Revealed Multi-Objective Utility Aggregation in Human Driving

Atrisha Sarkar
University of Toronto
Toronto, Canada
atrisha.sarkar@utoronto.ca

Kate Larson
University of Waterloo
Waterloo, Canada
kate.larson@uwaterloo.ca

Krzysztof Czarnecki
University of Waterloo
Waterloo, Canada
krzysztof.czarnecki@uwaterloo.ca

ABSTRACT

A central design problem in game theoretic analysis is the estimation of the players' utilities. In many real-world interactive situations of human decision making, including human driving, the utilities are multi-objective in nature; therefore, estimating the parameters of aggregation, i.e., mapping of multi-objective utilities to a scalar value, becomes an essential part of game construction. However, estimating this parameter from observational data introduces several challenges due to a host of unobservable factors, including the underlying modality of aggregation and the possibly boundedly rational behaviour model that generated the observation. Based on the concept of rationalisability, we develop algorithms for estimating multi-objective aggregation parameters for two common aggregation methods, weighted and satisficing aggregation, and for both strategic and non-strategic reasoning models. Based on three different datasets, we provide insights into how human drivers aggregate the utilities of safety and progress, as well as the situational dependence of the aggregation process. Additionally, we show that irrespective of the specific solution concept used for solving the games, a data-driven estimation of utility aggregation significantly improves the predictive accuracy of behaviour models with respect to observed human behaviour.

KEYWORDS

Empirical game theory, Bounded rationality, Multi-objective utility

1 INTRODUCTION

In part due to interest in autonomous vehicle (AV) research, recent years have seen a rise in the use of game-theoretic models for modelling human driving behaviour. In this context, the estimation of the utilities of the agents is one of the main steps involved in the design of game-theoretic models based on observational data. However, estimating utilities of players, in this case, human drivers, from purely observational data raises several challenges. First, the problem involves multi-objective utilities; during driving, humans balance different potentially conflicting objectives, such as safety, progress, and comfort, in the process of selecting their desired action. The process of aggregation, defined as the mapping of multi-objective utilities into a scalar value [28], is often context dependent and individual specific. Second, the underlying human

reasoning model is often diverse and cannot be modelled through a specific notion of an equilibrium behaviour in all situations [31, 34]. Additionally, there may also be aspects of bounded rationality in play [32]. One way to estimate utilities from observational data is by using the concept of *rationalisability*, that is, identifying the aggregation parameters that would make the observed agent decision optimal with respect to a reasoning model. However, this is still challenging since not only are there different modalities of aggregation, but also the definition of optimality depends on the various strategic and non-strategic reasoning models involved. When there are so many unobservable factors, the problem is inherently under-specified, meaning that the same observation can be explained by different combinations of reasoning model and utility aggregation. In this paper, we address the problem of estimating the parameters of multi-objective utility aggregation from observational data by developing different methods for popular classes of utility aggregation, namely weighted and satisficing aggregation, and different underlying reasoning models that include both strategic and non-strategic reasoning.

Analysing the aggregation process helps us gain a basic understanding of drivers' preferences under different driving situations, and answering questions such as observed association between the situational state and the utility aggregation. For example, if a driver is observed to have a higher than expected speed close to an intersection, can we infer something about their aggregation parameter based on that observation — maybe that a driver in that context may weigh progress more than safety? From a modelling point of view, we obtain a more accurate identification of agent utilities, which is especially relevant for behavioural models, since in the absence of such an analysis, any behavioural model, no matter how incorrect the utilities are, can explain away deviations of observed behaviour under boundedly rational behaviour. The line of inquiry of estimating some aspect of agent utilities from observations is related to the problem of inverse reinforcement learning [20, 29], inverse game theory [21], theory of revealed preference [10], and multi-criteria decision making in operations research [13, 19].

In the context of estimating utilities, it is important to note that there are two separate questions. First, the question of form, i.e., estimating the form and parameters of the utility function, say $u_{safety}(\delta) : \mathbf{R}^+ \to [-1, 1]$ that maps the choices (e.g., the distance gap δ) into a utility interval [-1,1]. This has been well studied within the literature of revealed preference with specific behavioural theories such as time discounting of utilities [14], risk aversion [15], and prospect theory of loss aversion [37], some of which have been applied to driving [32] and robotics [22]. However, a second question around multi-objective utilities, i.e. how to estimate aggregation parameters of multi-objective utilities based on consistency of observations and reasoning model, has received comparatively less

attention in the context of AV or human driving— and that is the focus of this work. Specifically, we address the following questions.

- **Aggregation**: Given a multi-objective utility U and a parametric scalarization function $S(U, \theta)$, how do we estimate θ that is *rationalisable* with a set of observed choices of all agents conditioned on a model of reasoning?
- **Bounded rationality**: How can the estimation of θ accommodate nonstrategic reasoning models such as maxmax or maxmin?
- **State association**: Is there an observed association between state factors such as velocity, traffic situation, etc., and θ? In other words, are the parameter values stable across different traffic situations?
- **Model performance**: How does the performance of different behaviour models change when utilities U are constructed based on a learning-based technique that infers the aggregation parameter θ from the data?

The question of aggregation is addressed by constructing axiomatic conditions under which a set of observations is *rationalisable* using a given parameterized aggregation method, namely weighted aggregation and satisficing aggregation. As a part of the second question, we show that such a construction is different for strategic and nonstrategic models, where the former can be formulated by a set of linear constraints and the latter as a set of nonlinear constraints. For the third question, we estimate the rationalisable parameters for different traffic situations and evaluate whether there are significant situational differences in how drivers aggregate the utilities, namely safety and progress. Finally, by treating the state factors as independent variables and the aggregation parameter as dependent, we use the data to learn a regression model (CART) that can predict the aggregation parameter in new situations, and use that method to evaluate the performance of the popular classes of behaviour models proposed in the literature for driving.

2 RELATED WORK

This section spans three different fields of research that deal with similar problems in their own right. Namely, literature on the theory and applications of reveled preference from economics, multicriteria decision making from operations research, and inverse reinforcement learning from robotics and computer science.

Theory of revealed preferences: The main problem addressed in this paper, i.e., estimation of agent preferences given a set of observations and a model, falls under the scope of the theory of revealed preferences. The literature on connection between revealed preference and utilities has built upon Afriats's approach [1], which defines axioms of existence of a utility function u that can *rationalise* a set of observed behaviours. Although most of the literature is focused on aggregate consumer demand problems [12, 36], revealed preference conditions can also be constructed for noncooperative strategic models such as Nash equilibrium [8]; and Chambers et al. [7] lay the universality and existence conditions of such a construction for any model beyond just equilibrium. Covering the extensive literature on revealed preference in economics is outside of the focus and scope of this work; therefore, we refer to [12] as a good reference for that general literature.

