Solution Concepts in Hierarchical Games Under Bounded Rationality With Applications to Autonomous Driving

Atrisha Sarkar [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,k2czarne@uwaterloo.ca

Abstract

With autonomous vehicles (AV) set to integrate further into regular human traffic, there is an increasing consensus of treating AV motion planning as a multi-agent problem. However, the traditional game theoretic assumption of complete rationality is too strong for the purpose of human driving, and there is a need for understanding human driving as a *bounded rational* activity through a behavioral game theoretic lens. To that end, we adapt three metamodels of bounded rational behavior; two based on Quantal level-k and one based on Nash equilibrium with quantal errors. We formalize the different solution concepts that can be applied in the context of hierarchical games, a framework used in multi-agent motion planning, for the purpose of creating game theoretic models of driving behavior. Furthermore, based on a contributed dataset of human driving at a busy urban intersection with a total of ~4k agents and ~44k decision points, we evaluate the behavior models on the basis of model fit to naturalistic data, as well as their predictive capacity. Our results suggest that among the behavior models evaluated, modeling driving behavior as pure strategy Nash Equilibria with quantal errors at the level of maneuvers with bounds sampling of actions at the level of trajectories provides the best fit to naturalistic driving behavior, and there is a significant impact of situational factors on the performance of behavior models.

Introduction

Motion planners are a critical component of autonomous vehicle (AV) architecture, and the decisions made by the algorithms impact the safety of road users, such as pedestrians, cyclists, and other human-driven vehicles. Traditional approaches to motion planning have typically treated the problem as a single-agent problem; in this perspective, a vehicle interacts with the environment (in simulation or on-field setting), possibly with the help of recorded human-driven trajectories, and plans its actions by optimizing over its objectives while taking into account the dynamic obstacles in the vicinity (Schwarting, Alonso-Mora, and Rus 2018; Ilievski et al. 2019). However, in reality human driving is a complex system with a symbiotic relation among agents, where actions of a vehicle influence the future actions of other road users and vice versa. More recently, there has been a focus towards treating motion planning of AVs as a multi-agent

problem with game-theoretic solutions to AV decision making (Fisac et al. 2019; Sadigh et al. 2016; Camara et al. 2018; Li et al. 2018). Such approaches can account for heterogeneous objectives in a group of vehicles in a traffic scene and identify equilibrium solutions that guide the actions of the AV. Given that the movement dynamics of a vehicle is in continuous domain, it is intuitive to model the dynamics as a differential game, an approach adopted by multiple models in the literature (Fridovich-Keil et al. 2019; Sadigh et al. 2016; Wang et al. 2019). However, the applicability of such games as a general purpose planner is limited by the trade off between the computational burden and expressivity; cases where efficient solutions exist in a multi-agent setting restrict the behavior of the agents to only linear dynamics (Fridovich-Keil et al. 2019). As an alternative, (Fisac et al. 2019) introduced the concept of a hierarchical game for AV planning where the game is decomposed into two levels; a long-horizon strategic game that can model richer agent behavior and a short-horizon tactical game with simplified information structure. Although hierarchical games are well suited and show promising results for planning in AV, for the models to be applicable in real world situations, we need to understand how well the stationary concepts in the game match naturalistic human driving behavior. It is well known that in many realistic settings, the theoretical fixed point of Nash equilibrium is a poor predictor of human behavior (Goeree and Holt 2001); therefore, it is necessary to investigate if the same is true for human driving behavior too. In absence of that information, we do not know *what* to optimize for.

Behavioral game theory provides a framework to analyse decision making in a naturalistic setting and models of behavior that often have higher predictive power than Nash equilibria (Camerer 2011). A key element in behavioral game theory is *bounded rationality*, where the conventional game-theoretic notion of agents as fully rational is relaxed to allow for sub-optimal behavior. Such behavior may arise from limitations in cognitive reasoning, or error-prone actions (Samuelson 1995). Driving is a cognitively demanding job that requires situational awareness and sophisticated visuomotor co-ordination, added on to individual habits, biases, and preferences; and it is not hard to imagine that driving at its core is a bounded rational activity. Consequently, it becomes essential for AV game theoretic planners to be

able to characterize the bounded rational behavior in human driving; for example, if humans are prone to making error in judgement when the signal is about to turn red from amber at a busy intersection, then the AV planner should take that into account since the safety of the AV decision is conditioned on the error made by the human driver. Therefore, developing a game-theoretic planner for an AV is a multi-step process, broadly involving a) selection of the right behavior model and equilibrium concepts for other road agents, b) estimation of the parameters of the model, and c) generation of a safe maneuver and trajectory after accounting for the model and its parameters. In this paper, we primarily focus on the first two aspects.

Wright and Leyton-Brown developed a general framework of analysing and estimating parameters of popular behavioral game theory models based on observations of game play. They focus on two models of behavior, i.e. Quantal Level-k (QLk) and Poisson-Cognitive Hierarchy (P-CH) (Wright and Leyton-Brown 2012), which model iterated reasoning where agents have a limited capacity to maintain higher order belief about other agents. Although QLk and P-CH do not capture all types of bounded rationality that one can think of in the case of human driving, such as the ones that arise from sampling the actions of other agents, the framework developed in (Wright and Leyton-Brown 2012) nevertheless can be applied to a wider set of behavior models including the ones we develop in this paper.

In this paper we make the following contributions: (i) Building upon (Fisac et al. 2019), we further formalize a general model of a hierarchical game for the purpose of modeling human driving behavior. (ii) We extend three models of behavior from behavioral game theory, two based on Quantal level-k and one based on Nash equilibrium with quantal errors, and demonstrate the possible solution concepts that can be applied to hierarchical games. (iii) In order to better understand strategic and non-strategic decision making in human driving, we compare 25 game theoretic models based on a cross sectional study of human drivers at a busy urban intersection with a total of 3913 agents and 43765 decision points. We make this dataset, which is one of the largest multi-agent behavior dataset of human driving, publicly available as a contribution.

Hierarchical Games

Prior to recent focus in autonomous driving, there has been considerable body of research on modeling driving behavior within the field of traffic psychology with a long history of treating driving behavior as a hierarchical model (Keskinen et al. 2004; Van der Molen and Bötticher 1988; Lewis-Evans 2012; Michon 1985). One of the more influential models, the Michon hierarchy of driving tasks (Michon 1985), decompose driving into three levels of control; a strategic plan such as a route and general speed choice of going from point A to B is decomposed into several tactical decisions of choosing the right maneuvers, which is further decomposed into high fidelity actions that control the steering and acceleration. A primary motivation of a hierarchical decomposition is that drivers have different motivations and risk judgements

in each level of the hierarchy, and the functional decomposition into a hierarchical system allows for modelling the risk and safety considerations separately at each level. Motion planners in autonomous vehicles also follow a similar hierarchical pattern of decomposition; a high level *route planner* plan is given to a *behavior planner*, which sets up the tactical maneuvers for a lower level *trajectory planner*, which in turn generates the trajectory profile for the vehicle *controller* after respecting its nonholonomic constraints. In addition to the motivation mentioned earlier, treating the problem of planning as a hierarchical system is also driven by computational efficiency as previously shown in (Fisac et al. 2019).

