaluate the outcome of the game based
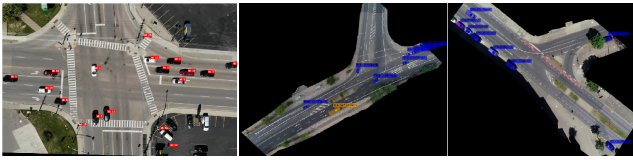
(a) Snapshot of naturalistic datasets (WMA and inD)

(b) Simulation of critical scenarios: intersection clearance, merge before intersection, parking pullout.

Figure 3: Evaluation setup based on naturalistic datasets and simulation scenarios.

| | WMA | | inD | |
	LT (1103)	RT (311)	LT (187)	RT (77)
maxmax	0.33438 (0.02)	0.43023 (0.27)	0.37975 (-0.2)	0.43506 (0.1)
AC	0.82053 (0.82)	0.90698 (0.84)	0.92089 (0.75)	0.81818 (0.79)
NAC	0.17947 (0.87)	0.09302 (0.82)	0.07911 (0.88)	0.18182 (0.85)
QLk(λ=1)	0.18262 (-0.07)	0.43265 (0.33)	0.37658 (0.37)	0.43506 (-0.1)
QLk(λ=0.5)	0.34131 (-0.03)	0.43023 (0.26)	0.37658 (0.37)	0.43506 (-0.1)
dLk($\mathcal{A}$)	0.5529 (0.19)	0.65449 (0.3)	0.51266 (0.51)	0.53247 (-0.4)
SSPE	0.69144 (-0.84)	0.90033 (-0.94)	0.6962 (-0.86)	0.53247 (-0.85)
MSPE	0.30479 (-0.1)	0.44518 (-0.13)	0.21519 (0.21)	0.27273 (-0.7)
Robust	0.56045 (0.20)	0.66944 (0.34)	0.51582 (0.51)	0.53247 (-0.4)

Table 1: Overall match rate of the models for each dataset and scenario. Mean agent type (γ) noted in parenthesis. LT: Left turn, RT: Right turn. Number of games noted in the header.

on each model. All the games in the experiments are 2 agent games with the exception of one of the simulation of critical scenario (intersection clearance), which is a 3 agent game. Code and supplementary videos are available at https://git.uwaterloo.ca/a9sarkar/repeated_driving_games.
Baselines. We select multiple baselines depending on whether a model is strategic or non-strategic. For non-strategic models, we compare the automata based model with a maxmax model, shown to be most promising from a set of alternate elementary models with respect to naturalistic data (Sarkar and Czarnecki 2021). For the strategic models (level-1 in dLk($\mathcal{A}$), SSPE, MSPE), we select a QLk model used in multiple works within the context of autonomous driving (Li et al. 2018, 2019; Tian et al. 2018, 2021). We use the same parameters used in (Tian et al. 2021) for the precision parameters in the QLk model.

Naturalistic data We evaluate the model performance on naturalistic driving data based on match rate, i.e. the number of games where the observed strategy of the human driver matched a model's solution divided by the total number of games in the dataset. More formally, let D be the set of games in the dataset, an indicator function $I : D \rightarrow \{0,1\}$ is 1 if in the game g, there exists a combination of agent types (γ_i, γ_{-i}), such that the observed strategy is in the set of strategies as predicted by the model, or 0 otherwise. The overall match rate of a model is given by $\sum_{\forall g \in D} I/|D|$. QLk (baseline) models being mixed strategy models, we count a match if the observed strategy is assigned a probability of $\geqslant 0.5$. Table 1 shows the match rate of each model for each dataset and scenario. It also shows in parenthesis the mean γ_i, i.e. the agent type value for each model when the strategy matched the observation. Since the agent types are a free parameter within each model, this gives the models flexibility to match the observed action to a driver's safety aspiration level. The overall numbers in the table show that there is variation both with respect to match rate as well as the matched safety aspiration level (agent type) for a given model. We observe that the match rate, especially for WMA

dataset, is better for right turning scenarios than for left turning ones. The models sometimes find harder to match a consistent safety aspiration level at left turns in WMA when vehicles exhibit impatient behavior by creeping forward. Notice that the difference is less stark in the inD dataset because in inD (Fig. 3a), the LT vehicle is still in the turn lane at point of initiation (i.e., before the stopline), and therefore has less incentive to take a risk and creep.