Most economics models are based on rational choices, and given that this work builds models that include non-strategic behaviour (boundedly rational agents), it is relevant to include literature on revealed preference that is based on behavioural economics. Crawford [10] presents a review of the literature on revealed preference that covers behavioural theories and links to empirical evidence. Dziewulski [14] constructs the revealed preference conditions based on a model in which a single agent uses time-based discounting of their utilities with various discounting models such as quasi-hyperbolic and exponential. In contrast, this paper uses a model that is simpler in some way (one-shot game as opposed to dynamic game) and more complex in other way (multi-agent behaviour). Application of the construction from [14] for the case of driving in dynamic semi-cooperative setting is an interesting future direction of research, especially since discounted utilities are standard in reinforcement learning (RL) based methods, and RL has received a lot of attention from the AV community in recent years. Another behavioural attribute, altruism, has been consistently observed in an empirical setting, especially in the context of dictator games [3]. Andreoni and Miller [2] set the construction of the revealed preference with respect to altruistic behaviour in a dictator game and find that only a quarter of the participants were selfish money maximisers and the rest passed the test of altruistic behaviour. More recently, Porter and Adams [26] study revealed preference with respect to altruistic behaviour in the context of intergenerational wealth transfer, that is, transfer of money from an adult child to ageing parents. The study in [26] varies different models of utility, from pure selfish behaviour to pure altruism, and finds that although more than 90% of the participants pass the test of revealed preference (i.e., behaviour consistent with the models and utilities), there were differences observed based on whom they were playing the game against, whether parents or strangers. Similarly to [26], we vary the utility construction (different models of aggregation), construct the revealed preference conditions, and test on empirical data ([26] is based on a laboratory experiment) to evaluate what proportion of behaviour passes those conditions. However, the models and applications in this paper are, of course, quite different. Overall, although the above works have treated different behavioural attributes with respect to theory of revealed preference well, to our understanding there is no existing work on revealed preference that is based on multi-objective utilities and non-strategic reasoning models in the context of driving behaviour.

Multi-criteria decision analysis (MCDA) Another strand of literature that is related to this paper is on multi-criteria decision making from operations research [13]. Compared to the theory of revealed preference, where the focus is more on the model of decision making, in MCDA, the focus is on multiobjective nature of the utilities. The process of estimating the parameters of the aggregation process that an agent uses is called preference disaggregation (a terminology we retain in the paper), and Jacquet-Lagreze and Siskos [19] provide a review of the tools and techniques for that purpose until the year 2000. From a set of datapoints of ranked choices made by an individual, typical algorithms solve the general minimisation problem $\arg\min_{\mathbf{w}} ||\mathcal{R}(X)^o, \mathcal{R}(X, \mathcal{A}_{\mathbf{w}})||$, where $\mathcal{R}(X)^o$ is the observed ranking of the alternatives by the agent, $\mathcal{R}(X, \mathcal{A}_{\mathbf{w}})$ is the ranking based on the aggregation model $\mathcal{A}$ parameterized on

w on the same set of alternatives S. Standard algorithms, such as UTA [18] formulate the solution as a mathematical programming problem, and in recent years, statistical learning methods similar to those we use (CART) have also been used [13]. There are few differences between the MCDA methods and those in this paper. First, in our case, the models we study are strategic (and non-strategic) decisions, thereby adding another layer of complexity. Second, our problem in this paper is also less well-defined, since we do not have access to the drivers' ranking of the preferences, but rather only a singular choice of the observed action.

Inverse reinforcement learning (IRL): Although IRL [25] is conceptually different from the methods presented in this paper, it is relevant to include some recent works in the literature due to the interest and application of IRL for autonomous driving. IRL formulates the problem of estimating agents' behaviour as a single agent problem as opposed to the game theoretic approach of treating the problem as one of multi-agent behaviour with support for different reasoning processes. Another salient distinction in IRL is that it typically retrieves the utility U that fits the observed behaviour best without referencing utility to prespecified dimensions of safety, progress, comfort, etc., but rather uses a single objective function that may or may not have a semantic meaning. Sadigh et al. [29] use IRL to first learn a policy of behaviour from demonstrations and subsequently use that in a game theoretic based planning module using a level-k (k = 2) [9] type solution concept, although the solution concept is not explicitly stated as such in the paper. As a mathematical formulation, such an approach works well in practise because IRL can provide a best-response type behaviour to the (other) agent action; however, the implicit assumption that the agents adhere to a single model of reasoning throughout every interaction might be a strong one. Nevertheless, the authors show practical ways to integrate single agent method like IRL into a game-theoretic setting.

A recent work on learning preference along multiple criteria with a game theoretic view and also in the context of driving is by Bhatia et al. [4], where agent utilities are learnt with respect to a solution concept developed based on the Blackwell approachability theorem [5]. Compared to [4], in this work, we use non-zero sum games and pure strategies in terms of the game constructs, as well as focus on multiple solution concepts. Additionally, in this paper, we also learn the preferences of drivers based on real-world observational data.

3 AGGREGATION

The general problem of aggregation for an agent is the transformation of a vector valued utility function U_i to a scalar valued function u_i in order to solve the game in question. In other words, this involves the construction of a scalarization function $S(U_i(a_i, a_{-i}), \theta_i)$ that maps the multiobjective vector of utilities for agent i, $U_i(a_i, a_{-i})$, to the real value utility $u_i(a_i, a_{-i})$ based on the parameter θ_i.

3.1 Weighted Aggregation

Weighted aggregation is a linear combination of individual utility objectives as follows.

$$S(U_i(a_i, a_{-i}), \mathbf{w}_i) = \mathbf{w}_i \cdot U_i(a_i, a_{-i}) \tag{1}$$

The above equation is simply the dot product between the aggregation parameter ($\mathbf{w}_i$) and the vector valued utility function. The disaggreagation process involves the estimation of the weight vector $\mathbf{w}_i$ based on the observed actions of the agents in the game.

3.2 Satisficing Aggregation

A driver always operating at their own subjective tolerance level of risk has been a well established model of behaviour in traffic psychology [17, 38], and the model has also been empirically validated [23]. Lexicographic thresholding is a method of aggregation that is based on *satisficing* and encapsulates two concepts, namely, ordered criteria of objectives and a thresholding effect [24]. In lexicographic thresholding, an agent ranks the objective criteria based on a fixed and strict total order, for example, safety > progress > comfort. In this work, we focus on safety and progress with a lexicographic ordering of safety > progress. The aggregation of the two utilities into a scalar value is given by

$$S(U_i(a_i, a_{-i}), \gamma_i) = \begin{cases} u_{s,i}(a_i, a_{-i}), & \text{if } u_{s,i}(a_i, a_{-i}) \leq \gamma_i \\ u_{p,i}(a_i, a_-), & \text{otherwise} \end{cases} \tag{2}$$

where γ_i is the *safety aspiration level* of agent i; $u_{s,i}(a_i, a_i)$ is the safety component of the vector valued function U_i; and $u_{p,i}(a_i, a_i)$ is the progress component. Based on the above formulation, an agent evaluates an action of multivalued utility based on progress rather than safety only when the safety utility of that action is greater than γ_i. The disaggregation process for the lexicographic thresholding method involves estimating the parameter γ_i based on the observed action of the agent i in the game.