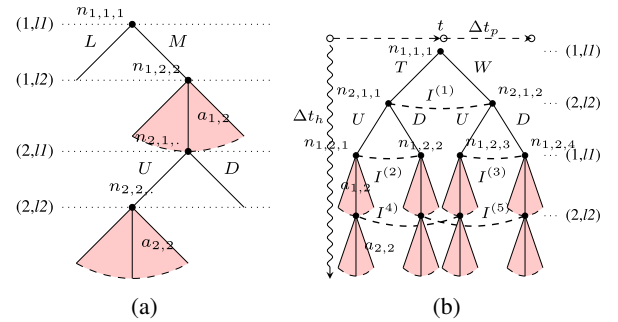In a multi-agent setting, this means that the planning prob-

Figure 1: Illustration of two instances of hierarchical games. (a) As a Stackelberg game modeling a lane change maneuver and (b) simultaneous move game modeling intersection navigation. A hierarchical game is instantiated every Δt_p seconds with action plan of Δt_h seconds.

lem has to be extended to the notion of a hierarchical game, which we formalize further below. A hierarchical game is formulated by

- Set of N agents indexed by $i \in \{1, 2, 3, ..N\}$.
- A set of $\mathcal{K}$ levels indexed by $\kappa \in \{1, 2, 3, ..\mathcal{K}\}$.
- Set of actions $A_{i,\kappa}$ available to each agent i at level κ.
- A strategy s_i for agent i is a $\mathcal{K}$-tuple $s_i = (a_{i,1}, a_{i,2}, .., a_{i,\mathcal{K}})$ where $a_{i,\kappa} \in A_{i,\kappa}$ and the strategy space of s_i is $\prod_{\kappa \in \mathcal{K}} A_{i,\kappa}$.
- A set of states X_i of agent i in level 1, and an initial mapping function $f_{i,1} : X_i \to \mathcal{P}(A_{i,1})$ that maps the initial state of the agent to the available actions in level 1, where $\mathcal{P}(\cdot)$ is the power set.
- Set-valued functions $f_{i,\kappa} : \prod_{j=1}^{\kappa-1} A_{i,j} \to \mathcal{P}(A_{i,\kappa})$ for each agent i that maps a partial strategy $(a_{i,1}, a_{i,2}, .., a_{i,\kappa-1})$ to $\mathcal{P}(A_{i,\kappa})$ and gives the set of available actions to i in level $\kappa > 1$ for the partial strategy till level $\kappa - 1$.
- Set of N pay-off (utility) functions $U = \{u_i(s_i, s_{-i})\}$, where $-i$ refers to all agents other than i.

The hierarchical game imposes a total ordering in actions $A_i = \{A_{i,1}, A_{i,2}, .., A_{i,\mathcal{K}}\}$ of a given agent, and along with

$f_{i,\kappa}$ induces a game tree, as shown in Fig. 1b. The frequency at which a hierarchical game is instantiated (Δt_p) and the time horizon of each strategy (Δt_h) are exogenous to the model. Each node is labeled $n_{i,\kappa,j}$, where i and κ are the agent and level indices, and j is the node identifier within level κ. This general formulation of a hierarchical game does not prescribe a fixed information structure, and allows the designer to set an information structure that is appropriate to the environment and situation they want to model. For example, (Fisac et al. 2019) models a lane change scenario where an AV merges into a lane occupied by a human driven vehicle. The game is modeled as a Stackelberg game with the AV being the leader and the human driven vehicle responding to the action of the AV. Fig. 1a shows a 2-agent 2-level the game tree for such a scenario where each decision node is a singleton information set since for every decision node, the agent who owns the decision node has perfect information on where they are in the game. At node $n_{1,1,1}$, the AV (indexed as agent 1) has the choice of either staying in its current lane (L) or merging into the adjoining lane (M) $A_{1,1} = \{L, M\}$. Conditioned on this choice, the vehicle has to generate a trajectory $a_{1,2} \in f_{1,2}(a_{1,1})$ to execute the maneuver chosen in level 1. Whereas actions in $A_{1,1}$ are discrete choices, the agent can choose from a continuum of actions (shaded region in the figure) at node $n_{1,2,2}$. The human driven vehicle after having observed the actions of the AV, can respond by deciding to speed up (U) to dissuade the merging AV cut-in the front, or slow down (D) followed by a trajectory that corresponds to the choice. In situations where assignment of a leader and a follower is unclear or that assumption is too strong, the agents might not have perfect information on the state of the play. Fig. 1b illustrates a 2-agent 2-level scenario as an example where an AV (indexed as 1) executes a free right turn on red at a signalized intersection (in a situation similar to id:14 in Fig. 2), while a human driven vehicle (id:26 and re-indexed as 2 in Fig. 1b) approaches cross path from left to right. The AV can either decide to turn (T) or wait (W) for the cross path vehicle to pass, i.e., $f_{1,1}(X_1) = A_{1,1} = \{T, W\}$. The human driven vehicle (id:26) can either slow down (D) or choose not to slow down (U), $f_{2,1}(X_2) = A_{2,1} = \{D, U\}$. Since either agent does not have perfect information about what the other agent is about to do next, agent 2 does not know whether they are in node $n_{2,1,1}$ or $n_{2,1,2}$ (connected by the information set $I^{(1)}$). This imperfection of information is also reflected at the trajectory level (level 2 actions), where each agent can only distinguish between the nodes in level 2 that follow from their own chosen actions in level 1, but not from the ones that follow from the other agent's level 1 decision ($I^{(2)}$-$I^{(5)}$).

It becomes apparent from this structure that the game has no proper subgame, and the game reduces to a simultaneous move game. It is well understood that a way to solve such games is by reduction to normal form. However, as we shall see, the hierarchical game has additional constraints that allow solving the game in Fig. 1 also through backward induction. To designate the nodes where utilities accumulate at each level in the backward induction process,

---

Algorithm 1: Backward induction for a hierarchical game

Result: S_1^*, V_1^*

1 **for** $\kappa := \mathcal{K}; \kappa = 1; \kappa := \kappa - 1$ **do**
2   **for** $n \in \mathcal{L}(\kappa)$ **do**
3          $S_{\kappa,n}^*, V_{\kappa,n}^* \leftarrow \text{solve } \mathcal{G}_\kappa(\prod_{i=1}^{N} f_{i,\kappa}(\sigma_i(n)),$
4          $\kappa = \mathcal{K}?U; V_{\kappa+1,\mathcal{L}(\kappa+1)}^*)$
5   **end**
6 **end**

---

we label a set of nodes in each level κ as *level roots* $\mathcal{L}(\kappa) = \{n_{i,\kappa,j} | parent(n_{i,\kappa,j}) \notin \mathcal{N}_\kappa\}$ where $\mathcal{N}_\kappa$ is the set of nodes in level κ. In other words, the set of level roots contain nodes in each level κ whose parent is not in level κ. Therefore, $\mathcal{L}(1) = \{n_{1,1,1}\}$ and $\mathcal{L}(2) = \{n_{1,2,1}, n_{1,2,2}, n_{1,2,3}, n_{1,2,4}\}$. Algorithm 1 shows the standard backward induction process adapted to the hierarchical game. The algorithm starts at the bottom most level ($\mathcal{K}$) and recursively moves up the tree by solving the level games $\mathcal{G}_\kappa$ at every level. At each level, a simultaneous move *level game* $\mathcal{G}_\kappa$ is instantiated from each node in $\mathcal{L}(\kappa)$. These level games are constructed by first extracting $\sigma_i(n)$, which gives the partial pure strategy for agent i that lies on the branch from the root node of the game tree $\mathcal{L}(1)$ to node $n \in \mathcal{L}(\kappa)$. $f_{i,\kappa}$ gives the available actions for each agent i in the current level κ, and these actions form the domain of available strategies in the level game $\mathcal{G}_\kappa$. The utilities depend on the level of the game; for level game $\mathcal{G}_{\kappa=\mathcal{K}}$ the utilities are same as the game utility U, whereas for level games $\mathcal{G}_{\kappa<\mathcal{K}}$ are solved based on the game values $V_{\kappa+1,\mathcal{L}(\kappa+1)}^*$ from the game $\mathcal{G}_{\kappa+1}$ solved in the previous iteration. Note that the pseudocode shows only the case where a single solution and game value ($S_{\kappa,n}^*, V_{\kappa,n}^*$) is propagated up the hierarchy. In the case of multiple solutions for the level games, the strategies and values have to be tracked and repeated for each solution. The solutions and game value $S_{\kappa,n}^*, V_{\kappa,n}^*$ depend on the solution concept used for the individual level game, and this is discussed in detail later under Solution concepts.

