

The Green Choice: Learning and Influencing Human Decisions on Shared Roads

Erdem Bıyık¹, Daniel A. Lazar², Dorsa Sadigh^{1,3}, and Ramtin Pedarsani²

Abstract—Autonomous vehicles have the potential to increase the capacity of roads via platooning, even when human drivers and autonomous vehicles share roads. However, when users of a road network choose their routes selfishly, the resulting traffic configuration may be very inefficient. Because of this, we consider how to influence human decisions so as to decrease congestion on these roads. We consider a network of parallel roads with two modes of transportation: (i) human drivers who will choose the quickest route available to them, and (ii) ride hailing service which provides an array of autonomous vehicle ride options, each with different prices, to users. In this work, we seek to design these prices so that when autonomous service users choose from these options and human drivers selfishly choose their resulting routes, road usage is maximized and transit delay is minimized. To do so, we formalize a model of how autonomous service users make choices between routes with different price/delay values. Developing a preference-based algorithm to learn the preferences of the users, and using a vehicle flow model related to the Fundamental Diagram of Traffic, we formulate a planning optimization to maximize a social objective and demonstrate the benefit of the proposed routing and learning scheme.

I. INTRODUCTION

Road congestion is a major and growing source of inefficiency, costing drivers in the United States billions of dollars in wasted time and fuel [1]. The emergence of autonomous vehicles promises to improve the efficiency of road usage to some extent by forming platoons and smoothing traffic flow. However, if drivers make their routing decisions *selfishly* and seek to minimize their experienced delay, this can result in suboptimal network performance [2]. Moreover, in the presence of selfish users, the increase in road capacity induced by replacing some human-driven vehicles with autonomous ones can paradoxically worsen average transit user delay [3]. The adverse effects arise from the way humans make decisions about how to route themselves, and the resulting inefficiency can be very large, even unbounded [4].

It is therefore important to consider the effect that self-interested users will have on the system. Previous works have shown that if autonomous users are *altruistic*, average latency can be greatly reduced [4]. However, it's not realistic to assume altruism in many scenarios; in this paper we consider how to monetarily incentivize such prosocial behavior. We consider the setting in which use of autonomous vehicles is offered by some benevolent ride hailing service or social planner which chooses prices for each of the routes. The

users of the autonomous service will choose their routes, as well as whether or not they even want to travel, based on their time-money tradeoff. Then the remaining population of human drivers will choose routes that minimize their delay. The role of the social planner is then to choose prices that will optimize some social objective, which includes maximizing road usage and minimizing average delay. Our model can be viewed as an indirect Stackelberg game – a game in which a social planner controls some fraction of the population's actions and the remainder of the population responds selfishly [5], [6]. However, in our model the planner can only control its portion of the vehicle flow via pricing.

In order to do so effectively, the social planner needs a model for how people make decisions between routes with various prices and latencies. We model these choices as being made based on a modification of Luce's choice axiom [7], [8] and we model human drivers as selfish agents who reach a *Wardrop Equilibrium*, a configuration in which no one could decrease their travel time by switching paths [9].

Moreover, we use a method of actively learning the choices of autonomous service users via a series of queries [10], [11]. This enables the planner to predict how autonomous service users will react to a set of options after a relatively low number of training queries. We verify that this method accurately predicts human choice, and that our planning algorithm can use this to improve network performance, via experiment.

We summarize our contributions as follows.

- *Learning Human Decisions*: We develop an active preference-based algorithm to learn how people value the time-money tradeoff, and choose their preferred transportation option. This enables learning a model of humans' routing choices in a data-efficient manner.
- *Influencing Human Decisions*: We formulate and solve an optimization for ride hailing service to minimize congestion and maximize the road network utilization based on the learned model of human choice, while constraining a minimum profit for the service supplier.
- We validate the learning algorithm and the planning optimization formulation via simulations and user studies. Our results show carefully designed pricing schemes can significantly improve traffic throughput.

Related Work. Previous works have shown the potential benefits that autonomous vehicles can have for traffic networks, by increasing road capacity through platooning [12], [13], damping shockwaves of slowing vehicles [14]–[16], managing merges [17], and decongesting freeways in the event of accidents [18]. In [3], the authors show that the ca-

¹Department of Electrical Engineering, Stanford University

²Department of Electrical and Computer Engineering, UC Santa Barbara

³Department of Computer Science, Stanford University

ebyik@stanford.edu, dlazar@ece.ucsb.edu,
dorsa@cs.stanford.edu, ramtin@ece.ucsb.edu

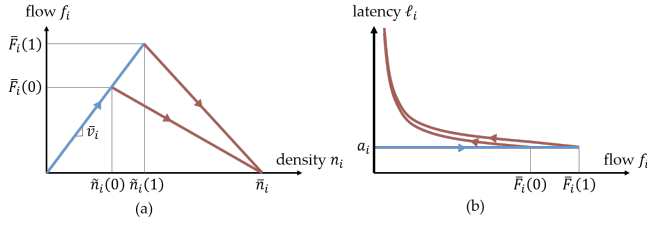


Fig. 1: Vehicle flow model for mixed autonomy based on the Fundamental Diagram of Traffic. (a) and (b) show flow vs density and latency vs flow characteristics if all vehicles are either human-driven or autonomous, where autonomy increases the critical density \tilde{n}_i and the maximum flow \bar{F}_i of a road. The blue and red regions correspond to the road being free-flow and congested states, respectively, and the arrow denotes increasing density.

capacity improvement from platooning can actually *worsen* the total delay experienced by users of the networks. Relatedly, many works (e.g. [19], [20], on shared roads, [21]) analyze and bound the inefficiency that can arise from network users choosing their routes selfishly.

As we use financial incentives to influence this behavior, our formulation is related to work in tolling [22], including works that consider users with different sensitivities to tolls [23], [24]. In contrast to many of these works, as well as some works that develop efficient algorithms for rebalancing an autonomous fleet [25], we consider a probabilistic model for human choice that has been empirically shown to reflect human choice making processes [26], and we model vehicle flow on shared roads based on the foundational Fundamental Diagram of Traffic (FDT) [4], [27], which we will describe in more detail in Section II.

On the human choice models, there have recently been a lot of effort on learning human reward functions which are assumed to be sufficient to model the preferences. Inverse reinforcement learning [28]–[31] and preference-based learning [10], [11], [32], [33] are the most popular choices. In this paper, we employ preference-based learning as it is a natural fit to our problem. We actively synthesize queries as a non-trivial generalization and extension of [10] for data-efficiency and better usability.