4 MULTIOBJECTIVE DISAGGREGATION

We address the problem of estimating the aggregation parameters for an agent given their observed action in a game. This involves estimating the weight parameters **w** for the case of the weighted aggregation method, and the safety aspiration level parameter γ for the case of the satisficing aggregation method. Additionally, since the choice of reasoning model (strategic or non-strategic) influences the behaviour of the agent in a game, we will develop separate methods based on strategic and non-strategic reasoning assumptions.

We start with the following definition of what *rationalisability* means in the context of disaggregation of multiobjective utilities.

DEFINITION 1. *Given a normal form game $\mathcal{G}$, a vector-valued utility U_i, a solution concept $\mathcal{B}$, and a tuple of observed action (a_i^o, a_{-i}^o), an aggregation parameter θ_i is rationalisable iff (a_i^o, a_{-i}^o) is in the solution set of $\mathcal{G}$ solved with the solution concept $\mathcal{B}$ with scalarized utility $S(U_i, \theta_i)$.*

Based on a dataset of observations (a_i^o, a_{-i}^o) in various game situations, the goal is to estimate the rationalisable θ_i for each agent in each game situation. We first do this for the case where $\mathcal{S}$ is the weighted aggregation function, followed by the case where it is the satisficing aggregation function.

4.1 Weighted Aggregation

Since the observed actions depend on the underlying reasoning models used by the player, we first present the methods for strategic
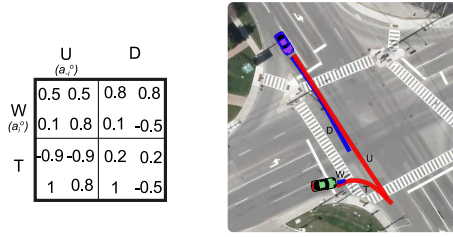
Figure 1: An example scenario with actions *Turn* (T) and *Wait* (W) (right-turn) and *Speed up* (U) or *Slow Down* (D) (oncoming). Each ell has safety and progress utility indicated.

reasoning models (taking Nash equilibrium as an example), followed by non-strategic reasoning models of maxmax and maxmin [40].

4.1.1 Strategic models. For strategic models, for the observed action, a_i^o, of a strategic agent i to be in the solution set, the action needs to be the best response to the action that i believes $-i$ will play. We use the case of Nash equilibrium in this section, where a_i^o and a_{-i}^o are best responses to each other. A running example of a right-turning scenario (Fig. 1) elaborates the estimation process. A prototypical game for the scenario is shown on the right. For each combination of pure strategies, the top row utility values in each cell represent the safety utility and the bottom row represent the progress utility of the right turning vehicle (row player) and straight through (column player) vehicle, respectively. Let's say the observed strategy in this game was (W, U). In that case, for the right turning vehicle, for W to be the best response to the observed action of $-i$, (i.e., U), the necessary and sufficiency conditions are $w_{i,s} \times 0.5 + w_{i,p} \times 0.1 \geq w_{i,s} \times -0.9 + w_{i,p} \times 1$ and $w_{i,s} + w_{i,p} = 1$, where $w_{i,s}$ and $w_{i,p}$ are agent i's weights for safety and progress utilities respectively. In the general case of arbitrary number of finite discrete actions and $|O|$ being the number of objectives, this can be formulated as a linear program (LP) and the rationalisable weights of agent i can be estimated as the solution to the following

$$\text{maxmize} \sum_{j=1}^{|O|} w_{i,j} u_{i,j}(a_i^o, a_{-i}^o)$$

$$\text{subject to} \sum_{j=1}^{|O|} w_{i,j}(u_{i,j}(a_i^o, a_{-i}^o) - u_{i,j}(a_i^{'}, a_{-i}^o)) \geq 0, \quad \forall a_i^{'} \neq a_i^o$$

$$\sum_{j=1}^{|O|} w_{i,j} = 1$$

In the above LP, we select weights in the feasible set that maximise the utility of the chosen action; however, any combination of weights that fall into the feasible set based on the constraints would be consistent with the conditions of rationalisability.

4.1.2 Non-strategic models. In the case of a non-strategic model, an agent is not *other responsive* but only *dominant responsive* [40]. In other words, since they do not reason about the actions of the other agents, it is not possible to pin down a specific action of the other agent (a_{-i}^o) with respect to which agent i calculates its best response. In this model, the players do not best respond based on a specific belief about other agents' actions, but rather evaluate actions based only on their own utility values, and in our case, choose

an action based on an elementary maxmax or maxmin model. This makes the process of estimating the weights slightly more complicated (read nonlinear) compared to the strategic case. Following from the example of Fig. 1, for action W to be the optimal action for agent i (based on the non-strategic model maxmax), the maximum utility for agent i that can be realised by choosing W in the aggregate form post scalarization should be greater or equal to the maximum utility that can be realized by choosing T. Therefore, the necessary and sufficiency conditions for the non-strategic case in this example are $\max\{w_{i,s} \times 0.5 + w_{i,p} \times 0.1, w_{i,s} \times 0.8 + w_{i,p} \times 0.1\} \geq \max\{w_{i,s} \times -0.9 + w_{i,p} \times 1, w_{i,s} \times 0.2 + w_{i,p} \times -0.5\}$ and $w_{i,s} + w_{i,p} = 1$. The left term in the inequality gives the maximum realised utility for the action W and the right term is the maximum realised utility for the action T. The process of estimating the weights in the non-strategic case can therefore be formulated as a nonlinear optimisation problem as follows.

$$\text{maxmize} \sum_{j=1}^{|O|} w_{i,j} u_{i,j}(a_i^o, \arg\max_{a_{-i}} w_{i,j} u_{i,j}(a_i^o, a_{-i}))$$

$$\text{subject to} \sum_{j=1}^{|O|} w_{i,j}(u_{i,j}(a_i^o, \arg\max_{a_{-i}} w_{i,j} u_{i,j}(a_i^o, a_{-i}))$$

$$- u_{i,j}(a_i^{'}, \arg\max_{a_{-i}} w_{i,j} u_{i,j}(a_i^{'}, a_{-i}))) \geq 0, \quad \forall a_i^{'} \neq a_i^o$$

$$\sum_{j=1}^{|O|} w_{i,j} = 1$$

Due to the presence of the argmax operator, the above problem changes to a nonlinear optimisation problem. Similarly, for maxmin non-strategic models, the process of estimating the weights is identical except that the argmax operator is replaced by the argmin operator. In the latter case, the argmin operator gives the minimum realisable utility of the observed action. In our experiments, for both maxmax and maxmin models, we solve the above optimisation problem using a trust region based method [11].