One can see that the backward induction process is very similar, if not same as solving for subgame perfect equilibria in multi-stage games with stages being replaced by levels in the hierarchy (Tadelis 2013). However, we cannot call it that since the level games are not subgames in the game tree. The reason why the backward induction works though is because the mapping functions $f_{i,\kappa}$ eliminate strategies for every agent i that are not direct successors of the partial strategies $\sigma_i(n) \cdot \sigma_{-i}(n)$, essentially breaking any information set within a level κ that spans across two separate level roots in $\mathcal{L}(\kappa)$. More intuitively, this mimics the elimination of hypothetical strategies where in level 1 a vehicle may think about slowing down, but in level 2 chooses a trajectory that speeds up; and the fact that this cannot happen is part of the common knowledge among the agents in the game.
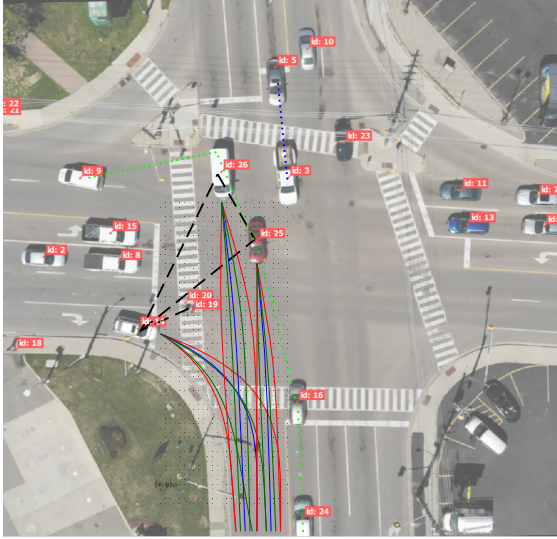
Figure 2: A snapshot of the intersection traffic scene. Representative trajectories based on the three sampling schemes over a R^3. The figure shows the path (R^2) projection of the trajectories.

Game Structure

In this section, we describe the details of the game structure used in our study, including the number of agents, actions/strategies, and utilities.

Relevant agents and available actions. Since we are interested in investigating decision making in the most critical tasks at a signalized intersection (such as unprotected left turns and right turns on red), at each time step Δt_p=1s, we setup a hierarchical game with an action plan horizon of Δt_h=5s into the future from the perspective of each vehicle that is turning left or right in the scenario. For example, in the snapshot of Fig. 2, the black dashed line shows the game from the perspective of the vehicle 14, which is turning right. The process of including the relevant vehicles in the game is as follows: we identify the conflict points on the map with respect to all the lanes in the intersection that cross each other. Since a game is initiated with respect to a 'subject vehicle', we first locate the conflict points corresponding to the lane that the subject vehicle is currently in, and include all agents in the scene that are on the lanes in conflict with the subject vehicle's lane. We also include the leading vehicle of these conflicting vehicles. This set of relevant vehicles along with the subject vehicle form the set of agents in each game. Pedestrian actions are not modeled explicitly in the game tree; however, their influence is modeled in the utility structure of the game, which is described later in the section. Each game is a N-player 2-level hierarchical game where level 1 actions for each agent are high level maneuvers that are relevant to the task under execution, and level 2 actions are the corresponding trajectories. We setup the set of maneuvers with the help of a rule engine that takes into account the task of the vehicle and its situational state (position, velocity, etc.). The complete list of level 1 actions

is documented in the code repository (c.f. Experiments section). Level 2 actions ($A_{i,2}$) are trajectories that are generated based on the actions in level 1. To generate the trajectories for each vehicle, we use a lattice sampling based trajectory generation similar to one presented in (Ziegler and Stiller 2009). First a set of lattice endpoints are sampled on R^2 cartesian co-ordinate centered on the vehicle's current position. Each lattice sample point on R^2 is then extended with a temporal lattice which is re-sampled to form the final lattice points in R^3 that contain the (x, y) positions and the target velocity at each lattice point after accounting for acceleration and jerk limits of passenger vehicles (Bae, Moon, and Seo 2019). Finally, the sampled lattice points are connected with a smooth cubic spline representing the vehicle trajectory (Fig. 3).

Since the trajectory generation is in continuous space with infinite actions for the drivers to reason over, combined with the time constraints to make a decision (which is in the order of milliseconds), the situation is ripe for bounded rationality to be in play. Osborne and Rubinstein takes a view of bounded rationality that emerges from agents' employing a mental process to sample other agents' actions and respond based on the imagined outcome of those samples (Osborne and Rubinstein 1998). In our case, this is akin to a vehicle sampling a set of trajectories of other agents and responding in accordance to the sampled trajectories. Naturally, one may imagine that some sampling procedures make more sense than others. We now briefly mention the sampling procedures used in our experiments, and the intuitive reasoning behind each.
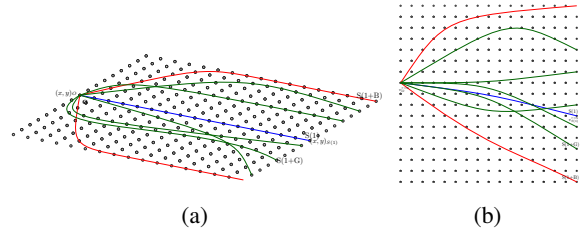

(a)   (b)

Figure 3: Representative trajectories based on the 3 sampling schemes over a R^3 lattice showing the spatial representation of the (a) path and (b) velocity profiles. Lattice points are connected with cubic splines.

At each time step when the game tree is instantiated, agents observe the current attributes (such as position, velocity and acceleration) of other relevant agents in the game tree. In the most basic case, an agent i may sample a single trajectory (level 2 action) of every agent $-i$ that that they, i.e., i, think is most representative of the level 1 action of the agent they are currently reasoning over. To construct the trajectory sample, we select lattice endpoints along the lane centerline and use a piecewise constant acceleration model to generate the final trajectory. In the subsequent sections, we refer to this sampling scheme that produces a single trajectory as $S(1)$ sampling. With a little more cognitive bandwidth, along with the $S(1)$ trajectory sample, they can also sample trajectories that form the extreme ends of

the bounded level-2 action space of other agents. These trajectories are bounded spatially by the lane boundaries and temporally by the upper and lower bounds on the velocity limits of the level 1 action they correspond to. We refer to this scheme as bounds sampling $S(1+B)$. This set of trajectories indicate what other agents might do in normative (i.e. following the rules as captured by the piecewise constant acceleration model) as well as in the extreme case but still within the physical limitations of the vehicle. The final sampling scheme lies in between the two schemes. Similar to $S(1+B)$, this scheme includes the $S(1)$ trajectory; however, the rest of the trajectories are sampled from a multivariate Gaussian distribution with $\mu = [x_{S(1)}, y_{S(1)}, v_{S(1)}]^\top$ and an unit diagonal covariance matrix, where $(x_{S(1)}, y_{S(1)}, v_{S(1)})$ is the lattice endpoint corresponding to the $S(1)$ trajectory. We refer to this scheme as $S(1+G)$ and the samples include the normative behavior that comes from $S(1)$ along with variations in the path and velocity of the vehicle but not to the extremes that were captured in the $S(1+B)$ scheme.