A major takeaway is that for non-strategic models, automata models show much higher match rate thereby reflecting high alignment with human driving behavior compared to the maxmax model. In fact, as we can see from the table that the entries for AC and NAC sum up to 1. Combination of AC and NAC although being non-strategic, can capture all observed driving decisions in the dataset, which indicate that automata models are very well suited for modelling level-0 behavior in a dynamic game setting for human driving. The performance of an accommodating model is not very surprising since for left and right turning scenarios, most drivers generally give way to oncoming vehicles. For the strategic models, dLK($\mathcal{A}$) and SSPE models show better performance than QLk and MSPE models. However, when we compare the mean agent types, we observe that when SSPE strategies matches the observation, it is based on agents with very low safety aspiration (reflected in low mean agent type values). If we assume that the population of drivers on average have moderate safety aspiration, say in the range [-0.5,0.5] (estimating the true distribution is out of the current scope), dLk($\mathcal{A}$) is a more reasonable model of strategic behavior compared to SSPE. We include the robust model comparison for the sake of completeness (and it shows performance comparable to dLk($\mathcal{A}$) model), but as mentioned earlier, ro-
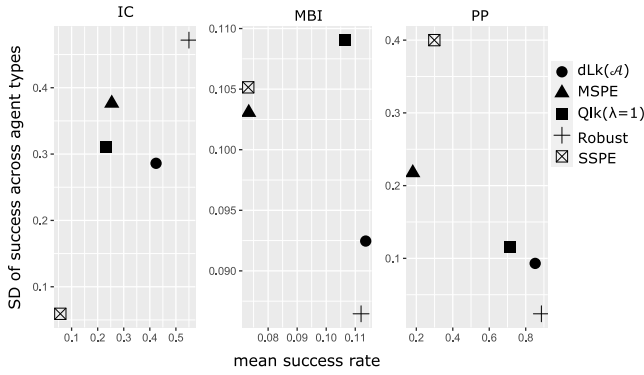
Figure 4: Mean and SD of success for each model in each scenario across all agent types.

bust model is a model of behavior planning for an AV, and therefore ideally needs to be evaluated on criteria beyond just comparison to naturalistic human driving, which we discuss in the next section.

Critical scenarios While evaluation based on a naturalistic driving datasets helps in the understanding of how well a model matches human driving behavior, in order to evaluate the suitability of a model for behavior planning of an AV, the models need to be evaluated on specific scenarios that encompass the operational design domain (ODD) of the AV (Ilievski 2020). Since the models developed in this paper are not specific to an ODD, we select three critical scenarios from ten most frequent crash scenarios in the NHTSA pre-crash database (Najm et al. 2007).

Intersection Clearance (IC): Left turn across path (LTAP) scenario where the traffic signal has just turned from green to yellow at the moment of the game initiation. There is a left turning vehicle is on the intersection and two oncoming vehicles from the opposite direction close to the intersection who may chose to speed and cross or wait for the next cycle. The expectation is that the left turning vehicle should be able to clear the intersection by the end of the game horizon without crashing into either oncoming vehicles, and no vehicles should be stuck in the middle of the intersection.

Merge Before Intersection (MBI): Merge scenario where a left-turning vehicle (designated as the merging vehicle) finds itself in the wrong lane just prior to entering the intersection, and wants to merge into the turn lane in front of another left-turning vehicle (designated as the on-lane vehicle). The expectation is that the on-lane vehicle should allow the other vehicle to merge.

Parking Pullout (PP): Merge scenario where a parked vehicle is pulling out of a parking spot and merges into traffic while there is a vehicle coming along the same direction from behind. The expectation is that the parked vehicle should wait for the coming vehicle to pass before merging into traffic.

For each scenario, we run simulations with a range of approach speeds as well as all combination of agent types from the set of agent types $\Gamma = \{-1, -0.5, 0, 0.5, 1\}$. One way to compare the models is to evaluate them based on the mean success rate across all initiating states and agent types. Fig. 4