4.2 Satisficing Disaggregation

The estimation process for the satisficing method involves estimating the parameter γ_i based on the observed action of the agent i in the game. Similar to the weighted aggregation case, the method of estimation depends on the underlying model due to the assumption an agent has over other agents' behaviour and the subsequent impact on the optimality calculations based on that agent's perspective. However, unlike in the weighted aggregation case, due to the thresholding effect, it is not straightforward to construct a functional form of the scalar utility over which the optimality conditions can be built. Instead, we develop an algorithmic estimation process that helps estimate the complete set of rationalisable aggregation parameter γ.

4.2.1 Strategic model. Based on Eqn. 2, the aggregation process for lexicographic thresholding can be expressed in a parametric form as $u_i(a_i, a_{-i}) = \mathcal{S}(U_i(a_i, a_{-i}), \gamma_i)$ where $U_i(a_i, a_{-i}) = [u_{i,s}(a_i, a_{-i}), u_{i,p}(a_i, a_{-i})]$ and $\mathcal{S}$ is the scalarization function of Eqn. 2. We present an adapted definition of rationalisability for strategic models as follows:

Algorithm 1: Estimation of Γ_i based on consistency with respect to satisficing aggregation

---

Input : (a_i^o, a_{-i}^o)

1   $P \leftarrow$ partition($[-1, 1]$, $<$);

2   $\Gamma_i \leftarrow \{\varnothing\}$;

3   **for** $I \in P$ **do**

4      $\gamma \leftarrow$ sample(I);

5   **if** *is_rationalisable(γ)* **then**

6         $\Gamma_i \leftarrow \Gamma_i \cup I$

7   **end**

8   **end**

DEFINITION 2. *For any agent i, a safety aspiration level $\gamma_i \in [-1, 1]$ is equilibrium rationalisable with strategy profile (a_i^o, a_{-i}^o) iff*
$$\mathcal{S}(U_i(a_i^o, a_{-i}^o), \gamma_i) \geqslant \mathcal{S}(U_i(a_i', a_{-i}^o), \gamma_i) \; \forall a_i' \neq a_i^o$$

The above definition follows from the definition 1 with an explicit reference to the condition of optimality of the equilibrium solution, that is, for the safety aspiration level of the agent i to be rationalisable, their observed action a_i^o must be the best response to the action of the other agents a_{-i}^o. Algo. 1 presents the general algorithm to estimate the rationalisable parameter. The intuition behind the algorithm is as follows: the value of the parameter γ_i lies within the utility interval [-1,1]. Let $P = \{I_1, I_2, ..., I_P\}$ be an ordered partition of the interval [-1,1]; the process of constructing the partition depends on the underlying models of reasoning and is explained later. We sample a single value of $\gamma \in I$ and check if the scalarization $\mathcal{S}$ based on that sampled value is rationalisable with respect to the definition 2. If so, we include the partition I from which γ was sampled in the set of rationalisable parameter set Γ, and the union of these sets is the set of rationalisable γ. Next, we discuss the conditions for sound and completeness.

PROPOSITION 1. *Algorithm 1 is sound and complete based on a partition P iff $\forall I \in P$, is_rationalisable(γ) $\leftrightarrow$ is_rationalisable(γ')$\forall \gamma, \gamma' \in I$*

Implementing the *is_rationalisable* method based on the definition 2 ensures soundness; this is because the utility maximizing action (condition of definition 2) in response to an equilibrium action means that the said action is in equilibrium, and therefore (correctly) rationalisable. The bidirectional implication condition of proposition 1 ensures that if we sample only a single value γ from an interval $I \in P$ and check for rationalisability, then any γ' in that interval that was not sampled is also rationalisable. Next, we construct a partition for which the double-implication condition of proposition 1 holds. Given a game, let the partition P_{eq} consist of the ordered safety utility of the agent i's action as follows

$$P_{eq} = \{[-1, u_{i,s}(a_{i,1}, a_{-i}^o)), [u_{i,s}(a_{i,1}, a_{-i}^o), ...), [u_{i,s}(a_{i,|A_i|}, a_{-i}^o), 1]\}$$

where $u_{i,s}(a_{i,1}, a_{-i}^o) \leqslant ... \leqslant u_{i,s}(a_{i,|A_i|}, a_{-i}^o)$ is the ordered sequence of the safety utility values of agent i. An example partition for the game with respect to the row player (right turning vehicle) in response to the action U (the observed action) of the column player of Fig. 1 is shown in Fig. 2.

Figure 2: The partition intervals $I_0, I_1, ..$ of P (game in Fig. 1).

THEOREM 1. *For any interval $I \in P_{eq}$, if $\gamma \in I$ is equilibrium rationalisable, then $\forall \gamma' \in I$, γ' is equilibrium rationalisable. Conversely, if $\gamma \in I$ is not equilibrium rationalisable, then $\forall \gamma' \in I$, γ' is not equilibrium rationalisable.*

The proof is based on the intuition that if the partitions are constructed using ordered safety values of different actions, then any given threshold that falls between two such utilities imposes the same ordering of actions after scalarization since the conditions of Eqn. 2 remain unchanged.

PROOF. Since the equilibrium rationalisability is based on the condition $\mathcal{S}(U_i(a_i^o, a_{-i}^o), \gamma_i) \geqslant \mathcal{S}(U_i(a_i', a_{-i}^o), \gamma_i)$, we first show that $\mathcal{S}(U_i(a_i, a_{-i}), \gamma_i) = \mathcal{S}(U_i(a_i, a_{-i}), \gamma_i') \; \forall \gamma, \gamma' \in I$, i.e., the scalarized value based on any two parameters that fall in the same interval is equal.

Case $u_{s,i}(a_i, a_{-i}^o) \leqslant \min I$: In this case, $u_{s,i}(a_i, a_{-i}^o) \leqslant \gamma, \forall \gamma \in I$. Therefore, $\mathcal{S}(U_i(a_i, a_{-i}), \gamma_i) = \mathcal{S}(U_i(a_i, a_{-i}), \gamma_i')$ since both evaluate to $u_{s,i}(a_i, a_{-i}^o)$ based on Eqn. 2.

Case $u_{s,i}(a_i, a_{-i}^o) > \min I$: In this case, $u_{s,i}(a_i, a_{-i}^o) \geqslant \sup I$, since for any $u_{s,i}(a_i, a_{-i}^o) \neq 1$, $u_{s,i}(a_i, a_{-i}^o) = \min I$ when $u_{s,i}(a_i, a_{-i}^o) \in I$ based on the construction of P_{eq}. Therefore, $\forall \gamma \in I$, $\gamma < u_{s,i}(a_i, a_{-i}^o)$, and $\mathcal{S}(U_i(a_i, a_{-i}), \gamma_i) = \mathcal{S}(U_i(a_i, a_{-i}), \gamma_i')$ since both evaluates to $u_{p,i}(a_i, a_{-i}^o)$ based on Eqn. 2

Therefore for any $I \in P_{eq}$, the condition $\mathcal{S}(U_i(a_i, a_{-i}), \gamma_i) = \mathcal{S}(U_i(a_i, a_{-i}), \gamma_i')$ holds true for all $\gamma, \gamma' \in I$.