II. PROBLEM SETTING

Vehicle Flow Model. We consider a nonatomic model in which individual users control an infinitesimal unit of vehicle flow and this quantity is treated as a continuous variable. To model flow of mixed-autonomous traffic on roads, we use our model in [4], summarized in this section. This is an extension of the FDT [27], in which flow inhabits one of two regimes – free flow, in which the flow on a road *increases* as vehicle density increases, and congestion, in which flow *decreases* as density increases. The flow decreases until the *jam density*, \tilde{n}_i , at which point the flow ceases entirely. Similarly, the relationship between latency and flow on a road shows a similar divide: in free flow latency is constant with respect to flow, and in congestion latency *increases* as flow decreases. These behaviors are shown in Fig. 1. \bar{v}_i denotes the free-flow speed of the road.

The transition between free flow and congestion occurs when the density on a road exceeds the *critical density* of that road, denoted \tilde{n}_i . We model the effect of autonomous

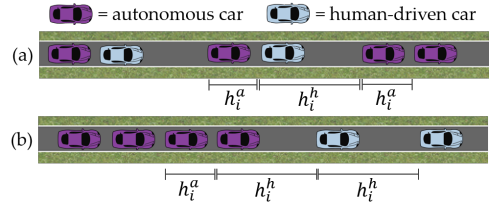


Fig. 2: Illustration of vehicle spacing if (a) autonomous vehicles maintain the same headway behind any vehicle they follow, regardless of type, or (b) autonomous vehicles can reorder themselves to form a large platoon.

vehicles as increasing this critical density, as autonomous cars can maintain shorter headways than human drivers:

$$\tilde{n}_i(\alpha_i) := \frac{b_i}{\alpha_i h_i^a + (1 - \alpha_i) h_i^h}, \quad (1)$$

where $\alpha_i \in [0, 1]$ is the *autonomy level*, or fraction of vehicles on road i that are autonomous, b_i is the number of lanes on that road, and h_i^a and h_i^h are the space occupied by an autonomous and human driven car, respectively, when traveling at the nominal velocity of the road. This model [21] assumes that either the autonomous vehicles maintain the same headway to any vehicle they may follow or that they can rearrange themselves on a road to form large platoons [34], [35], as shown in Fig. 2. Note that if $h_i^a < h_i^h$, the presence of autonomous cars increases the critical density and therefore also increases the maximum flow possible on a road, i.e. $\bar{F}_i(1) > \bar{F}_i(0)$ where \bar{F}_i is a function of autonomy level that gives the maximum capacity of road i . Fig. 1 shows the effect of increased autonomy on a road.

We use s_i to denote the state of road i such that $s_i = 0$ denotes it is in free flow and $s_i = 1$ denotes it is congested. We use n_i^h and n_i^a to denote the density of human-driven and autonomous vehicles, respectively on road i . Accordingly,

$$s_i = \begin{cases} 0 & n_i^h + n_i^a \leq \tilde{n}_i(n_i^a/(n_i^h + n_i^a)) \\ 1 & \text{otherwise} \end{cases}$$

We model the road as first-in-first-out, so $\alpha_i = n_i^a/(n_i^h + n_i^a) = f_i^a/(f_i^h + f_i^a)$, where f_i^h and f_i^a respectively are the human-driven and autonomous vehicle flow on road i .

The flow on a road, $f_i = f_i^h + f_i^a$ is a function of the density of each vehicle type as follows.

$$f_i(n_i^h, n_i^a) := \begin{cases} \bar{v}_i \cdot (n_i^h + n_i^a), & \text{if } n_i^h + n_i^a \leq \tilde{n}_i(\alpha_i) \\ \frac{\bar{v}_i \cdot \tilde{n}_i(\alpha_i) \cdot (\tilde{n}_i - (n_i^h + n_i^a))}{\tilde{n}_i - \tilde{n}_i(\alpha_i)}, & \text{if } \tilde{n}_i(\alpha_i) \leq n_i^h + n_i^a \leq \tilde{n}_i \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

We can then write the latency as a function of vehicle flow and the state of the road [4], [6]:

$$\ell_i(f_i^h, f_i^a, s_i) = \begin{cases} \frac{d_i}{\bar{v}_i} & \text{if } s_i = 0 \\ d_i \left(\frac{\tilde{n}_i}{f_i^h + f_i^a} + \frac{\tilde{n}_i(\alpha_i) - \tilde{n}_i}{\bar{v}_i \cdot \tilde{n}_i(\alpha_i)} \right) & \text{if } s_i = 1 \end{cases}$$

where d_i is the length of road i .

Problem Overview. We assume that the demand of human drivers is fixed, and that the demand of people using the autonomous service is elastic – *if prices and latencies are high, some people may choose not to use the autonomous mobility service*. Our goal is then to simultaneously maximize the number of autonomous service users that can use

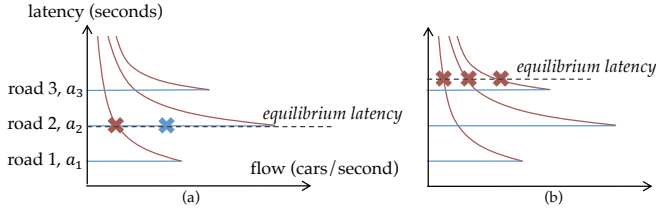


Fig. 3: Some possible equilibria of a three-road network with fixed flow demand. Blue and red lines denote the free-flow and congested regimes, respectively. Equilibria may have (a) one road in free-flow or (b) all used roads congested. An equilibrium has an associated *equilibrium latency* experienced by all selfish users. By considering a given equilibrium latency, we can reason about which roads must be congested at that equilibrium as well how much flow is on each road.

the road, and minimize the average delay experienced by all the people using the roads.

Our control variables in this optimization are the *latencies* on each road and the *prices* offered to the users for traveling on the routes. However, we can't arbitrarily choose prices and latencies – we need to respect

- 1) *the characteristics of the roads*, in terms of how the flow demand for a road corresponds to the latency on that road, and
- 2) *how people make decisions*, making sure that the number of people who choose each option corresponds to the latencies of the roads described in the options.

Moreover, we want to be fair and must therefore offer the same pricing and routing options to all users of the autonomous service.

How a population of users chooses between a variety of price and latency pairs depends on their valuation of time and money. Without knowing this choice model, we cannot plan vehicle flows and will not be able to ensure the resulting configuration matches our vehicle flow models for the roads. Also, since different populations may have different parameter distributions, we need to learn this tradeoff for our population before we can estimate how many people will choose which option. To untangle these constraints, in the subsequent sections we describe our models for vehicle flow and human choices. Following that we return with a mathematical formulation of this optimization.