shows the mean success rate (success defined as the desired outcome based on expectation defined in the description for each scenario) for all the strategic and robust models. We see that the mean success rate of the robust and dLk(A) model is higher compared to the equilibrium models or the QLk model. However, this is only part of the story. With varying initiation conditions, it may be harder or easier for a model to lead to a successful outcome. For example, in the parking pullout scenario a vehicle with low safety aspiration coming at a higher speed is almost likely to succeed in all models when facing a parked vehicle with high safety aspiration at zero speed. Therefore, to tease out the stability of models across different safety aspirations (i.e. agent type combinations), Fig. 4 also plots on y-axis, the standard deviation of the mean success rate across different agent types. Ideally, a model should have a high success rate with low SD across types indicating that with different combinations of agent type population (from extremely low safety aspiration to very high), the success rate stays stable. As we see from Fig. 4, the robust and dLk(A) models are broadly in the ideal lower right quadrant (high mean success rate low SD) for parking pullout and merge before intersection scenarios. For IC scenario, however, we observe that the success rate comes at a price of high SD (for all models) as indicated by the linearly increasing relation between mean success rate and its SD across agent types. This means that the success outcomes are skewed towards a specific combination of agent types; specifically, the case where the left turning vehicle has low safety aspiration. It is intuitive to imagine that in a situation like IC, agents with high safety aspiration would be stuck in the intersection instead of being able to navigate out of the intersection quickly.

Finally, the failure of models to achieve the expected outcome can also be due to a crash (minimum distance gap between trajectories $\leqslant 0.1$m) instead of an alternate outcome (e.g. getting stuck in the intersection). In all the simulations, for the parking pullout and intersection clearance we did not observe a crash for any of the models. However, for the merge before intersection, on account of starting at a more riskier situation than the other two in terms of chance of a crash, the crash rate (ratio of crashes across all simulations) for the models across all initial states and agent types were as follows: (dlk($\mathcal{A}$): 0.052, MSPE: 0.022, SSPE: 0.007, QLk(level-1): 0.026, Robust: 0.053). QLk (level-1) demonstrates conservative behavior primarily due to max-max behavior of level-0 agent, where always best responding to maxmax behavior ends up being more conservative than best-responding to diverse models by belief updating (as in robust and dlk(A)). This is reflected in the lower crash rate for QLk model compared to Robust. Equilibrium models lead to reduced collision rate compared to dlk(A) and robust, likely a result of working under a complete information setting where there is no scope for misinterpreting the other agents' type. In terms of success rate, MBI also shows lower success rate for all models. This is mainly because the situation (low distance gap between vehicle during game initiation) is setup as such that for most models the optimal strategy profile is merging vehicle waiting for the on lane vehicle to pass. In fact, the WMA dataset had one instance of

MBI scenario, which resulted in failure. Similarly, we note that collisions are also observed only in the MBI scenario.

Overall, whether or not an AV planner can succeed in their desired outcome depends on a variety of factors, such as, the assumption the vehicle and the human drivers hold over each other, the safety aspiration of each agent, as well as the specific state of the traffic situation. The analysis presented above helps in quantifying the relation between the desired outcome and the criteria under which it is possible.

Conclusion

We developed a unifying framework of modeling human driving behavior and strategic behavior planning of AVs that supports heterogeneous models of strategic and non-strategic reasoning. The model consists of three layers of increasing ability for strategic reasoning, and each layer can hold multiple behavior models. For the non-strategic layer, we also developed a model of level-0 behavior for level-k type models through the use of automata strategies ($dLk(\mathcal{A})$) that is suitable as a non-strategic model for the context of modeling driving behavior. The evaluation on two large naturalistic datasets shows that a combination of a rich level-0 behavior can capture most of the driving behavior as observed in the dataset. On the other hand, for the problem of behavior planning, with the awareness that there can be different types of reasoners in the population, an approach of robust response to heterogeneous behavior models is not only effective, but also is stable across a population of drivers with different levels of risk tolerance.

References

Aghassi, M.; and Bertsimas, D. 2006. Robust game theory. *Mathematical programming*, 107(1): 231–273.

Al Jawahiri, A. 2018. *Spline-based Trajectory Generation for Autonomous Truck-Trailer Vehicles in Low Speed Highway Scenarios*. Master's thesis, Delft University of Technology.

Bae, I.; Moon, J.; and Seo, J. 2019. Toward a comfortable driving experience for a self-driving shuttle bus. *Electronics*, 8(9): 943.

Bock, J.; Krajewski, R.; Moers, T.; Runde, S.; Vater, L.; and Eckstein, L. 2020. The inD Dataset: A Drone Dataset of Naturalistic Road User Trajectories at German Intersections. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, 1929–1934.

Camerer, C. F.; Ho, T.-H.; and Chong, J.-K. 2004. A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, 119(3): 861–898.

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

Fuller, R. 2008. Recent developments in driver control theory: from task difficulty homeostasis to risk allostasis. In *Proceedings of the International Conference on Traffic and Transport Psychology, Washington, USA*, volume 31.

Geanakoplos, J. 1992. Common knowledge. *Journal of Economic Perspectives*, 6(4): 53–82.