By the above equality condition, $\mathcal{S}(U_i(a_i^o, a_{-i}^o), \gamma_i) \geqslant \mathcal{S}(U_i(a_i', a_{-i}^o), \gamma_i) \leftrightarrow \mathcal{S}(U_i(a_i^o, a_{-i}^o), \gamma_i') \geqslant \mathcal{S}(U_i(a_i', a_{-i}^o), \gamma_i')$, which establishes the biconditional relationship of the theorem (Defn. 2). □

Theorem 1 helps significantly reduce the number of consistency checks that we need to perform, since we need to check only one value in each interval in P to determine whether all the values in that interval are rationalisable or not. This keeps the run-time complexity of Algo. 1 linear in the number of actions of the agent in the worst case (that is, $O(|A_i|)$), since the run time depends on the size of the partition P_{eq}, which in turn depends on the number of unique safety utilities, i.e., $|A_i|$ in the worst case.

4.2.2 Non-strategic models. Recall that for non-strategic models, an agent i does not hold a specific belief about the action another agent might play, and therefore, similar to the weighted aggregation case, we cannot pin down a specific action a_{-i}^o in response to which the parameters can be estimated. This leads to a revision of the rationalisability definition of Def. 2 to make it independent of the actions of other agents for the maxmax and maxmin models.

DEFINITION 3. *For any agent i, a safety aspiration level $\gamma_i \in [-1, 1]$ is* maxmax *rationalisable with action a_i^o iff*
$$\max_{a_{-i}} \mathcal{S}(U_i(a_i^o, a_{-i}), \gamma_i) \geqslant \max_{a_{-i}} \mathcal{S}(U_i(a_i', a_{-i}), \gamma_i) \forall a_i' \neq a_i^o.$$

DEFINITION 4. *For any agent i, a safety aspiration level $\gamma_i \in [-1, 1]$ is maxmin rationalisable with action a_i^o iff $\min_{a_{-i}} \mathcal{S}(U_i(a_i^o, a_{-i}), \gamma_i) \geqslant \min_{a_{-i}} \mathcal{S}(U_i(a_i', a_{-i}), \gamma_i) \forall a_i' \neq a_i^o$.*

For the strategic case, we needed to check a_i^o for rationalisability only as a response to a fixed action a_{-i}^o, and therefore it was sufficient to construct the partition P_{eq} based only on the safety utilities for all actions that were in response to a_{-i}^o. However, for non-strategic models, the rationalisability of a_i^o involves comparison with all entries of the safety utilities of agent i in the game matrix. Therefore, to apply Prop. 1 for the non-strategic case, the partition points of P need to include all the entries of the table as follows:

$$P_{ns} = \{[-1, u_{i,s}(a_{i,1}, a_{-i})), [u_{i,s}(a_{i,2}, a_{-i}), u_{i,s}(a_{i,k}, a_{-i})),$$
$$..., [u_{i,s}(a_{i,|A_i|}, a_{-i}), 1]\}$$

where $u_{i,s}(a_{i,1}, a_{-i}) \leqslant u_{i,s}(a_{i,2}, a_{-i}) \leqslant ... \leqslant u_{i,s}(a_{i,|A_i|}, a_{-i})$ is the ordered sequence of the safety utility values of agent i, and (with a minor abuse of notation) a_{-i} steps through all the corresponding actions of the other agents based on that ordering. The only difference between P_{eq} and P_{ns} is that P_{ns} is partitioned based on the safety utilities of i in the entire game matrix, whereas P_{eq} was based on the column corresponding to a_{-i}^o. This also has an impact on the runtime of the algorithm, which is $O(|A|^N)$, where N is the number of players in the game and $|A|$ is the number of actions for a player. This value is the same as the size of the game matrix since the partition is constructed from each safety utility value for each agent. The corresponding corollaries of Theorem 1 for the non-strategic case are as follows:

COROLLARY 1. *For any interval $I \in P_{ns}$, if $\gamma \in I$ is maxmax rationalisable, then $\forall \gamma' \in I$, γ' is maxmax rationalisable. Conversely, if $\gamma \in I$ is not maxmax rationalisable, then $\forall \gamma' \in I$, γ' is not maxmax rationalisable.*

COROLLARY 2. *For any interval $I \in P_{ns}$, if $\gamma \in I$ is maxmin rationalisable, then $\forall \gamma' \in I$, γ' is maxmin rationalisable. Conversely, if $\gamma \in I$ is not maxmin rationalisable, then $\forall \gamma' \in I$, γ' is not maxmin rationalisable.*

PROOF. The proof of the above corollaries is similar to the proof of Theorem 1. Observe that for the partition set P_{ns}, for any agent i, the aggregation of the utilities of i is the same for any pair of $\gamma, \gamma' \in I$. This follows from the equality condition $\mathcal{S}(U_i(a_i, a_{-i}), \gamma_i) = \mathcal{S}(U_i(a_i, a_{-i}), \gamma_i')$, which holds for the partition P_{ns} in the same way as was established for P_{eq} earlier in Theorem 1. This means that the pairwise comparison between the utilities of actions of i is invariant to the value of $\gamma \in I$, thus establishing the conditions of definitions 3 and 4. □

5 EXPERIMENTS AND EVALUATION

We evaluate the utility aggregation estimation methods using a multi-agent traffic drone data set from [30], which includes several real-world traffic interactions. From the dataset, we include interaction scenarios of unprotected right-turn and left-turn across path at a busy four way signalized intersection (1667 2-player games), decisions about entering a roundabout (2441 N-player games), and decisions about waiting for a pedestrian at a crosswalk (288 N-player games). Each of these games are constructed from the perspective

of a *focal* agent about to enter the intersection, roundabout, and crosswalk, respectively. The intersection and roundabout games contain the interaction of the focal agent (a vehicle) with other vehicles, and the crosswalk scenario include games in which the other players are pedestrians. The relevant code is available at git.uwaterloo.ca/a9sarkar/single-shot-hierarchical-games. As part of our analysis, first, we evaluate the proportion of games in which a rationalizable parameter was found for each aggregation and reasoning model combination. Second, we evaluate the improvement in accuracy of player's predicted actions when the utilities are constructed using the proposed estimation methods.