Utilities. To determine the utility structure, we draw from motivational aspects of driver behavior modelling in traffic psychology literature (Summala 1988). In general, driving motivations can be broadly classified into *inhibitory* and *excitatory*. Whereas excitatory motivations drive a driver to make progress towards reaching the destination, inhibitory motivations are the balancing factors that account for mitigating crashes and mental stress. In our case, the degree of progress a driver can make based on a selected trajectory $a_{i,2}$ is the excitatory utility $u_{\text{v\_exc}}(a_{i,2})$ as determined by the trajectory length $\|a_{i,2}\|$, $u_{\text{v\_exc}}(a_{i,2}) = \min(\frac{\|a_{i,2}\|}{d_g}, 1)$, where d_g is a constant and can be interpreted as the distance to goal or crossing the intersection. Inhibitory utility is based on the minimum distance gap of the trajectory to other vehicles $u_{\text{v\_inh}}$ as well as respecting pedestrian's right of way $u_{\text{p\_inh}}$. The final form of the utility function is

$$u_i(a_{i,2}, a_{-i,2}) =$$
$$\mathbf{W} \cdot [u_{\text{v\_inh}}(a_{i,2}, a_{-i,2}) \quad u_{\text{p\_inh}}(a_{i,2}) \quad u_{\text{v\_exc}}(a_{i,2})]^\top$$
$$u_{\text{v\_inh}}(a_{i,2}, a_{-i,2})$$
$$= \int \text{erf}\left[\frac{d(a_{i,2}, a_{-i,2}) - \theta}{\sigma\sqrt{2}}\right] \mathcal{N}(\theta; d^*_{a_{i,2}, a_{-i,2}}, \sigma) d\theta$$

Sigmoidal functions are a popular family of functions that map a safety surrogate metric, e.g., distance gap $d(a_{i,2}, a_{-i,2})$, into an utility interval (Fishburn 1970). For $u_{\text{v\_inh}}$, we first fix a minimum safe distance gap $d^*_{a_{i,2}, a_{-i,2}}$ based on the task (left turn, right turn, etc.) of the agents in the game. The value of the safe distance gap determines the location θ of the sigmoidal function (erf). However, since the conception of what is considered safe may vary in a population of drivers, we let θ to be a random variable that is normally distributed with $\mu = d^*_{a_{i,2}, a_{-i,2}}$ and constant variance σ determining the scale of the sigmoidal function. The choice of erf as the sigmoidal function is a mathematical convenience since the Gaussian integral of the erf in $u_{\text{v\_inh}}(a_{i,2}, a_{-i,2})$ evaluates to another sigmoidal $\text{erf}(\frac{d(a_{i,2}, a_{-i,2}) - d^*_{a_{i,2}, a_{-i,2}}}{2\sigma})$. $u_{\text{p\_inh}}$ is a step function over [-

1,1] such that $u_{\text{p\_inh}}(a_{i,2}) = -1$ if $a_{i,2}$ is a trajectory that does not wait for a pedestrian when the pedestrian is in the vicinity having a right of way, or is on the crosswalk to be traversed; and 1 otherwise. $\mathbf{W}$ is the weight parameter that combines the inhibitory and excitatory utilities together to produce a single real value. Utilities for the actions in $\mathcal{G}_1$ can be calculated as follows. $u_i(a_{i,1}, a_{-i,1}) = V^*_{2,\eta}(i)$, where η is the leaf node of the branch $a_{i,1}, a_{-i,1}$ and $V^*_{2,\eta}(i)$ is the utility of agent i following the pure strategy response $a^*_{i,2}$, where $a^*_{i,2}$ is the solution to the underlying level-2 game.

Solution Concepts

A key element that influences solution concepts in games is the manner in which each agent reasons over the strategies of other agents. In non-strategic behavior models, agents do not explicitly model other agents in the game and respond solely on the basis of their own utility structure (Wright and Leyton-Brown 2020). Strategic agents, on the other hand, perform some reasoning over the strategies of other agents and respond accordingly.

The first category of behavior models we consider is the Quantal level-k (QLk) model (Wright and Leyton-Brown 2012). QLk models the population of agents as a mix of strategic and non-strategic agents, with strategic agents having an iterated cognitive hierarchy of reasoning. Strategic agents in QLk use Quantal Best Response (QBR) function, often expressed as a logit response $\pi_i^{\text{QBR}}(a_i, s_{-i}, \lambda) = \frac{\exp[\lambda \cdot u_i(a_i, s_{-i})]}{\Sigma_{a'_i} \exp[\lambda \cdot u_i(a'_i, s_{-i})]}$, where s_{-i} represent the pure or mixed strategies of other agents and λ is the *precision* parameter that can account for errors in agent response with respect to utility differences[1]. When $\lambda \to 0$, the mixed response is a uniform random distribution, whereas $\lambda \to \infty$ makes the response equivalent to best response. Level-0 agents are non-strategic (NS) agents who choose their actions uniformly at random, whereas Level-1 agents are strategic (S) agents who believe that the population consists solely of Level-0 agents, and their response is a QBR response to Level-0 agents' actions. In the original QLk model, level-0 agents follow an uniform distribution mixed strategy; however, in our case we use an expanded definition of level-0 agents presented in (Wright and Leyton-Brown 2014), where instead of an uniform distribution, the level-0 agents' strategies follow more intuitive yet non-strategic response, such as maxmax response (MX) or maxmin (MM) response. We believe that the expanded definition of the level-0 agents suit our situation much better, since it is unrealistic to expect a driver to choose actions purely at random from their available actions. Even with this expanded definition, these are still *non-strategic* since agent responses depend purely on their own utilities and do not rely on a strategic reasoning over other agents' utilities (Wright and Leyton-Brown 2020).

In a hierarchical game, since the agent strategies are factored into levels $s_i = (a_{i,1}, a_{i,2})$, the manner in which an agent reasons over strategies in one level might not be the

---

[1] In this formulation, the symbols s_i and a_i are strategies and actions of a game in a general sense, and not related to the symbols used specifically in the formulation of hierarchical games earlier.

QL0		QL1		PNE-QE	
$\mathcal{G}_1$	$\mathcal{G}_2$	$\mathcal{G}_1$	$\mathcal{G}_2$	$\mathcal{G}_1$	$\mathcal{G}_2$
NS	NS	S+NS	NS	S	NS

Table 1: Distribution of strategic (S) and non-strategic (NS) behavior in level games $\mathcal{G}_1$ and $\mathcal{G}_2$ in three metamodels QL0, QL1, and PNE-QE.

same as the reasoning process in another level. Therefore, instead of a single solution concept in the game of Fig. 1, the level games $\mathcal{G}_2$ can have a different solution concept than the one in game $\mathcal{G}_1$. In our models, we let agents have a cognitively less demanding non-strategic response in $\mathcal{G}_2$, and a more deliberative strategic response in $\mathcal{G}_1$. This choice is similar to one taken in (Fisac et al. 2019), and reflects the natural process where it is easier for drivers to reason strategically over the strategy space of discrete maneuvers than over the space of infinitely many trajectories.