Geiger, P.; and Straehle, C.-N. 2021. Learning Game-Theoretic Models of Multiagent Trajectories Using Implicit Layers. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35: 4950–4958.

Ilievski, M. 2020. *WiseBench: A Motion Planning Benchmarking Framework for Autonomous Vehicles*. Master's thesis, University of Waterloo.

Kelly, A.; and Nagy, B. 2003. Reactive nonholonomic trajectory generation via parametric optimal control. *The International Journal of Robotics Research*, 22(7-8): 583–601.

Lewis-Evans, B. 2012. *Testing models of driver behaviour*. Ph.D. thesis, University Library Groningen.

Li, C.; and Czarnecki, K. 2019. Urban Driving with Multi-Objective Deep Reinforcement Learning. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 359–367.

Li, N.; Kolmanovsky, I.; Girard, A.; and Yildiz, Y. 2018. Game theoretic modeling of vehicle interactions at unsignalized intersections and application to autonomous vehicle control. In *2018 Annual American Control Conference (ACC)*, 3215–3220. IEEE.

Li, S.; Li, N.; Girard, A.; and Kolmanovsky, I. 2019. Decision making in dynamic and interactive environments based on cognitive hierarchy theory, Bayesian inference, and predictive control. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2181–2187. IEEE.

Marks, R. E. 1992. Repeated games and finite automata. In *Recent Developments in Game Theory*. Edward Elgar.

Michon, J. A. 1985. A critical view of driver behavior models: what do we know, what should we do? In *Human behavior and traffic safety*, 485–524. Springer.

Najm, W. G.; Smith, J. D.; Yanagisawa, M.; et al. 2007. Pre-crash scenario typology for crash avoidance research. Technical report, United States. National Highway Traffic Safety Administration.

Pruekprasert, S.; Zhang, X.; Dubut, J.; Huang, C.; and Kishida, M. 2019. Decision making for autonomous vehicles at unsignalized intersection in presence of malicious vehicles. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2299–2304. IEEE.

Rubinstein, A. 1986. Finite automata play the repeated prisoner's dilemma. *Journal of economic theory*, 39(1): 83–96.

Sadigh, D.; Sastry, S.; Seshia, S. A.; and Dragan, A. D. 2016. Planning for autonomous cars that leverage effects on human actions. In *Robotics: Science and Systems*, volume 2, 1–9. Ann Arbor, MI, USA.

Sagberg, F.; Selpi; Bianchi Piccinini, G. F.; and Engström, J. 2015. A review of research on driving styles and road safety. *Human factors*, 57(7): 1248–1275.

Sarkar, A.; and Czarnecki, K. 2021. Solution Concepts in Hierarchical Games Under Bounded Rationality With Applications to Autonomous Driving. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 5698–5708.

Schwarting, W.; Pierson, A.; Alonso-Mora, J.; Karaman, S.; and Rus, D. 2019. Social behavior for autonomous vehicles. *Proceedings of the National Academy of Sciences*, 116(50): 24972–24978.

Stirling, W. C. 2003. *Satisficing Games and Decision Making: with applications to engineering and computer science*. Cambridge University Press.

Sun, L.; Cai, M.; Zhan, W.; and Tomizuka, M. 2020. A Game-Theoretic Strategy-Aware Interaction Algorithm with Validation on Real Traffic Data. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 11038–11044. IEEE.

Tian, R.; Li, S.; Li, N.; Kolmanovsky, I.; Girard, A.; and Yildiz, Y. 2018. Adaptive game-theoretic decision making for autonomous vehicle control at roundabouts. In *2018 IEEE Conference on Decision and Control (CDC)*, 321–326. IEEE.

Tian, R.; Liting, S.; Masayoshi, T.; and David, I. 2021. Anytime Game-Theoretic Planning with Active Reasoning About Humans' Latent States for Human-Centered Robots. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE.

Vaa, T. 2011. Drivers' information processing, decision-making and the role of emotions: Predictions of the Risk Monitor Model. In *Human Modelling in Assisted Transportation*, 23–32. Springer.

Wright, J. R.; and Leyton-Brown, K. 2010. Beyond equilibrium: Predicting human behavior in normal-form games. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*.

Wright, J. R.; and Leyton-Brown, K. 2020. A formal separation between strategic and nonstrategic behavior. In *Proceedings of the 21st ACM Conference on Economics and Computation*, 535–536.

Ziegler, J.; and Stiller, C. 2009. Spatiotemporal state lattices for fast trajectory planning in dynamic on-road driving scenarios. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1879–1884. IEEE.