III. LEARNING THE HUMAN CHOICE MODEL

In this section we describe how both human drivers and the autonomous service users make routing decisions. We then provide a data-efficient learning algorithm for estimating the way that the autonomous service users make routing decisions, allowing us to predict the response a population will have to a set of latency/price options.

Human Drivers. We consider a network of N parallel roads. We assume that no two roads have the same free-flow latency and denote the free-flow latency of road i by a_i . We order the indices such that $a_1 < a_2 < \dots < a_N$, and we use $[k]$ to denote the set of the first k roads.

For the human drivers, we assume their only consideration is minimizing their commute time. This leads to a *Wardrop Equilibrium* [36], [37], which, on parallel roads, means that if one road has positive flow on it, all other roads must have

higher or equal latency. Formally,

$$f_i^h > 0 \implies \ell_i(f_i^h, f_i^a, s_i) \leq \ell_{i'}(f_{i'}^h, f_{i'}^a, s_{i'}) \quad \forall i, i' \in [N].$$

This implies that all selfish users experience the same latency. It is therefore useful to consider the flow-latency diagrams of roads when studying which equilibria exist – by fixing the latency on the y-axis, one can reason about which roads must be congested to achieve that equilibrium. As shown in Fig. 3, equilibria may have one road in free-flow and rest congested, or all may be congested [4], [6].

Autonomous Service Users. Though human drivers are motivated directly only by delay, autonomous service users experience cost in both delay and the price of the ride. We model the users as having some underlying reward function, which is parameterized by their time/money tradeoff, as well as the desirability of the option of traveling by some other means such as walking or public transit. We assume that a strictly dominated option in terms of latency and prices is completely undesirable. Formally,

$$D = \{i \in [N] \mid (p_i > p_{i'} \wedge \ell_i \geq \ell_{i'}) \vee (p_i \geq p_{i'} \wedge \ell_i > \ell_{i'})\} \quad \text{for some } i' \in [N]\}$$

denotes this set of dominated roads. We also define the set of undominated roads, $\bar{D} = [N] \setminus D$. We model the reward function of user j for choosing road i as follows:

$$r_j(\ell, \mathbf{p}, i) = \begin{cases} -\omega_{j1}\ell_i - \omega_{j2}p_i & \text{if user } j \text{ chooses road } i \text{ and } i \in \bar{D} \\ -\infty & \text{if user } j \text{ chooses road } i \text{ and } i \in D \\ -\zeta_j \ell^w & \text{if user } j \text{ declines the service,} \end{cases}$$

where ℓ denotes the vector of road latencies, \mathbf{p} denotes prices, $\omega_j = [\omega_{j1} \ \omega_{j2}]^T$ characterizes the users' time/money tradeoff and ζ_j specifies their willingness to use an alternative option, which has delay ℓ^w . This alternate option could be walking, biking, or public transportation.

We don't assume that users are simple reward maximizers. Rather, we draw from Luce's Choice Axiom [7] to model the probability with which users choose each option, in combination with a softmax rule [26], formalized as follows.

$$P(\text{user } j \text{ chooses route } i) = \frac{\exp(r_j(\ell, \mathbf{p}, i))}{\sum_{i' \in [N]} \exp(r_j(\ell, \mathbf{p}, i'))}. \quad (3)$$

In order to determine optimal pricing, we want to know how many users will choose each route as a function of the route prices and latencies. To this end, we define $q_i(\ell, \mathbf{p})$ as the expected fraction of autonomous service users that will choose route i . If the parameter distribution for autonomous service users is $g(\omega, \zeta)$, then

$$q_i(\ell, \mathbf{p}) = \int_0^\infty \int_0^\infty \int_0^\infty g(\omega, \zeta) P(i|\omega, \zeta) d\omega_1 d\omega_2 d\zeta.$$

where $P(i|\omega, \zeta)$ is the probability that a user with reward function parameters (ω, ζ) will choose road i . This expression relates prices and latency to human choices, enabling us to determine the prices that will maximize our social objective. This will be important in constraining our optimization to only consider latency/price options that correspond to the desired vehicle flows.

Data-Efficient Learning of Human Reward Functions.

While the routes in a specific network can be fully modeled with the physical properties and the speed limits, user's decision models must be learned in a data-driven way. As described above, reward function parameters (ω, ζ) are sufficient to model the decisions of a user.

The parameters will be different among the users. While a business executive might prefer paying extra to reach their company quickly, a graduate student may decide to go to the lab a little later in order to save a few dollars. Therefore, we have to learn personalized parameters ω and ζ .

We learn the parameters from users' previous choices, which is known as *preference-based learning*. If user j chooses from a variety of options, the user's choice gives us a noisy estimate of which road $i \in [N]$ maximizes their reward function $r_j(\ell, \mathbf{p}, i)$. We could start from either uniform priors or priors informed by domain knowledge, then sample from the distribution $g(\omega, \zeta)$.

However, a major drawback of doing so is how quickly we learn the user preferences. Preference-based learning suffers from the small amount of information that each query carries. For example, if we show 4 options to a user (including the option to decline the service), then the maximum information we can get from that query is just 2 bits. To tackle this issue, previous works pose the query synthesis problem as an optimization and maximize a measure of the expected change in the learned distribution after the queries [10], [11].

While those works focused only on pairwise queries, in this case we expect to pose several route options to the users and therefore need more general query types. By using these general queries that offer a variety of routes with varied latencies and prices, we can consider various ways of using this learning framework to learn the human preferences.

- We could do a user study on a few people to learn a good prior distribution.
- We could do an exploration/exploitation strategy if we are allowed to break the fairness constraint a few times for some small portion of the users. This could be made through user-specific campaigns; for each user we may either choose to use the learned model or to offer special rates that would help us profile the user better.
- We could do an initial profiling study for each new user, where we ask them queries synthesized by the active learning framework.

To implement any of these options, we formulate the following active learning optimization. First, we discuss the general preference-based learning framework. Given the data from previous choices of user j , which we denote as \mathcal{D}_j , we can formalize the probability of (ω_j, ζ_j) being true parameters for that user as follows:

$$P(\omega_j, \zeta_j | \mathcal{D}_j) \propto P(\omega_j, \zeta_j) \prod_m P(\mathcal{D}_{j_m} | \omega_j, \zeta_j)$$

where \mathcal{D}_{j_m} denotes the road user j chose in their m^{th} choice (with $\mathcal{D}_{j_m} = 0$ meaning that the user preferred the alternative option). The relation is due to the assumption that the users' choices are conditionally independent from each

other given the reward function parameters.

