Congestion and Incentives in the Age of Driverless Cars*

Federico Boffa† Alessandro Fedele‡ Alberto Iozzi§

January 30, 2020

Abstract

Following the development of autonomous vehicles (AVs) and GPS systems, fleets will gain prominence over private vehicles. We analyze the welfare effects of the transition from a fully decentralized regime, in which all travelers are atomistic and do not internalize the congestion externality, to a centralized regime, where all travelers are supplied by a fleet of AVs controlled by a monopolist. In our model, heterogeneous individuals differing in the disutility from congestion may travel on one of two lanes, which may endogenously differ in the level of congestion, or they may not travel. We show that the monopolist sorts travelers across the two lanes differently than the decentralized regime. Moreover, depending on the severity of congestion costs, it may also exclude some travelers. We find that centralization is always welfare detrimental when the monopolist does not ration travel. If instead rationing occurs, centralization may be welfare beneficial, provided that congestion costs are sufficiently high. We then analyze how to restore first best with road taxes. While congestion charges are optimal under decentralization, taxes differ markedly in a centralized regime, where restoring first best may require subsidizing the monopolist.

Keywords: autonomous vehicles, congestion externality, fleets, sorting, rationing.
JEL Codes: R41, R11.

*We are grateful to Toomas Hinnosaar, Ryan Kellogg, Aleksey Tetenov and Matthew Turner for useful discussions and helpful comments. We thank participants to the NBER Workshop on "The economics of autonomous and electric vehicles" in Stanford, to seminars at Collegio Carlo Alberto, Free U. of Bolzano, Hong Kong Polytechnic, IMT Lucca, U. of Siena, SOAS U. of London, U. of Rome 'Tor Vergata', to workshops at Católica Porto Business School, U. of Sassari (IBEO), and U. of Bergamo, and to session conferences at AISRE, ASSET, EARIE, ERSA, IAERE, ITEA, JEI, SIE, and SIET. All errors are ours.

†Free University of Bolzano and Collegio Carlo Alberto: federico.boffa@unibz.it.
‡Free University of Bolzano: alessandro.fedele@unibz.it.
§Università di Roma 'Tor Vergata' and SOAS University of London: alberto.iozzi@uniroma2.it.

1 Introduction

Technological advancement is rapidly changing the mobility industry. One of the most prominent innovations is the development of Autonomous Vehicles (AVs), that is, vehicles driven by a software that does not require human intervention. AVs open a host of relevant technological, legal and moral issues (Awad et al., 2018). But significant, and as of now underappreciated, impacts will come from the changes in the structure of the mobility market. The provision of mobility services through fleets will likely gain prominence over private vehicles, at least in urban contexts. Fleets will turn cheaper due to savings in the drivers' labor costs. At the same time, their utilization rate will remain higher than that of private vehicles, contributing to a superior efficiency. This trend is already underway, with the emergence of fleets of robotaxis, i.e., taxis operated by ride service companies through AVs. Waymo, an Alphabet's subsidiary and one of the leading companies in the development of self-driving technology, has partnered with Lyft and is currently offering robotaxi services in selected locations in the United States.[1] General Motor's subsidiary Cruise is also expected to roll out its robotaxis fleet soon.[2] Tesla has recently announced that it will soon stop selling cars to private owners, and use them for its own - comparatively more profitable - robotaxis fleet instead.[3] Consumers' investment into private cars is thus bound to shrink (Fagnant and Kockelman, 2015). As a result, urban transport will likely be centralized, and managed by a handful of transport service suppliers.

Such centralization will have dramatic consequences on how congestion will be dealt with. Congestion in the mobility industry not only derives from transport infrastructures being inadequate relative to demand, but is also the result of a standard externality. In the current decentralized setting, drivers do not factor in their travel decisions the external effect in terms of congestion they impose on fellow travelers.