5.1 Agent Utility Aggregation Parameters

The first point of analysis is the pass rate for each model, that is, the percentage of games in which a rationalisable parameter was found for each model and aggregation method combination (Table 1). Based on the estimation procedure, we see that the chosen action of drivers can be rationalised by a weighted aggregation parameter in all cases for all models. For satisficing aggregation, the pass rate is sensitive to the specific traffic situation and the choice of the reasoning model. In the roundabout and crosswalk scenarios, rationalisable parameters for strategic models could be estimated for almost all games (96.9% and 100%, respectively), whereas for non-strategic models it could only be found for 56.8% to 66.9% of the games depending on the specific solution concept. There are two possible reasons why weighted aggregation parameters show higher rationalisability. First, at least for non-strategic models, the optimisation method involves an approximate procedure (in the form of the use of trust region based method of [11]), which ends up finding a solution, albeit approximate, more easily than the corresponding exact estimation procedure for satisficing based methods. Second, it might be possible that a weighted aggregation methodology as opposed to a satisficing based procedure makes the chosen action more optimal with respect to rationalisability constraints. The actual parameter values of the weights shed more light on this aspect, which is discussed next.

5.1.1 Parameter values. Next, we study the values of the rationalisable parameters that were estimated under each model. Figures 3 and 4 show the violin plots of the safety weight (safety weight of **w**) and safety aspiration level (γ) for weighted and satisficing aggregation methods, respectively, for the focal player in roundabout and crosswalk games, and both players in intersection games. For cases where rationalisable parameters are an interval, for the purpose of the figure we select the center of the interval as the representative sample. To study the association between the state factors, the figures are stratified based on the velocity of the agent as well as the scenario. The first observation about the weighted aggregation method is that irrespective of the reasoning model, the distributions of the weight parameters are multimodal with the modes being concentrated towards lower and higher values in most cases. This means that in most cases, with safety and progress as the two objectives, drivers tend to weigh heavily one or the other rather than weighing both together in some mixed proportion at the same time. Intuitively, this makes sense because what the revealed preference estimation *rationalises* are the weights of the two objectives that would make the chosen action of the driver optimal. For example, in a given game, if the driver chose to proceed, then

		Intersection		Roundabout		Crosswalk	
		Weighted	Satisficing	Weighted	Satisficing	Weighted	Satisficing
Strategic	Nash	100%	68%	100%	96.9%	100%	100%
Non strategic	maxmax	100%	72.2%	100%	56.8%	100%	66.9%
	maxmin	100%	72.28%	100%	56.87%	100%	64.3%

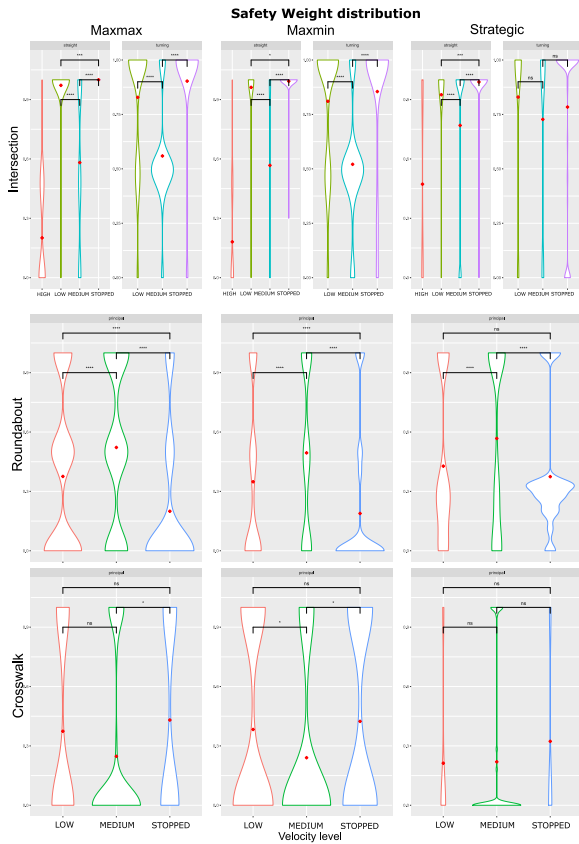Table 1: Pass rate of estimated preferences for each model, aggregation method, and dataset.

Figure 3: Safety weight parameter distribution stratified by vehicle speed, scenario, and task. Significance levels $p \leqslant 0.05(*), \leqslant 0.01(**), \leqslant 0.001(***), \leqslant 0.0001(****)$, and *ns.*

Figure 4: Safety aspiration level distribution stratified by vehicle speed, scenario, and task. Significance levels $p \leqslant 0.05(*), \leqslant 0.01(**), \leqslant 0.001(***), \leqslant 0.0001(****)$, and *ns.*

what we find is that evaluating that action only with respect to its, say, progress utility, makes it more optimal than if the driver had evaluated that action based on both safety and progress utility. On the other hand, for the satisficing method of aggregation, the distribution of γ is not multimodal in most cases. Rather, the mean values (shown in red) are concentrated near 0, thereby indicating more stability across reasoning model or the game situation.

5.1.2 Subgroup analysis. Next, we study the association between the vehicle speed and the estimated parameters values. The median values of the parameters are shown in red within the violin plots. Within each dataset and scenario, we perform subgroup analysis based on discretised velocity levels [1], and significance

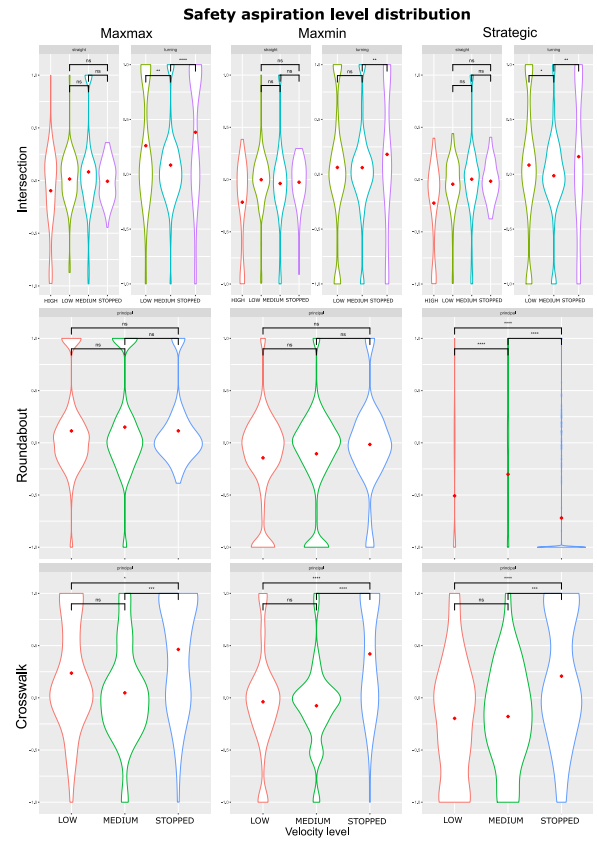[1] Some velocity levels were excluded due to not enough instances in the data

between groups is noted above the horizontal lines according to Wilcoxon *t-test* at significance levels $p \leqslant 0.05(*), p \leqslant 0.01(**), p \leqslant 0.001(***), p \leqslant 0.0001(****)$ (null hypothesis of equal means between groups). In general, significant differences in parameter values with respect to velocity levels were found for 72% and 42% of pairwise group comparisons for weighted and satisficing aggregation, respectively, where a group is a combination of scenario and model. This shows that the safety aspiration levels of drivers show more stability at different velocity levels compared to the weight parameters. Additionally, within each scenario and reasoning model, for the cases where there is a significant difference, higher velocities are associated with lower safety weights and lower safety aspiration level for weighted and satisficing aggregation, respectively.