We can readily obtain the second term from the human choice model. For the prior, we can simply use a uniform distribution over nonnegative values of the parameters. The prior can be crucial especially when we do not have enough data for a new user. In such settings, we incorporate domain knowledge to start with a better prior.

We can then use this unnormalized $P(\omega_j, \zeta_j | \mathcal{D}_j)$ to obtain the samples of (ω_j, ζ_j) using Metropolis-Hastings algorithm. Doing this for each user, which can be easily parallelized, we can directly obtain the samples $(\bar{\omega}, \bar{\zeta}) \sim g(\omega, \zeta)$.

Next we formulate the active learning framework, which is needed so that it will not take an excessive number of queries to learn human preferences. For this, we want to maximize the expectation of the difference between the prior and the unnormalized posterior:

$$\begin{aligned} \text{query}_m^* &= \arg \max_{\text{query}_m} \mathbb{E}_{\mathcal{D}_{j_m}} \left[P(\omega_j, \zeta_j | \mathcal{D}_{j_{1:m-1}}) - \right. \\ &\quad \left. P(\omega_j, \zeta_j | \mathcal{D}_{j_{1:m-1}}) P(\mathcal{D}_{j_m} | \omega_j, \zeta_j, \mathcal{D}_{j_{1:m-1}}) \right] \\ &= \arg \min_{\text{query}_m} \mathbb{E}_{\mathcal{D}_{j_m}} \left[P(\mathcal{D}_{j_m} | \omega_j, \zeta_j, \mathcal{D}_{j_{1:m-1}}) \right] \end{aligned}$$

As we will use the sampled (ω_j, ζ_j) to compute the probabilities of road choices, we can write the optimization as:

$$\text{query}_m^* \doteq \arg \min_{\text{query}_m} \mathbb{E}_{\mathcal{D}_{j_m}} \left[\sum_{\bar{\omega}_j, \bar{\zeta}_j} P(\mathcal{D}_{j_m} | \bar{\omega}_j, \bar{\zeta}_j, \mathcal{D}_{j_{1:m-1}}) \right]$$

where we have M samples denoted as $(\bar{\omega}_j, \bar{\zeta}_j)$, \doteq denotes asymptotic equality as $M \rightarrow \infty$, and the term $1/M$ is canceled. Using the law of total probability,

$$P(\mathcal{D}_{j_m} | \mathcal{D}_{j_{1:m-1}}) \doteq \frac{1}{M} \sum_{\bar{\omega}_j, \bar{\zeta}_j} P(\mathcal{D}_{j_m} | \bar{\omega}_j, \bar{\zeta}_j, \mathcal{D}_{j_{1:m-1}})$$

which then leads to

$$\begin{aligned} \text{query}_m^* &\doteq \arg \min_{\text{query}_m} \sum_{\mathcal{D}_{j_m}} P(\mathcal{D}_{j_m} | \mathcal{D}_{j_{1:m-1}}) \sum_{\bar{\omega}_j, \bar{\zeta}_j} P(\mathcal{D}_{j_m} | \bar{\omega}_j, \bar{\zeta}_j, \mathcal{D}_{j_{1:m-1}}) \\ &\doteq \arg \min_{\text{query}_m} \sum_{\mathcal{D}_{j_m}} \left(\sum_{\bar{\omega}_j, \bar{\zeta}_j} P(\mathcal{D}_{j_m} | \bar{\omega}_j, \bar{\zeta}_j) \right)^2 \end{aligned}$$

We can easily compute this objective value for any given query_m . This optimization is nonconvex due to the human choice model. As in previous works, we assume local optima is good enough [10], [11]. We then use a Quasi-Newton method (L-BFGS [38]) to find the local optima, and we repeat this for 1000 times starting from random initial points.

IV. INFLUENCING HUMAN DECISION MAKING

Problem Formulation. We are now ready to formulate our objective. Since the demand of autonomous service users is elastic, we cannot just minimize the average latency. That would result in extremely high prices to reduce latency by keeping autonomous service users off the network. Hence, we consider the objective to be a combination of maximizing road usage and minimizing average travel time. We param-

terize this tradeoff with parameter $\theta \geq 0$ in the cost function

$$J(\mathbf{f}^h, \mathbf{f}^a, \mathbf{s}) = \frac{\sum_{i \in [N]} (f_i^h + f_i^a) \ell_i(f_i^h, f_i^a, s_i)}{\sum_{i \in [N]} (f_i^h + f_i^a)} - \theta \sum_{i \in [N]} (f_i^h + f_i^a).$$

Given $\mathbf{q}(\ell, \mathbf{p})$, the cost function $J(\mathbf{f}^h, \mathbf{f}^a, \mathbf{s})$, inelastic flow demand of human drivers λ^h , and elastic demand of autonomous users λ^a , we are ready to formulate the planning optimization. The most straightforward way is to optimize jointly over $\mathbf{f}^h, \mathbf{f}^a, \ell$ and \mathbf{p} . However, ℓ is fully defined by $\mathbf{f}^h, \mathbf{f}^a$ and \mathbf{s} . Hence, instead of ℓ , we use \mathbf{s} so that we can search over the possible values of \mathbf{s} . The problem is then:

$$\min_{\mathbf{f}^h, \mathbf{f}^a, \mathbf{p} \in \mathbb{R}_{\geq 0}^N, k \in [N], s_{k:N} \in \{0, 1\}^{N-k+1}} J(\mathbf{f}^h, \mathbf{f}^a, \mathbf{s}) \quad (4)$$

$$\text{subject to } \sum_{i \in [N]} f_i^h = \lambda^h \quad (5)$$

$$f_i^a = \lambda^a q_i(\ell(\mathbf{f}^h, \mathbf{f}^a, \mathbf{s}), \mathbf{p}), \forall i \in [N] \quad (6)$$

$$a_k \leq \ell_k(f_k^h, f_k^a, s_k) \leq a_{k+1} \quad (7)$$

$$f_i^h = 0 \text{ for } i \in [N] \setminus [k] \quad (8)$$

$$\ell_i(f_i^h, f_i^a, 1) = \ell_k(f_k^h, f_k^a, s_k) \text{ for } i \in [k] \quad (9)$$

$$f_i^h + f_i^a \leq \bar{F}(f_i^a / (f_i^a + f_i^h)), \forall i \in [N] \quad (10)$$

$$\sum_{i \in [N]} (f_i^a p_i - f_i^a d_i c) \geq \bar{P} \quad (11)$$

with $s_{1:k-1} = 1$ due to selfishness of human-driven vehicles, where k is the longest road with human-driven vehicle flow, \bar{P} is the minimum profit per unit time that we want to make from autonomous service users and c is the constant fuel cost per unit length, which can be easily calculated by multiplying the vehicles' fuel consumption per unit distance and the fuel price. We can describe the constraints as follows.