The paper studies how the transition to a centralized mobility market, organized around companies that manage their own fleets, affects congestion, and, as a result, welfare. It may seem intuitive that firms in a centralized setting internalize (at least partly) the congestion externality inherent in the fully decentralized setting, thereby contributing to an increase in welfare. We show that this intuition is incomplete, and we emphasize the distortions occurring in a centralized market. We then analyze the taxation schemes that allow to restore first best.

Understanding the impact of traffic centralization on congestion is important because congestion costs, while hidden and hard to measure because of their nature of opportunity costs, represent a major component of the traveling costs. Congestion not only increases the average travel time, but it also raises its variance. The yearly con-

[1] https://www.wsj.com/articles/lyft-to-offer-waymo-self-driving-taxis-in-suburban-phoenix-11557259648 (last accessed January 29, 2020).

[2] https://www.wsj.com/articles/gm-s-driverless-car-unit-cruise-delays-robot-taxi-service-11563971401 (last accessed January 29, 2020).

[3] https://www.wsj.com/articles/a-tale-of-two-teslas-elon-musk-to-tout-robot-cars-amid-sales-slump-11555848000 (last accessed January 29, 2020).

gestion cost has been estimated to amount to more than one hundred billion dollars in the US, and to be steadily increasing over time (Schrank, Lomax and Eisele, 2011). In the absence of policies or of market design interventions, congestion costs are not bound to disappear when AVs will be deployed. On the one hand, with AVs, consumers may spend more productively their time on vehicles, thereby reducing the disutility of congestion when one holds the level of congestion constant. On the other hand, however, the ability to use time in the car more fruitfully will increase congestion, by inducing more travel demand (Gucwa, 2014). The overall effect on disutility of congestion will depend on the relative strength of the two effects, something hard to predict now. It is well possible that congestion costs will ultimately increase, reproducing an effect similar to the positive association between new infrastructure and kilometers traveled expressed in the fundamental law of road congestion (Downs, 1962; Duranton and Turner, 2011).

A proper analysis of congestion requires to consider travelers' heterogeneity in the disutility from congestion. This is substantial (see, for instance, Small, 2012) and reflects heterogeneity in individuals' value of time, as well as in value of reliability. Small, Winston and Yan (2005) find that the difference in value of time between the 25th and the 75th percentile is about \$10 per hour (with the median being about \$21 per hour). Even more starkly, the difference in value of reliability between the 25th and the 75h percentile is \$13 per hour (higher than the median of \$12 per hour). Also, as intuitive, disutility from congestion is known to be tightly positively associated to income. Estimates of the elasticity of time value to income range from about 0.5 to 1, and they are increasing at higher levels of income (Börjesson, Fosgerau and Algers, 2012). The current trends of increasing inequality in many Western countries, in the context of a rising average income, should only exacerbate such heterogeneity.

With heterogenous travelers, the reduction in aggregate congestion costs (as well as welfare maximization) requires to act not only on the margin of the total number of vehicles that travel, but also on the efficient sorting of travelers. To see this, consider a highway with two lanes. With heterogeneous disutility from congestion, efficiency may require to differentiate the speed across the two lanes, thereby creating a faster lane with fewer vehicles for travelers that dislike congestion a lot, and a slower lane with more vehicles for drivers who are less bothered by congestion.

Efficient sorting of travelers is economically very relevant. The combined evidence in Small, Winston and Yan (2005, 2006) shows that a little less than the average hourly disutility from congestion can be mitigated through an efficient consumer sorting. They simulate the effects of several alternative lane management schemes (through tolls and/or high occupancy vehicles lanes), which give rise to different allocations of heterogenous travelers on a California State highway. They find that the welfare-maximizing scheme yields about \$2.50 welfare gain in a 15 minute trip over the alternative with no tolls and no lane utilization rules. This is equivalent to \$10 per hour per passenger, which is more than half of the average hourly wage in the United States, and just less than half of the median value of time that they estimate.