We consider two metamodels of behavior under QLk: QL0 and QL1. We refer to them as metamodels, since they can be further refined based on the choice of response function and sampling schemes to create concrete models. In QL0 models, we restrict the population to be solely level-0 responders in both $\mathcal{G}_1$ and $\mathcal{G}_2$. In QL1, the population consist of a mix of level-0 and level-1 responders in $\mathcal{G}_1$ and level-0 responders in $\mathcal{G}_2$ (Table 1). For models of non-strategic behavior, we use two response functions; maxmax response (MX) and best worst-case or maxmin response (MM). The model of MX is:

$$a_{i,\kappa}^* = \underset{\forall a_{i,\kappa}, a_{-i,\kappa}}{\operatorname{argmax}} u_i(a_{i,\kappa}, a_{-i,\kappa}) \tag{1}$$

$$\pi_i(a_{i,\kappa})$$
$$= \frac{\exp[\lambda_i \cdot u_i(a_{i,\kappa}, \operatorname{argmax}_{\forall a_{-i,\kappa}} u_i(a_{i,\kappa}, a_{-i,\kappa}))]}{\Sigma_{\forall a_{i,\kappa}} \exp[\lambda_i \cdot u_i(a_{i,\kappa}, \operatorname{argmax}_{\forall a_{-i,\kappa}} u_i(a_{i,\kappa}, a_{-i,\kappa}))]} \tag{2}$$

where a_i^* is the pure strategy utility maximizing action for i. The model for non-strategic MM response is:

$$a_{i,\kappa}^* = \underset{\forall a_{i,\kappa}}{\operatorname{argmax}} \underset{\forall a_{-i,\kappa}}{\operatorname{argmin}} u_i(a_{i,\kappa}, a_{-i,\kappa}) \tag{3}$$

$$\pi_i(a_{i,\kappa})$$
$$= \frac{\exp[\lambda_i \cdot u_i(a_{i,\kappa}, \operatorname{argmin}_{\forall a_{-i,\kappa}} u_i(a_{i,\kappa}, a_{-i,\kappa}))]}{\Sigma_{\forall a_{i,\kappa}} \exp[\lambda_i \cdot u_i(a_{i,\kappa}, \operatorname{argmin}_{\forall a_{-i,\kappa}} u_i(a_{i,\kappa}, a_{-i,\kappa}))]} \tag{4}$$

Equations 2 and 4 are relaxations that translate the pure strategy action to a noisy response $\pi_i(a_{i,\kappa})$ based on the precision parameter λ_i and sensitivity to i's utility difference with respect to opponent actions that maximizes i's utility for MX and minimizes for MM.

In QL1 metamodel, the population consists of a mix of level-0 and level-1 agents. Level-0 agents in this population follow non-strategic behavior as formulated earlier and level-1 agents best responds quantaly to level-0 agents' behavior. With the expanded definition of level-0 agents as non-strategic bounded rational agents, there is a design choice to be made on what level-1 agents believe

about level-0 agents. They can either consider level-0 agents bounded rational having mixed response of Equations 2 and 4, or level-1 agents can consider level-0 agents to be pure strategy rational responders based on Equations 1 and 3. We choose the later to align with the original QLk model, where agents modeling other agents as bounded rational agents are observed only at a higher cognitive level (level-2 and above). In QLk models, mixed population is modeled as uniform population of bimodal mixture behavior. Therefore, if the proportion of level-0 and level-1 agents is α and $1 - \alpha$ respectively, then the QL1 model response in $\mathcal{G}_1$ is the mixed strategy response

$$\pi_i^{\text{QL1}}(a_{i,1}) = \alpha \cdot \pi_i^{\text{QL0}}(a_{i,1}) + (1 - \alpha) \cdot \pi_i^{\text{QBR}}(a_{i,1}, a_{-i,1}^*, \lambda_i) \tag{5}$$

where $\pi_i^{\text{QL0}}(a_{i,1})$ is the left hand side of the equation 2 or 4 and $a_{-i,1}^*$ is the solution set to equations 1 or 3 for each of the other agents.

The final metamodel we consider is a generalization of pure strategy Nash equilibrium with noisy response. In this metamodel, agents follow a non-strategic model in $\mathcal{G}_2$, and a strategic model in $\mathcal{G}_1$ as described below.

$$a_{i,1}^* = \underset{\forall a_{i,1}}{\operatorname{argmax}} u_i(a_{i,1}, a_{-i,1}^*) \tag{6}$$

$$\pi_i(a_{i,1})$$
$$= \frac{\exp[-\lambda_i \cdot \min_{\forall(a_{i,1}^*, a_{-i,1}^*)}(u_i^* - u_i(a_{i,1}, a_{-i,1}^*))]}{\Sigma_{\forall a_{i,1}} \exp[-\lambda_i \cdot \min_{\forall(a_{i,1}^*, a_{-i,1}^*)}(u_i^* - u_i(a_{i,1}, a_{-i,1}^*))]} \tag{7}$$

where $u_i^* = u_i(a_{i,1}^*, a_{-i,1}^*)$. In the above model, agents respond according to pure strategy Nash equilibria $a_{i,1}^*$, but in error may choose actions $a_{i,1} \notin a_{i,1}^*$ based on the sensitivity to the difference in the utility of the action and an equilibrium action. We refer to this model as pure strategy Nash equilibria with quantal errors (PNE-QE). The formulation is similar to Quantal Response Equilibrium (QRE), yet with key differences. In QRE, strategic reasoning occurs in a space of mixed responses and the precision parameter is part of common knowledge in the game. In our model, reasoning over opponent strategies is in pure strategy action space and the precision parameter is endogenous to each agent; therefore, when an agent reasons about the strategies of other agents, their parameters do not play a role (Crawford, Costa-Gomes, and Iriberri 2013). Based on the choice of the metamodel, the response function, and the sampling scheme, we get 25 different behavior models ($\mathcal{B}$), cf. Fig. 4, which we evaluate in the next section.

Estimation of game parameters. Our dataset contains instances of $\mathcal{D}$ (~23k) hierarchical games, instantiated with $\Delta t_p = 1s, \Delta t_h = 5s$ and with the state variables X_i along with the observed strategy $s_i^o = (a_{i,1}^o, a_{i,2}^o)$ for every agent i in the game. For each behavior model $b \in \mathcal{B}$, we note the errors in actions with respect to the pure strategy responses in the games as $\Delta \mathcal{U}_b = \{\epsilon_{i,b} | \epsilon_{i,b} = \min_{\forall a_i^*}[u_i(a_i^*, a_{-i}^*) - u_i(a_i^o, a_{-i}^*)]\}$, where a_i^* are the solutions to Equations 1 or 3 for non-strategic models and 6 for PNE-QE model (we verified the existence of pure strategy NE for all $\mathcal{G}_1$ games