(5) The human-driven vehicle flow demand is fulfilled.

(6) Autonomous flow will be distributed based on the choice model described in the preceding section.

(7) The "longest equilibrium road" has latency on the given interval of free flow latencies.

(8),(9) Human-driven cars are selfish.

(10) The maximum capacities of the roads are respected.

(11) The minimum profit per unit time is satisfied.

We can further improve the search space by relying on the heuristic that the roads that are not used by the human-driven vehicles will be in free-flow, i.e. $s_{k+1:N} = 0$. While we do not have a proof for this, we also note constructing counterexamples seems to require extremely careful design, which suggests the heuristic holds in general. Furthermore, the following proposition shows we could also set $s_k = 0$ under an additional assumption.

Proposition 1: Under the assumptions that the reward function is homogeneous among the users and $\omega_2 > 0$, there exists a free-flow road k in the optimal solution to the problem such that $\ell_i = \ell_k$ for $\forall i \in [k]$, and $f_i^h = 0$ for $\forall i \in [N] \setminus [k]$ as long as the optimization is feasible.

The proof is given in Appendix.

Generalizations. We assumed all autonomous cars are controlled by a centralized planner. To extend our framework to scenarios where the planner has the control over a fraction

of autonomous cars, we can simply do the following modifications: The optimization will also be over $\mathbf{f}^b \in \mathbb{R}_{\geq 0}^N$, which will now represent the autonomous flow that does *not* belong to the planner. We add the corresponding constraint of 5 for f_i^b . Similar to 8, we will also have $f_i^b = 0$ for $i > k$ due to selfishness. We will also replace f^a 's in 7, 9 and 10 with $f^a + f^b$ with appropriate subscripts.

Solving the Optimization. We first take M samples $(\bar{\omega}, \bar{\zeta}) \sim g(\omega, \zeta)$ as we previously described in Section III. We then approximate the expected fraction of autonomous service users that will choose road i as:

$$q_i(\ell, \mathbf{p}) \doteq \frac{1}{M} \sum_{\bar{\omega}, \bar{\zeta}} P(\text{user chooses route } i | \bar{\omega}, \bar{\zeta})$$

We then locally solve the nonconvex planning optimization using interior point algorithms [39], [40]. We execute the algorithms with 100 random initial points.

V. EXPERIMENTS & RESULTS

To validate our framework, we conducted different simulations and a user study.

Hypotheses. We test three hypotheses that together suggest our framework successfully reduces traffic congestion through pricing, after it learns the humans' choice models:

H1: Our active learning algorithm can learn the autonomous service users' preferences in a data-efficient way.

H2: Our planning optimization reduces the latencies by creating altruistic behavior through pricing.

H3: When used by humans, the overall framework works well and is advantageous over non-altruistic algorithms.

Implementation Details. In the planning optimization, we used the heuristic that $s_{k+1:N} = 0$, but did not set $s_k = 0$. We assumed there is only walking alternative for autonomous service users. We set $c = 6 \times 10^{-5}$ USD/meter, $\mathbf{b} = 1$. We assumed the human-driven cars keep a 2-second headway distance with the leading car, whereas autonomous cars keep 1-second distance. We set the length of the cars as 5 meters, and the minimum gap between two cars as 2 meters.

Experiments and Analyses. To validate **H1**, we simulated 5 autonomous service users with different preferences. We tested our active learning framework by asking two sets of 200 queries, each of which consisted of 4 road options, similar to Fig. 6, and a walking option. The queries were generated actively in the first set and randomly in the second. After each query, we recorded the sample $(\bar{\omega}, \bar{\zeta})$ which has the highest likelihood as our estimates.

Fig. 4 shows how the estimates evolved within active learning setting for one of the users. All values are overestimated initially. This is intuitively because getting noiseless responses has higher likelihood in the beginning. As we query more, accepting some of the responses as noisy maximizes the likelihood. Therefore, the values start decreasing.

Another important observation is that the estimates of the parameters increase and decrease together even in the early iterations. This suggests we are able to learn the ratio between the parameters, e.g. ω_1/ω_2 very quickly. To check this claim, we used the following error metric:

$$e_{x,y} = \|x/y - \hat{x}/\hat{y}\|_1$$

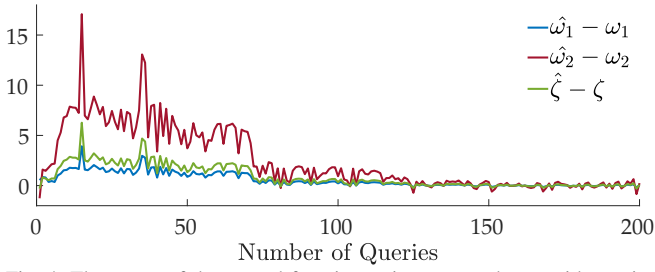


Fig. 4: The errors of the reward function estimates are shown with varying number of queries. $\hat{\omega}_1$, $\hat{\omega}_2$ and $\hat{\zeta}$ represent the estimates.

where $x, y \in \{\omega_1, \omega_2, \zeta\}$ and \hat{x}, \hat{y} represent the corresponding estimates. Figure 5 shows how this error decreases with increasing number of queries. It also shows how active querying enables data-efficient learning compared to the random querying baseline. We are able to learn the relationship between parameters even under 20 queries with active learning. All these results strongly support **H1**.

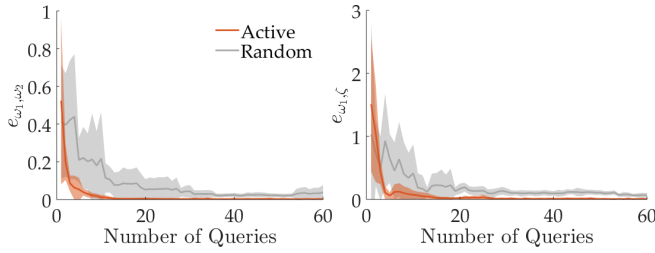


Fig. 5: The error metric is averaged over 5 different reward functions, and the results are plotted.