5.2 Predictive Accuracy Improvements

Since subgroup analysis shows an association is observed between factors such as speed and task with the values of the aggregation parameter, in this section, we evaluate whether the aggregation methods in conjunction with a statistical learning method, such as Classification and Regression Trees (CART) [6], can improve the predictive accuracy of various behaviour models and solution concepts with respect to naturalistic human behaviour. We first split the dataset of games into training and testing set in 80:20 proportion with $K = 30$ random subsampling, and use the training set to construct two separate regression trees, one each for weighted and satisficing aggregation, respectively. The independent variables in the models are driving task (left-turn, right-turn, straight-through, etc.), scenario (intersection, crosswalk, roundabout), reasoning model, and velocity, whereas the aggregation parameter estimated using the methods developed earlier is the dependent variable. We use the CART model to predict the aggregation parameters in the games in the testing set, and solve the games using different solution concepts. We then evaluate the improvement in accuracy of prediction of each solution concept when the games are constructed with multi-objective utilities aggregated using a fixed weights of $\mathbf{w} = [0.5, 0.5]$ compared to our method. We select solution concepts from existing literature that have been proposed as a model for human driving or autonomous vehicle behaviour. The selected solution concepts are as follows: Level-k model [30, 35], where $k = \{0, 1, 2\}$ indexed as L0, L1, and L2, respectively. L0 models are further separated based on the non-strategic solution concept into maxmax (L0:MX) and maxmin models (L0:MM) [39], pure strategy Nash equilibrium (PNE) with welfare maximizing solution selected in cases of multiple equilibria [27, 30, 33], Stackelberg solution for 2-player scenarios (Stack.) [16, 34], Rule based solution that says always wait for other vehicles or pedestrians, and a level-k, $k = 1$, model in which the level-0 behaviour is the rule based behaviour (LkR).

In the Stackelberg model, the focal agent, on account of not holding the right of way, is modelled as the follower. To match the observed (ground truth) manoeuvre with one of the two manoeuvres in our games (*wait* or *proceed*), we first select the trajectory generated in the game that is closest to the observed trajectory based on the trajectory length. The manoeuvre corresponding to that closest trajectory is selected to be the ground-truth manoeuvre.

Fig. 5 shows the mean accuracy, i.e., proportion of match between predicted and observed manoeuvre, across 30 runs. Compared to a baseline (weighted aggregation $\mathbf{w} = [0.5, 0.5]$ indicated as dashed line in figure Fig. 5), a prediction-based aggregation consistently shows better accuracy regardless of the choice of the solution concept. The performance of the models with respect to weighted and satisficing aggregation show some dependency on the specific scenario. For intersection and roundabout scenarios, weighted aggregation shows higher accuracy; whereas for crosswalk, satisficing based aggregation shows higher accuracy. The crosswalk scenario, which is vehicle-pedestrian interaction type, is quite different compared to the intersection or roundabout where there are only vehicle-vehicle interaction games. Drivers are also much more cautious when navigating a crosswalk, since there are pedestrians involved. Combining the insights from the accuracy result along
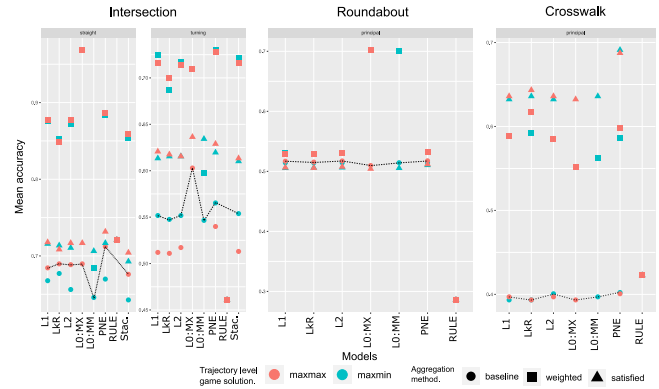

Figure 5: Mean accuracy of all the models. Dashed line is the accuracy of the models with baseline weights.

with the pass rate of Table 1, the data suggest that satisficing is much more effective as an aggregation method in scenarios where drivers exhibit higher levels of caution, such as crosswalk navigation. The final observation is the worse performance of pure rule following indicating that a model of pure rule following might not be best suited as a model of behaviour for the selected situations of high strategic interactions.

6 CONCLUSION

In this paper, we address the problem of estimation of multi-objective aggregation parameters of agents based on observed behaviour. The methods developed in the paper are based on the idea of *rationalisability*, i.e., a value of the aggregation parameter that makes the observed decision of a player optimal conditioned upon a reasoning model. The paper covers two processes of aggregation, namely weighted and satisficing aggregation, and the reasoning models cover strategic as well as non-strategic models. We show that the process of estimating aggregation parameters for weighted aggregation can be formulated as a linear and non-linear program for strategic and non-strategic models, respectively. Furthermore, we develop a novel algorithm for estimation of aggregation parameters for satisficing aggregation that is linear time for strategic models and polynomial time in the size of action space for non-strategic models. Based on a naturalistic dataset of three different traffic scenarios, rationalisable parameters for weighted aggregation were found for all games in the dataset, and for the majority of the games for satisficing aggregation. The paper also includes an extensive evaluation of several solution concepts and behaviour models and shows that, compared to a weighted aggregation with fixed set of weights, the proposed method of utility aggregation improves predictive accuracy across all the chosen solution concepts. In future, the hypotheses generated by the methods in the paper using observational data can be followed up with a controlled experiment to evaluate the aggregation process of human drivers in different situations.

REFERENCES

[1] Sydney N Afriat. 1967. The construction of utility functions from expenditure data. *International Economic Review* 8, 1 (1967), 67–77.

[2] James Andreoni and John Miller. 2002. Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* 70, 2 (2002), 737–753.

[3] Avner Ben-Ner, Louis Putterman, Fanmin Kong, and Dan Magan. 2004. Reciprocity in a two-part dictator game. *Journal of Economic Behavior & Organization* 53, 3 (2004), 333–352.

[4] Kush Bhatia, Ashwin Pananjady, Peter Bartlett, Anca Dragan, and Martin J Wainwright. 2020. Preference learning along multiple criteria: A game-theoretic perspective. *Advances in Neural Information Processing Systems* 33 (2020), 7413–7424.