In addition, sorting vehicles across different routes will become increasingly easy for a central decision-maker, be it a government or a company managing a fleet of vehicles. The software that drives the AVs may itself be used to sort vehicles across different routes, even based on other vehicles' behavior, without the possibility of interference by travelers. Also, GPS and mobile technologies will contribute to the diffusion of road pricing schemes. These schemes will be increasingly sophisticated, as prices may be conditioned on a variety of dimensions (route, time of the day, occupancy, etc.), and cheaper, as a result of the reduction in the costs of the required infrastructure (Ostrovsky and Schwarz, 2018).

We consider a stylized framework with individuals using AVs to travel on a road, segmented into two separate parallel lanes, both congested. The lanes are ex-ante identical, but can ex-post differ in the amount of congestion. Individuals are assumed to be heterogenous as to the utility they derive from the trip and to the disutility they derive from the congestion. Consistent with evidence pointing to a positive relation between income and value of time, we assume that individuals with a larger utility from the trip suffer from a larger cost of congestion. We look at the equilibrium assignment of individuals to one of the two lanes or to not traveling.

We first show that welfare maximization requires to differentiate the congestion level in the two lanes, reflecting the heterogeneity in individuals' value of time. Individuals with low disutility of congestion will travel in a slow lane, while those with a high disutility of congestion will travel in a fast lane. Furthermore, an individual travels as long as her benefit from traveling exceeds the increase in aggregate congestion costs she imposes on fellow travelers. Thus, if the congestion cost is sufficiently large, efficiency requires to prevent some low-value individuals from traveling.

We then study the welfare effects of moving from a fully decentralized to a fully centralized regime. In a fully decentralized regime, all individuals are atomistic and do not factor in their travel decisions the external congestion effect they impose on fellow travelers. In a fully centralized regime, all the mobility services are provided by a monopolist through its fleet. During the transition period, we look at a partially centralized regime, where individuals are exogenously assigned to being atomistic or to being supplied by the monopolist. This reflects the fact that the transition towards an economy fully organized around fleets will likely be gradual. We first focus on how travelers are sorted across the two lanes, holding their total number fixed. We then let the aggregate amount of travelers vary and allow for rationing.

When the total number of travelers is fixed, the two lanes have exactly the same number of travelers under full decentralization. When the market is fully centralized there is too much differentiation in congestion across lanes with respect to the social optimum. From the welfare standpoint, this excess differentiation turns out to be worse than the no differentiation prevailing in the decentralized setting. An additional welfare-reducing inefficiency emerges under the partially centralized regime, in that all atomistic travelers, including those with a relative low value of traveling and disutility from congestion, travel in the fast lane. Overall, we find that welfare monotonically decreases in moving from fully decentralized to fully centralized travel.

Instead, when the total number of travelers is allowed to vary, a monopolist may have an incentive to exclude some low-value individuals from traveling, in order to increase prices on, and extract more value from, travelers. By contrast, all individuals travel in the decentralized setting. If congestion costs are sufficiently severe that the social planner would efficiently exclude some low-value agents from traveling, the quantity reduction operated by the firm with respect to decentralized travel may be efficient. If, to the contrary, congestion costs are not so large in the first place, and the planner would dispatch all the travelers, the monopolist's screening is welfare-reducing over decentralized travel.

Part of our results parallel those obtained in the airline economics literature, under monopolistic air carriers, in terms of the congestion levels: see, for instance, Brueckner (2002) and Basso (2008), and the empirical counterparts estimating the relation between airport concentration and congestion (Mayer and Sinai, 2003; Rupp, 2009; Daniel and Harback, 2008; and Molnar, 2013). However, we crucially add travelers' heterogeneity in the disutility from congestion to the picture, and, as a result, we analyze how consumers sorting across lanes can be used to mitigate the negative effects of congestion on welfare. Our findings also relate to Czerny and Zhang (2015), who study third degree price discrimination in the presence of congestion externalities.