The fact that we are able to learn the ratios implies we can estimate which road the user is most likely to choose. We will only be unsure about how noisy the user is if the parameter estimates did not converge yet. Therefore, we can still use the estimates for our planning optimization even when we have small number of queries.

$$d_3 = 1000\pi \text{ m}, \bar{v}_3 = 25.00 \text{ m/s}$$

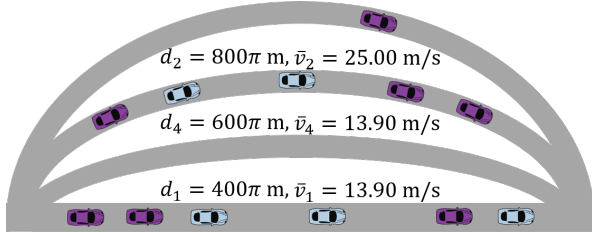


Fig. 6: The 4-road network from [4]. The roads are not to the scale and ordered with respect to the free-flow latencies.

To validate **H2**, we adopt 3 of the equilibria from [4] for comparison, namely, Nash Equilibrium where all drivers are selfish; Best Nash Equilibrium which gives the smallest average latency again with selfish drivers; and Fully Altruistic Best Nash Equilibrium which gives the smallest average latency where autonomous users are fully altruistic and human-driven cars route selfishly. Here, we give the average latencies for the specific 4-road network from that study which we visualize in Fig. 6, and where $\lambda^h = 0.4$, $\lambda^a = 1.2$ cars per second:

- Nash Equilibrium: 400.00 seconds
- Best Nash Equilibrium: 125.66 seconds
- Fully Altruistic Best Nash Equilibrium: 102.85 seconds

We then assumed we perfectly learned the preferences of the 5 simulated users. We ran the planning optimization with

$\bar{P} = 0$ and 3 different θ to show the trade-off. The results are summarized in Table I.

TABLE I: Results of Routing Simulation

θ	Avg. Latency (seconds)	Flow (cars/second)
1	90.41	0.4412
20	97.03	1.2746
10^6	111.28	1.5964

It can be seen we can adjust the trade-off between average latency and the served flow by tuning θ . Also, given the human preferences, even when we served (almost) all of autonomous demand, our framework outperforms Best Nash Equilibrium. This shows the effectiveness of our framework on creating altruism and supports **H2**.

For **H3**, we recruited 21 subjects (9 female, 12 male) with an age range from 19 to 60. In the first phase of the experiment, each participant was asked 40 queries (4 roads + 1 walking option) which are all actively synthesized. We then used their responses to get the sample $(\bar{\omega}, \bar{\zeta})$ that has the highest likelihood. Afterwards, we designed 5 different road networks each with 4 different roads and an additional route where people may choose to walk. The roads on the 5 different networks cover a range of different road lengths from 1.8 kilometers to 78 kilometers. For each network, we also set different θ , λ^h , λ^a , and \bar{P} . By assuming all autonomous flow is in the service of these 21 subjects, or of the groups that match with their preferences, we locally solved the planning optimization to obtain the pricing scheme for each traffic network. We refer to the results of this optimization as *anticipated* values.

In the second phase of the user study, we presented the route-price pairs and the walking option to the same 21 subjects. For each of the 5 networks, they picked the option they would prefer. Using these responses, we allocated the autonomous flow into the roads. However, it is technically possible that more users select a road than its maximum flow. To handle such cases, we assumed extra flow is transferred to the roads with smaller latencies without making the users pay more. If that is not feasible, extra flow is transferred to the slower roads, without any discount. While these break the fairness constraint, it rarely happens and affects only a very small portion of the flow. After autonomous flows are allocated, human driven cars selfishly chose their routes in a way to minimize the overall average latency. We refer to the results of this allocation as *actual* values.

Table II compares the anticipated and the actual values. We report latencies in seconds, flows in cars per second, and profit is a rate value with the unit USD per second. In order to show how our framework incentivizes altruistic behavior, we also added two other selfish best Nash equilibrium baselines: one where the same flow as actual flow is routed (BNE1) and one where all flow demand is routed (no walking option) again under best Nash equilibrium (BNE2).

It can be seen that there is generally an alignment between anticipated and actual values. While the mismatch may be further reduced by doing more queries or having more users, the difference with BNE methods is significant. In all cases, our framework achieved to incentivize altruism,

which yielded lower average latencies compared to BNE1. Especially in Case 3 and Case 5, our framework approximately halved the average latency compared to the best Nash equilibria. We visualize this difference in Fig. 7. Furthermore, our framework successfully reduced flow demand when satisfying the full demand is impossible (Case 1).

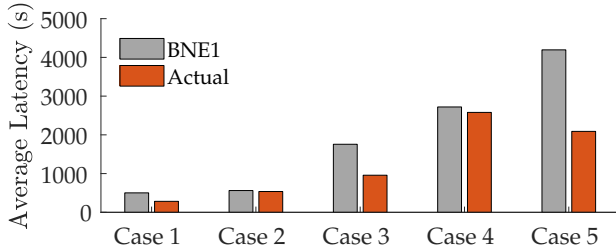


Fig. 7: The comparison of the actual results and BNE1, both of which allocate the same amount of flow.

One caveat is the small amount of actual flow in Case 1, which also caused an important profit loss. This is because the roads are relatively shorter, and most users preferred walking over paying for an autonomous car. Our framework could not predict this, because the learned reward functions failed to accurately estimate the probabilities. This issue will be further discussed in Section VI.

VI. CONCLUSION

In this work, we develop a method improving the efficiency of traffic networks shared between human-driven and autonomous vehicles by influencing the routing behavior of humans. Considering an autonomous service that offers a choice of a number of routes, we formulate an optimization to find the prices that influence user choices so that when they choose their routes, and the human-driven traffic routes solely based on travel delay, travel delay is minimized and road usage is maximized. In order to do so, we model how people choose between routes with differing prices and latencies, and develop a method to actively learn the choice model parameters for a population of users. We verify these mechanisms through a user study.

There are areas in which our mechanism can be improved. Specifically, one could further refine the reward function model for autonomous service users and not necessarily assume that it's linear. One could also consider the human-driven vehicle demand to be elastic, meaning that people may choose to stay home or choose public transportation if congestion is high. Further, we find that the profit constraint is often unsatisfied in the actual results with a small margin. We could instead include profit as part of the objective function or optimize for the worst-case performance.

We considered only a single alternative to the autonomous service. However, a population may choose from a variety of options such as walking, biking, or taking the bus, ζ will have a multimodal distribution; we then need to learn the mixture. Finally, it is unclear if it is better to use many samples from our posterior on (ω, ζ) or just the maximum likelihood sample while computing q for the planning optimization.