[5] David Blackwell. 1956. An analog of the minimax theorem for vector payoffs. *Pacific J. Math.* 6, 1 (1956), 1–8.

[6] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. 2017. *Classification And Regression Trees*. Routledge.

[7] Christopher P Chambers, Federico Echenique, and Eran Shmaya. 2017. General revealed preference theory. *Theoretical Economics* 12, 2 (2017), 493–511.

[8] Laurens Cherchye, Thomas Demuynck, and Bram De Rock. 2011. Revealed preference analysis of non-cooperative household consumption. *The Economic Journal* 121, 555 (2011), 1073–1096.

[9] Miguel Costa-Gomes, Vincent P Crawford, and Bruno Broseta. 2001. Cognition and behavior in normal-form games: An experimental study. *Econometrica* 69, 5 (2001), 1193–1235.

[10] Ian Crawford and Bram De Rock. 2014. Empirical revealed preference. *Annu. Rev. Econ.* 6, 1 (2014), 503–524.

[11] Frank E Curtis, Olaf Schenk, and Andreas Wächter. 2010. An interior-point algorithm for large-scale nonlinear optimization with inexact step computations. *SIAM Journal on Scientific Computing* 32, 6 (2010), 3447–3475.

[12] Thomas Demuynck and Per Hjertstrand. 2019. Samuelson's approach to revealed preference theory: Some recent advances. *IFN Working Paper No. 1274* (2019).

[13] Michael Doumpos and Evangelos Grigoroudis. 2013. *Multicriteria Decision Aid and Artificial Intelligence: Links, Theory and Applications*. John Wiley & Sons.

[14] Paweł Dziewulski. 2018. Revealed time preference. *Games and Economic Behavior* 112 (2018), 67–77.

[15] Federico Echenique and Kota Saito. 2015. Savage in the Market. *Econometrica* 83, 4 (2015), 1467–1495.

[16] Jaime F Fisac, Eli Bronstein, Elis Stefansson, Dorsa Sadigh, S Shankar Sastry, and Anca D Dragan. 2019. Hierarchical game-theoretic planning for autonomous vehicles. In *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 9590–9596.

[17] Ray Fuller. 1984. A conceptualization of driving behaviour as threat avoidance. *Ergonomics* 27, 11 (1984), 1139–1155.

[18] Eric Jacquet-Lagreze and Jean Siskos. 1982. Assessing a set of additive utility functions for multicriteria decision-making, the UTA method. *European Journal of Operational Research* 10, 2 (1982), 151–164.

[19] Eric Jacquet-Lagreze and Yannis Siskos. 2001. Preference disaggregation: 20 years of MCDA experience. *European Journal of Operational Research* 130, 2 (2001), 233–245.

[20] Beomjoon Kim and Joelle Pineau. 2016. Socially adaptive path planning in human environments using inverse reinforcement learning. *International Journal of Social Robotics* 8, 1 (2016), 51–66.

[21] Volodymyr Kuleshov and Okke Schrijvers. 2015. Inverse game theory: Learning utilities in succinct games. In *International Conference on Web and Internet Economics*. Springer, 413–427.

[22] Minae Kwon, Erdem Biyik, Aditi Talati, Karan Bhasin, Dylan P Losey, and Dorsa Sadigh. 2020. When Humans Aren't Optimal: Robots that Collaborate with Risk-Aware Humans. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 43–52.

[23] Ben Lewis-Evans, Dick De Waard, and Karel A Brookhuis. 2010. That's close enough—A threshold effect of time headway on the experience of risk, task difficulty, effort, and comfort. *Accident Analysis & Prevention* 42, 6 (2010), 1926–1933.

[24] Changjian Li and Krzysztof Czarnecki. 2019. Urban Driving with Multi-Objective Deep Reinforcement Learning. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. 359–367.

[25] Andrew Y Ng, Stuart J Russell, et al. 2000. Algorithms for inverse reinforcement learning.. In *International Conference on Machine Learning*, Vol. 1. 2.

[26] Maria Porter and Abi Adams. 2016. For love or reward? Characterising preferences for giving to parents in an experimental setting. *The Economic Journal* 126, 598 (2016), 2424–2445.

[27] Sasinee Pruekprasert, Xiaoyi Zhang, Jérémy Dubut, Chao Huang, and Masako Kishida. 2019. Decision making for autonomous vehicles at unsignalized intersection in presence of malicious vehicles. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. IEEE, 2299–2304.

[28] B. Roy. 1971. Problems and methods with multiple objective functions. *Mathematical Programming* 1, 1 (Dec. 1971), 239–266. https://doi.org/10.1007/bf01584088

[29] Dorsa Sadigh, Shankar Sastry, Sanjit A Seshia, and Anca D Dragan. 2016. Planning for autonomous cars that leverage effects on human actions.. In *Robotics: Science and Systems*, Vol. 2. Ann Arbor, MI, USA.

[30] Atrisha Sarkar and Krzysztof Czarnecki. 2021. Solution Concepts in Hierarchical Games Under Bounded Rationality With Applications to Autonomous Driving. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 35. 5698–5708.

[31] Atrisha Sarkar, Kate Larson, and Krzysztof Czarnecki. 2022. Generalized dynamic cognitive hierarchy models for strategic driving behavior. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 5173–5182.

[32] Martin Schmidt-Daffy. 2014. Prospect balancing theory: Bounded rationality of drivers' speed choice. *Accident Analysis & Prevention* 63 (2014), 49–64.

[33] Wilko Schwarting, Javier Alonso-Mora, and Daniela Rus. 2018. Planning and decision-making for autonomous vehicles. *Annual Review of Control, Robotics, and Autonomous Systems* (2018).

[34] Liting Sun, Mu Cai, Wei Zhan, and Masayoshi Tomizuka. 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)*. IEEE, 11038–11044.

[35] Ran Tian, Sisi Li, Nan Li, Ilya Kolmanovsky, Anouck Girard, and Yildiray Yildiz. 2018. Adaptive game-theoretic decision making for autonomous vehicle control at roundabouts. In *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 321–326.

[36] Hal R Varian. 2012. Revealed preference and its applications. *The Economic Journal* 122, 560 (2012), 332–338.

[37] Katarzyna M Werner and Horst Zank. 2019. A revealed reference point for prospect theory. *Economic Theory* 67, 4 (2019), 731–773.

[38] Gerald JS Wilde. 1982. The theory of risk homeostasis: implications for safety and health. *Risk Analysis* 2, 4 (1982), 209–225.

[39] James R Wright and Kevin Leyton-Brown. 2014. Level-0 meta-models for predicting human behavior in games. In *Proceedings of the 15th ACM Conference on Economics and Computation*. 857–874.

[40] James R Wright and Kevin Leyton-Brown. 2020. A formal separation between strategic and nonstrategic behavior. In *Proceedings of the 21st ACM Conference on Economics and Computation*. 535–536.