in $\mathcal{D}$). Within the context of a game, we assume that all agents follow the same behavior model, and the precision parameters ($\lambda_{i,b}$) in an individual game is a function of the agent's state X_i from whose perspective the game is initiated as well as the behavior model b of the game. Therefore, for a given state factor X_i, $\epsilon_{i,b}$ follows an exponential distribution based on the game's precision parameter for non-strategic and PNE-QE models, and a mixed exponential distribution (5) for QL1 in $\mathcal{G}_1$. To estimate the value of $\lambda_{i,b}$ we fit a generalized linear model $glm(\epsilon_{i,b} \sim \boldsymbol{\beta} X_i)|_{\Delta\mathcal{U}_b}$ with Gamma($k = 1$) family and inverse link, which models $\epsilon_{i,b}$ as an exponentially distributed random variable with $\mathrm{E}[\epsilon_{i,b}] = \frac{1}{\lambda_{i,b}}$ and $\mathrm{Var}[\epsilon_{i,b}] = \frac{1}{\lambda_{i,b}^2}$. $\boldsymbol{\beta}$ is the model co-efficient, solved through maximum likelihood estimate based on the data in $\Delta\mathcal{U}_b$. The prediction of the glm model gives the mean and standard error of $\lambda_{i,b}^{-1}$ based on the state observation X_i. For the mixed exponential distribution in QL1 model, once we estimate the individual precision parameters of 5, we use iterative gradient ascent to solve for α by maximizing the likelihood function $\Sigma_{\forall a_{i,1}^o} \ln(\pi_i^{\mathrm{QL1}}(a_{i,1}^o))$.

Experiments

Dataset. The dataset contains a total of 3649 vehicles and 264 pedestrians, including their centimetre-accurate trajectory estimates. We analyse the decision making in right turning and left turning vehicles, which results in a total of 12526 hierarchical games. The process of data collection was as follows: we recorded traffic footage from a busy intersection during mid-day traffic in Waterloo, Canada, with an overhead drone and used a third-party service to label the vehicles, pedestrians, bicycles, etc in the video recordings [2]. We added additional metadata information about the trajectories, including lane segments, the state of the traffic lights, the true level-1 maneuver of the vehicles, and conflict points in the map. Dataset along with the related documentation is available at git.uwaterloo.ca/a9sarkar/traffic_behavior_modeling.

In our experiment we study naturalistic driving behavior and evaluate which behavior model captures human driving better, both in terms of model fit and predictive accuracy. In our experiments, we study behaviors after setting $\mathbf{W} = \begin{bmatrix} 0.25 & 0.5 & 0.25 \end{bmatrix}$, thereby giving more importance to pedestrian inhibitory actions and set the value of $d_g = 100\,\mathrm{m}$. In particular we answer the following research questions:

- *RQ1*. Which solution concept provides the best explanation for the observed naturalistic data?

- *RQ2*. How do state factors influence the precision parameters in the games?

- *RQ3*. How does the choice of the response function in the lower level game $\mathcal{G}_2$ affect the higher level solutions in $\mathcal{G}_1$?

Models are indexed by their metamodel followed by the choices of response functions in $\mathcal{G}_1$:$\mathcal{G}_2$ followed by the sampling scheme used in $\mathcal{G}_2$. For models using S(1) sampling

---

[2] datafromsky.com

of trajectories, the response function in $\mathcal{G}_2$ is omitted since the hierarchical game only consists of $\mathcal{G}_1$ games; and in those cases each agent has a single choice under each level-2 roots. We perform our analysis of RQs 1 and 2 based on $\mathcal{G}_1$, and discuss the impact of the choice of $\mathcal{G}_2$ solution concepts as a part of RQ3. For the level-1 actions, we use the following maneuvers: wait-for-oncoming, proceed-turn, track-speed, follow-lead, decelerate-to-stop, wait-for-lead-to-cross, follow-lead-into-intersection, wait-on-red, wait-for-pedestrian. Most of the names of these maneuvers are self-explanatory; track-speed applies to straight through vehicles, where the generated trajectory either maintains or accelerates/decelerates to road speed limit. Each maneuver is further divided into aggressive and normal modes, thereby giving a total of 18 level-1 actions.

RQ1. We address this question in three ways; with respect to the (i) parameter values in the model, (ii) predictive accuracy in unseen data, and (iii) model fit. Fig. 4(c) shows the box-plot of $\Delta\mathcal{U}_b$ or the utility difference between the true utility (i.e. utility of observed maneuver) and the utility of the maneuver predicted by the game solutions for each behavior model. A lower value therefore indicates that the solutions are closer to the true maneuver executed by the vehicle. In general, PNE-QE metamodels show lower values in the utility difference compared to Ql0 and Ql1 metamodels, and within the PNE-QE metamodels, utilities of the actions selected by the PNE-QE:MM$_{\mathrm{S(1+B)}}$ model are closest to the utilities of real action selected by the vehicles.

The models can also be compared based on the precision parameter estimates ($\lambda_{i,b}$), where a higher value in $\lambda_{i,b}$ indicate the solutions of the models being closer to the true behavior. However, since $\lambda_{i,b}$ depends on the state of each agent (X_i), we can only compare the parameter across the models after controlling for the agent state. Therefore, for each X_i, we note the predicted $\lambda_{i,b}$ value by the glm model, and Figure 4 (a) shows their distribution across all states for each model b. Since Ql1 models are mixed exponential distributions with two precision parameters (one from level-0 behavior and the other from level-1 behavior) (Eqn 5), Figure 4(a) shows only one precision parameter (the one from level-1 behavior) for Ql1 models. The other parameter is the same as the values shown in the figure for Ql0 models with the same solution concept; for example Ql1:MX$_{\mathrm{S(1)}}$ is a mixed exponential model with two parameters; one shown under Ql1:MX$_{\mathrm{S(1)}}$ corresponding to level-1 behavior and another under Ql0:MX$_{\mathrm{S(1)}}$ model from level-0 behavior. Overall, the proportion of level-1 responders in Ql1 models were $\alpha = 0.519 \pm 0.02$. In general, similar to results in Fig. 4 (c), PNE-QE models show higher values of the precision parameter as well, thus reflecting better performance as a model of behavior in level-1 games, i.e. for selection of maneuvers. PNE-QE:MX$_{\mathrm{S(1+B)}}$ (PNE-QE model with maxmax response in $\mathcal{G}_2$ with bounds sampling) show highest value of the precision parameter ($\lambda = 190.8 \pm 0.57$). Next, we evaluate model fit using Akaike information criterion (AIC) values, which are noted in Fig. 4(a) in brackets. PNE-QE models with bounds sampling of trajectories have lowest AIC values (-263.96 and -267.56 for PNE-QE:MX$_{\mathrm{S(1+B)}}$ and PNE-QE:MM$_{\mathrm{S(1+B)}}$ respectively), indicating the best fit among the
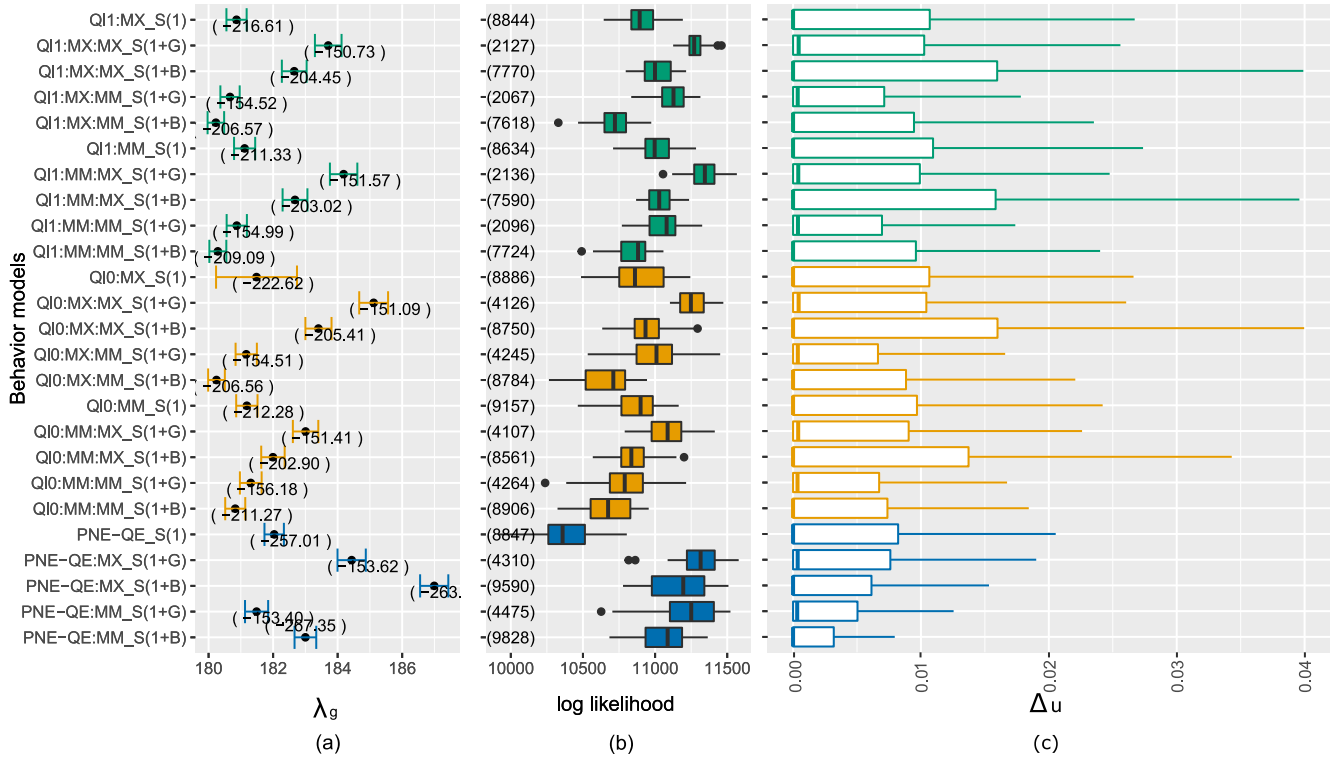
Figure 4: Comparison of the models based on (a) precision parameters ($\lambda_{i,b}$), fit (AIC values noted in brackets), (b) predictive accuracy (log likelihood of observations in test data after 30 runs). (c) Spread of utility differences between selected action ($a^o_{i,1}$) and the solution ($a^*_{i,1}$) in $\mathcal{G}_1$ games for each behavior model..