Future work can address these issues or expand in broader directions. It would be worthwhile to look at more general network topologies. Further, the information people have

access plays a huge role in decision making. It would be enlightening to develop choice models that incorporate this information availability as well as possibly other aspects of decision making, such as risk aversion. These expansions would do much to address congestion on roads that are soon to be shared between human drivers and autonomous cars.

ACKNOWLEDGMENTS

Toyota Research Institute (TRI) provided funds to assist the authors with their research but this article solely reflects the opinions and conclusions of its authors and not TRI or any other Toyota entity.

REFERENCES

- [1] D. Schrank, B. Eisele, T. Lomax, and J. Bak, "2015 urban mobility scorecard," 2015.
- [2] E. Koutsoupias and C. Papadimitriou, "Worst-case equilibria," in *Annual Symposium on Theoretical Aspects of Computer Science*. Springer, 1999, pp. 404–413.
- [3] N. Mehr and R. Horowitz, "Can the presence of autonomous vehicles worsen the equilibrium state of traffic networks?" in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 1788–1793.
- [4] E. Biyik, D. Lazar, R. Pedarsani, and D. Sadigh, "Altruistic autonomy: Beating congestion on shared roads," in *Workshop on Algorithmic Foundations of Robotics (WAFR)*, 2018.
- [5] C. Swamy, "The effectiveness of stackelberg strategies and tolls for network congestion games," *ACM Transactions on Algorithms (TALG)*, vol. 8, no. 4, p. 36, 2012.
- [6] W. Krichene, J. D. Reilly, S. Amin, and A. M. Bayen, "Stackelberg routing on parallel transportation networks," *Handbook of Dynamic Game Theory*, pp. 1–35, 2017.
- [7] R. D. Luce, *Individual choice behavior: A theoretical analysis*. Courier Corporation, 2012.
- [8] M. E. Ben-Akiva, S. R. Lerman, and S. R. Lerman, *Discrete choice analysis: theory and application to travel demand*. MIT press, 1985, vol. 9.
- [9] J. R. Correa and N. E. Stier-Moses, "Wardrop equilibria," *Wiley encyclopedia of operations research and management science*, 2011.
- [10] D. Sadigh, A. D. Dragan, S. Sastry, and S. A. Seshia, "Active preference-based learning of reward functions," in *Robotics: Science and Systems (RSS)*, 2017.
- [11] E. Biyik and D. Sadigh, "Batch active preference-based learning of reward functions," in *Conference on Robot Learning*, 2018, pp. 519–528.
- [12] J. Lioris, R. Pedarsani, F. Y. Tascikaraoglu, and P. Varaiya, "Platoons of connected vehicles can double throughput in urban roads," *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 292–305, 2017.
- [13] A. Askari, D. A. Farias, A. A. Kurzhanskiy, and P. Varaiya, "Effect of adaptive and cooperative adaptive cruise control on throughput of signalized arterials," in *Intelligent Vehicles Symposium (IV)*, 2017 IEEE. IEEE, 2017, pp. 1287–1292.
- [14] R. E. Stern, S. Cui, M. L. Delle Monache, R. Bhadani, M. Bunting, M. Churchill, N. Hamilton, H. Pohlmann, F. Wu, B. Piccoli, et al., "Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments," *Transportation Research Part C: Emerging Technologies*, vol. 89, pp. 205–221, 2018.
- [15] R. K. Bhadani, B. Piccoli, B. Seibold, J. Sprinkle, and D. Work, "Dissipation of emergent traffic waves in stop-and-go traffic using a supervisory controller," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 3628–3633.
- [16] C. Wu, A. M. Bayen, and A. Mehta, "Stabilizing traffic with autonomous vehicles," in *To appear, International Conference on Robotics and Automation*, 2018.
- [17] L. Jin, M. Čičić, S. Amin, and K. H. Johansson, "Modeling the impact of vehicle platooning on highway congestion: A fluid queuing approach," in *Proceedings of the 21st International Conference on Hybrid Systems: Computation and Control (part of CPS Week)*. ACM, 2018, pp. 237–246.

TABLE II: Results of Real-User Experiments

	Anticipated			Actual			BNE1		BNE2	
	Avg. Latency	Flow	Profit	Avg. Latency	Flow	Profit	Avg. Latency	Flow	Avg. Latency	Flow
Case 1	338.10	1.6622	11.11	283.30	1.0429	3.85	501.97	1.0429	Infeasible	
Case 2	510.97	1.5592	4.50	537.30	1.5048	3.95	562.20	1.5048	562.20	1.6000
Case 3	921.06	1.2677	4.59	957.71	1.3000	3.99	1756.89	1.3000	1756.89	1.3000
Case 4	2083.42	1.2383	48.48	2089.53	1.2048	46.32	4194.26	1.2048	4194.26	1.3000
Case 5	2576.36	1.5538	40.01	2579.52	1.6000	49.33	2720.00	1.6000	2720.00	1.6000