models based on this criteria.

Alternatively, model selection can also be guided by their predictive power in unseen situations. For evaluation based on this criterion, we use random subsampling with 75:25 training and testing split and 30 runs. The model parameters are estimated based on the observations in the training set, and the predictive accuracy is measured on the basis of the log likelihood of the observed actions in the testing set. Fig. 4 (b) shows the sum log likelihood of the observed $\mathcal{G}_1$ actions in the testing set as predicted by each model, along with the standard deviation. We observe that even in terms of predictive accuracy, PNE-QE models have better accuracy than Ql0 and Ql1 models with the exception of PNE-QE:$_{S(1)}$, which in fact has the worst predictive accuracy among all models. This means that when used in context of a behavior planner in an AV, if computational constraints inhibits the planner's capacity to sample more than a single baseline trajectory of other agent, the planner is better off selecting Ql0 or Ql1 as a model of behavior than PNE-QE. However, in other cases, PNE-QE is better suited than Ql0 or Ql1. Within the PNE-QE models, whereas the bounds sampling showed better performance in PNE-QE models with regards to precision parameter and AIC, Gaussian sampling shows higher predictive accuracy when maxmin response is used for $\mathcal{G}_2$ games (*p*-value 0.003; Welch test). With maxmax response, the difference between the sampling schemes is non-significant. The rational baselines are shown in braces

in Fig. 4 (b), which are the sum log-likelihood of the actions corresponding to the pure strategy solutions, i.e. the action with highest probability in the mixed strategy.

Overall, these results indicate that based on the three evaluation criteria combined (precision parameter, AIC, and predictive performance), pure strategy Nash equilibria, especially with bounds sampling of trajectories, is still a good model of decision making at the level of maneuvers, but with a noisy response; and this noise can be modelled with a quantal error model that is sensitive to the utility difference to a sample NE.

RQ2. In this research question we study the impact of the state factors on the precision parameter. The state factors are shown in the first column of Table 2. Most state factors are self explanatory; NEXT_CHANGE refers to the next change in the traffic signal and time in seconds till the change occurs, RELEV_VEHICLE refers to the type of relevant vehicle in the game, for example, whether there is a lead vehicle present or other vehicles in conflict which are not lead vehicles. The table shows the mean precision parameter of the behavior models for each state. Since $\lambda_{i,b}$ depends on the state X_i, which is a vector of the six categorical state factors, each row in the table shows the mean precision parameter for situations with the corresponding state factor value, but in isolation; i.e. without taking into account the interaction between the state factors like in the predictive *glm* model. The factors that are found to have the most impact are vehi-

STATE FACTOR	Ql1 MM MX S(1+G)	Ql0 MX MX S(1+G)	PNE_QE MX S(1+B)
SEGMENT			
prep-left-turn	22.4	23.6	**27.7***
exec-left-turn	71.7	72.4	56.8
OTHER LANES	101.8	106.9	111.5
prep-right-turn	**101.1***	95.7	79.0
exec-right-turn	730.6	801.2	385.9
NEXT_CHANGE			
$\geq$ 10sec. to Y/R	34.9	37.0	**37.7***
< 10sec. to Y/R	39.2	**41.7***	40.7
< 10sec. to G	314.6	283.2	263.4
$\geq$ 10sec. to G	204.5	198.5	195.8
SPEED			
LOW SPEED	65.9	68.7	**71.8***
MEDIUM SPEED	**190.6***	177.4	162.0
PEDESTRIAN			
pedestrian present	53.2	54.5	**57.1***
no pedestrian	82.8	76.5	**90.1***
RELEV_VEHICLE			
LEAD VEHICLE	29.3	31.0	**35.4***
OTHER VEHICLES	25.4	20.2	**48.1***
AGGRESSIVE			
Y	114.2	**117.6***	105.2
N	57.2	59.7	**65.0***

Table 2: Mean precision parameter ($\lambda_{i,b}$) of the representative behavior models for each state variable across all games.

$\mathcal{R}(\mathcal{G}_1)$:	MX			MM		
$\mathcal{R}(\mathcal{G}_2)$:	MX	MM	p	MX	MM	p
QL0	+15.3	–	.006	+6.3	–	$\ll 0.05$
QL1	+13.8	–	$\ll 0.05$	–	–	ns.

$\mathcal{R}(\mathcal{G}_1)$:	PNE-QE		
$\mathcal{R}(\mathcal{G}_2)$:	MX	MM	p
	+10.8	–	$\ll 0.05$

Table 3: Impact of response function choice in $\mathcal{G}_2$ on rationality parameters in $\mathcal{G}_1$.

indicated significant within group difference. We see that other than Ql1:MM models, the choice of response function in $\mathcal{G}_2$ is statistically significant. Even though there is an effect of the choice of $\mathcal{G}_2$ response function in the precision parameter values, the difference is not as pronounced differences due to the situational factors.