- [18] S. Sivaranjani, Y.-S. Wang, V. Gupta, and K. Savla, "Localization of disturbances in transportation systems," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE, 2015, pp. 3439–3444.
- [19] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, "A geometric approach to the price of anarchy in nonatomic congestion games," *Games Econ. Behavior*, vol. 64, no. 2, pp. 457–469, 2008.
- [20] T. Roughgarden and É. Tardos, "How bad is selfish routing?" *Journal of the ACM (JACM)*, vol. 49, no. 2, pp. 236–259, 2002.
- [21] D. A. Lazar, S. Coogan, and R. Pedarsani, "Routing for traffic networks with mixed autonomy," *arXiv preprint arXiv:1809.01283*, 2018.
- [22] M. Beckmann, C. B. McGuire, and C. B. Winsten, "Studies in the economics of transportation," Tech. Rep., 1956.
- [23] L. Fleischer, K. Jain, and M. Mahdian, "Tolls for heterogeneous selfish users in multicommodity networks and generalized congestion games," in *45th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, 2004, pp. 277–285.
- [24] P. N. Brown and J. R. Marden, "The robustness of marginal-cost taxes in affine congestion games," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3999–4004, 2017.
- [25] M. Salazar, M. Tsao, I. Aguiar, M. Schiffer, and M. Pavone, "A congestion-aware routing scheme for autonomous mobility-on-demand systems," in *European Control Conference (ECC)*, submitted, 2019.
- [26] N. D. Daw, J. P. O'doherty, P. Dayan, B. Seymour, and R. J. Dolan, "Cortical substrates for exploratory decisions in humans," *Nature*, vol. 441, no. 7095, p. 876, 2006.
- [27] C. F. Daganzo, "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory," *Transportation Research Part B: Methodological*, vol. 28, no. 4, pp. 269–287, 1994.
- [28] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 1.
- [29] A. Y. Ng, S. J. Russell, et al., "Algorithms for inverse reinforcement learning," in *ICML*, vol. 1, 2000, p. 2.
- [30] P. Abbeel and A. Y. Ng, "Exploration and apprenticeship learning in reinforcement learning," in *Proceedings of the 22nd international conference on Machine learning*. ACM, 2005, pp. 1–8.
- [31] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning," in *Aaai*, vol. 8. Chicago, IL, USA, 2008, pp. 1433–1438.
- [32] R. Akrou, M. Schoenauer, and M. Sebag, "April: Active preference learning-based reinforcement learning," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2012, pp. 116–131.
- [33] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," in *Advances in Neural Information Processing Systems*, 2017, pp. 4299–4307.
- [34] D. A. Lazar, S. Coogan, and R. Pedarsani, "The price of anarchy for transportation networks with mixed autonomy," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 6359–6365.
- [35] D. A. Lazar, R. Pedarsani, K. Chandrasekhar, and D. Sadigh, "Maximizing road capacity using cars that influence people," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 1801–1808.
- [36] J. Wardrop, "Some theoretical aspects of road traffic research," in *Inst Civil Engineers Proc London/UK*, 1900.
- [37] A. De Palma and Y. Nesterov, "Optimization formulations and static equilibrium in congested transportation networks," Tech. Rep., 1998.
- [38] G. Andrew and J. Gao, "Scalable training of l1-regularized log-linear models," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 33–40.
- [39] R. H. Byrd, J. C. Gilbert, and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming," *Mathematical programming*, vol. 89, no. 1, pp. 149–185, 2000.
- [40] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban, "An interior algorithm for nonlinear optimization that combines line search and trust region steps," *Mathematical programming*, vol. 107, no. 3, pp. 391–408, 2006.

APPENDIX

A. Proof of Proposition 1

Let us call the reward function that is homogeneous among the users as \bar{r} . Assume the optimal solution is $(\ell^*, \mathbf{p}^*, \mathbf{f}^h, \mathbf{f}^{a*})$, and $\ell_k^* > a_k$, where $k = \arg \max_i a_i$ subject to $\mathbf{f}^{h*} > 0$. Due to selfishness, we know $\ell_i = \ell_k$ for $\forall i \in [k]$. Also, let D^* (and \bar{D}^*) be the set of dominated (undominated) options under the optimal solution.

Let (ℓ', \mathbf{p}') be such that

- $\ell'_i = \ell_i^*$ and $p'_i = p_i^*$, for $\forall i \in [N] \setminus [k]$,
- $a_k \leq \ell'_k < \ell_k^*$,
- $\ell'_i = \ell'_k$, for $\forall i \in [k-1]$,
- $p'_i = p_i^* + \epsilon$ where $\epsilon > 0$ for $\forall i \in [k] \cup D^*$,
- $\mathbf{q}(\ell', \mathbf{p}') = \mathbf{q}(\ell^*, \mathbf{p}^*)$ for $i \in [k]$.

We also define D' and \bar{D}' similarly. The only two free variables of (ℓ', \mathbf{p}') are ℓ'_k and ϵ . To show such an (ℓ', \mathbf{p}') with the last constraint exists, we prove 1) For any ℓ'_k , there exists an ϵ such that $\bar{r}(\ell^*, \mathbf{p}^*, i) = \bar{r}(\ell', \mathbf{p}', i)$ for $\forall i \in [k]$, 2) $D^* = D'$. These together imply $\bar{r}(\ell^*, \mathbf{p}^*, i) = \bar{r}(\ell', \mathbf{p}', i)$ for $\forall i \in [N]$, which then implies the last constraint.

The first one is easy to show. $\bar{r}(\ell^*, \mathbf{p}^*, i) = \bar{r}(\ell^*, \mathbf{p}^*, k)$ for $\forall i \in [k]$. For any ℓ'_k , we can set $\epsilon = \frac{\omega_1}{\omega_2}(\ell_k^* - \ell'_k)$ to get $\bar{r}(\ell', \mathbf{p}', i) = \bar{r}(\ell^*, \mathbf{p}^*, i)$ for $\forall i \in [k]$ where ω is associated with \bar{r} . The equality also holds for $i \in [N] \setminus [k]$, because the prices and latencies are the same between (ℓ^*, \mathbf{p}^*) and (ℓ', \mathbf{p}') in those roads.

We show $D^* = D'$ in two steps:

- 1) $D^* \cap [k] = D' \cap [k]$, since the roads had the same latency drop and price increase, and they cannot be dominated by any of $[N] \setminus [k]$, because they have strictly lower latency.
- 2) $D^* \setminus [k] = D' \setminus [k]$, because the roads have the same latency between two configurations and the prices went up only for those in D^* , so they cannot dominate each other in a different way between the configurations. They also cannot dominate any of the roads in $[k]$, because they have strictly higher latency. And $i \in D' \setminus [k] \implies i \in D^* \setminus [k]$, because the roads in $[k]$ had already lower latency in the optimal solution and their prices only went up.

Having showed (ℓ', \mathbf{p}') exists, we note $(\ell', \mathbf{p}', \mathbf{f}^h, \mathbf{f}^{a*})$ is a feasible solution for some \mathbf{f}^h , because the distribution of autonomous service users remained the same and the prices went only up, which together means the profit constraint (11) is still satisfied. The other constraints are trivially satisfied by the construction of (ℓ', \mathbf{p}') and as the optimal solution has to be feasible. Then, since the demand of human-driven cars are inelastic, the total served flow is the same between two solutions. However, the latencies of

the roads in $[k]$ decreased, which means their capacities increased. Hence, $k \geq \arg \max_i a_i$ subject to $f_i^{\text{h}'} > 0$. Then, the average latencies of human-driven cars decreased. Similarly, the average latencies of autonomous service users may only go down. Hence, $J(\mathbf{f}^{\text{h}*}, \mathbf{f}^{\text{a}*}, \mathbf{s}^*) > J(\mathbf{f}^{\text{h}'}, \mathbf{f}^{\text{a}'}, \mathbf{s}')$, and $(\ell^*, \mathbf{p}^*, \mathbf{f}^{\text{h}*}, \mathbf{f}^{\text{a}*})$ cannot be the optimal solution. \square