Conclusion

We formalize the concept of a hierarchical game and develop the various solution concepts that can be applied to a hierarchical game by adapting popular behavioral game theoretic metamodels (QLk and PNE-QE). We evaluated the behavior models based on a large contributed dataset of human driving at a busy urban intersection. Our results show that among the behavior models evaluated, modeling driving behavior as pure strategy NE with quantal errors at the level of maneuvers along with bounds sampling of trajectories provides the best fit to naturalistic driving behavior. Additionally, we observe a significant impact of situational circumstances on the performance of the behavior models, and a moderate impact of the choice of solution concept and sampling strategy at the trajectory level of a hierarchical game. We identify two main directions for future work. The first one evaluate the behavior models with respect to different weights $\mathbf{W}$ of the excitatory and inhibitory motivations, ideally learned from the observational data itself and further addition of social utility norms (Schwarting et al. 2019) and traffic rules in the utility structure. The second one is to perform further analysis of the effect of more situational contexts on the behavior models to inform the choice of correct behavior model under different circumstances.

Ethics Statement

Research in self-driving cars or autonomous vehicles has broad impact on transportation and society in general. Members of the public have a stake in the development of AVs since the algorithms and the processes that go into the development of AVs impact the safety of everyone as road users. The main goal of our paper is to understand human driving behavior in a multi-agent setting in order to make it easier to evaluate how decisions made by AV motion planning algorithms impact other road users. Although the approaches developed in the paper are well suited to be used for the purpose of verification and testing of AV motion planners, there are ethical impacts that should be taken into consideration

cles being on right turn execution segment (exec-right-turn), with and average increase of 165.62 in the precision parameter across all models and when the state of the traffic signal is red/amber, with average increase of 136. Table 2 shows the precision parameter values for the best models (c.f. Fig. 4 (a)) from each metamodel when the corresponding state factor is present. When we contrast the values of the precision parameter in Table 2 to the values in Fig. 4(a), we observe that there is much more variation within individual models depending on agent's state compared to the variation across different models. This means that although PNE-QE models overall has the better performance compared to other models, there is no single model that is best suited across all situations; rather, the choice of the behavior model should be influenced by the agent's situational circumstances. For example, as shown in the table, when agents are observed to have an aggressive trajectory, a Ql0:MX:MX$_{S(1+G)}$ may be a better model *on average* than any other model.

RQ3. As a part of this research question we analyse how much impact the choice of solution concept in lower level game $\mathcal{G}_2$ has on the solution of the higher level game $\mathcal{G}_1$. For the five possible combinations of the metamodel and solution concept in $\mathcal{G}_1$, i.e. Ql0:MX, Ql0:MM, Ql1:MX, Ql1:MM, PNE-QE, Table 3 shows the relative change in the precision parameter of level games $\mathcal{G}_1$ based on the choice of the response function in $\mathcal{G}_2$. Significance is shown based on Dunn's pairwise comparison test after Kruskal-Wallis test

while applying the models.

First, driving behaviors that fall under errors or off-equilibrium behaviors are considered off-equilibrium only with respect to a specific utility structure. Our ability to predict the utility of motivations at an individual level is severely limited, and this limitation needs to be acknowledged and taken into account. For example, the quantitative value an individual driver, for whom their car is a vital commodity for their source of livelihood, assigns to driving safely, (as modelled through surrogate safety metics) may be very different from another individual who owns multiple cars or someone who uses their vehicle only for casual commute. In addition, there are several factors, such as, socioeconomic status, disability, access to insurance, etc., play a role in shaping the driving behavior of an individual.

Secondly, as shown in the paper, off-equilibrium behaviors that we observe in the behavior models can be modelled as an exponential distribution; i.e., the probability of behaviors that lie away from the equilibrium reduces the further the behavior is from on-equilibrium behavior. Since we can estimate the parameters of this distribution, it may be tempting to evaluate models solely through quantitative risk metrics that are derived from this distribution. However, along with such an analysis, there is also a need to be more transparent and investigate the situational context in which the low probability events occur. Due to the same factors mentioned earlier, choosing a certain behavior profile for an AV may adversely impact a segment of road-users, such as older people or people with disability, disproportionately while keeping the overall risk at a population level within a prescribed threshold. Therefore, the use of behavior models in practical AV development needs to be accompanied with not only the information about objective risk metrics but also how the chosen behavior profile impacts vulnerable sections of road users.

References

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

Camara, F.; Romano, R.; Markkula, G.; Madigan, R.; Merat, N.; and Fox, C. 2018. Empirical game theory of pedestrian interaction for autonomous vehicles. In *Proceedings of Measuring Behavior 2018*, 238–244. Manchester Metropolitan University.

Camerer, C. F. 2011. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Crawford, V. P.; Costa-Gomes, M. A.; and Iriberri, N. 2013. Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature* 51(1): 5–62.

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.

Fishburn, P. C. 1970. Utility theory for decision making. Technical report, Research analysis corp McLean VA.

Fridovich-Keil, D.; Ratner, E.; Dragan, A. D.; and Tomlin, C. J. 2019. Efficient iterative linear-quadratic approximations for nonlinear multi-player general-sum differential games. *arXiv preprint arXiv:1909.04694* .

Goeree, J. K.; and Holt, C. A. 2001. Ten little treasures of game theory and ten intuitive contradictions. *American Economic Review* 91(5): 1402–1422.

Ilievski, M.; Sedwards, S.; Gaurav, A.; Balakrishnan, A.; Sarkar, A.; Lee, J.; Bouchard, F.; De Iaco, R.; and Czarnecki, K. 2019. Design Space of Behaviour Planning for Autonomous Driving. *arXiv preprint arXiv:1908.07931* .

Keskinen, E.; Hatakka, M.; Laapotti, S.; Katila, A.; and Peräaho, M. 2004. Driver behaviour as a hierarchical system. *Traffic and Transport Psychology* 9–29.

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

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.

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.

Osborne, M. J.; and Rubinstein, A. 1998. Games with procedurally rational players. *American Economic Review* 834–847.

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. Ann Arbor, MI, USA.

Samuelson, L. 1995. Bounded rationality and game theory. *Quarterly Review of Economics and finance* 36: 17–36.

Schwarting, W.; Alonso-Mora, J.; and Rus, D. 2018. Planning and decision-making for autonomous vehicles. *Annual Review of Control, Robotics, and Autonomous Systems* .

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.

Summala, H. 1988. Risk control is not risk adjustment: The zero-risk theory of driver behaviour and its implications. *Ergonomics* 31(4): 491–506.

Tadelis, S. 2013. *Game theory: an introduction*. Princeton University Press.

Van der Molen, H. H.; and Bötticher, A. M. 1988. A hierarchical risk model for traffic participants. *Ergonomics* 31(4): 537–555.

Wang, M.; Wang, Z.; Talbot, J.; Gerdes, J. C.; and Schwager, M. 2019. Game Theoretic Planning for Self-Driving Cars in Competitive Scenarios. In *Robotics: Science and Systems*.

Wright, J. R.; and Leyton-Brown, K. 2012. Behavioral game theoretic models: a Bayesian framework for parameter analysis. In *AAMAS*, 921–930.

Wright, J. R.; and Leyton-Brown, K. 2014. Level-0 meta-models for predicting human behavior in games. In *Proceedings of the fifteenth ACM conference on Economics and computation*, 857–874.

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.