We then analyze the case of a tax authority able to impose taxes that restore social optimality. In a fully decentralized regime, a traditional congestion charge, i.e., a Pigouvian tax, equal to the marginal (external) cost imposed on the other vehicles, restores optimality. This mirrors the finding obtained in the bottleneck model (see Vickrey, 1969; and Arnott, de Palma and Lindsey, 1990). Instead, under full centralization, the tax that restores social optimum is very different. Since the monopolist considers in its pricing policy the effects of the congestion costs on all the vehicles it dispatches, there is no scope for a congestion charge. This result aligns with those obtained in the literature on airports when carries have market power (Daniel, 1995; Brueckner, 2002; Pels and Verhoef, 2004; Brueckner 2005; Basso and Zhang, 2007; and Silva and Verhoef, 2013). We then characterize a simple tax/subsidy scheme that restores the incentives to optimality both in the degree of differentiation across lanes and in the aggregate number of travelers. We show that, when the congestion problem is particularly severe, this scheme involves a net subsidy for the monopolist. This may prove politically challenging, even more than traditional congestion pricing schemes, and may require some countermeasures by the tax authority that improve its political feasibility. An example of them could be the collection of license fees from the monopolist, so as to balance the tax budget within the AV market.

To the best of our knowledge, there are only a few papers that consider congestion with reference to AVs. Lamotte, De Palma and Geroliminis (2016) develop a bottleneck model to investigate the commuters' choice between conventional and autonomous vehicles, while van den Berg and Verhoef (2016) focus on the impact of AVs on road capacity, studying the deployment of infrastructures resulting from the transition to AVs.[4] Finally, our paper is close to Ostrovsky and Schwarz (2018), who investigate the interplay between autonomous transportation, carpooling, and road pricing to achieve socially efficient outcomes.

The rest of the paper is organized as follows. Section 2 describes the model. Section 3 illustrates the first best. Section 4 and Section 5 characterize the equilibrium when the total number of travelers is fixed and is allowed to vary, respectively. Section 6 analyzes taxation to restore social optimality. Section 7 concludes. Derivations and proofs of all Propositions and Lemmata are relegated to an appendix.

2 The model

Lanes and travelers' utility. There is a unit mass of individuals, each with unit demand for a trip from a common origin to a common destination. Trips have zero production costs and occur along a single road connecting the origin and the destination. The road is divided into two ex-ante identical lanes that are congested at any positive mass of travelers. The lanes may, however, differ ex-post because of a different mass of travelers, leading to different levels of congestion. We refer to the (weakly) more congested lane as the *slow* lane (sometimes shortened as S) and denote its mass of traveler by s. Similarly, the (weakly) less congested lane is referred to as *fast* (sometimes shortened as F) and its mass of traveler is denoted by f.

Individuals are heterogeneous. Their type θ is assumed to be uniformly distributed in the $[0, 1]$ interval. A type-θ individual has the following utility function:

$$U(\theta) = \begin{cases} 0 & \text{when not traveling;} \\ B(\theta) - \theta g s & \text{when traveling in lane } S; \\ B(\theta) - \theta g f & \text{when traveling in lane } F. \end{cases} \quad (1)$$

The term $B(\theta)$ is the gross benefit from traveling, which depends only on the type θ. We let $B(\theta)$ be increasing and weakly concave in θ, so that $B'(\theta) > 0$ and $B''(\theta) \leq 0$. The terms $\theta g s$ and $\theta g f$ denote the disutility from congestion. They depend on: i) the mass of travelers in the same lane, either s or f; ii) a type-independent parameter $g > 0$ representing the common (across travelers) component of congestion disutility; iii) the type θ representing the idiosyncratic component of congestion disutility. Note that θ determines the travelers' value of travel and, at the same time, affects their cost of congestion. The assumption that both increase with θ is consistent with evidence that points at a positive relation between wage and value of time (see, for instance, Small, 2012).

[4]Two other papers analyze equilibria when drivers are non-atomistic (Silva et al. 2016, Lindsey, de Palma and Silva, 2019), while Simoni et al. (2019) use agent-based simulations to evaluate the impact of different congestion pricing and tolling strategies in the presence of AVs.

We assume that the common component of congestion disutility, g, is sufficiently low so that the net utility of travelers is increasing in θ for any s and f, i.e., $\frac{\partial U(\theta)}{\partial \theta} > 0$ when $U(\theta) > 0$. This low value of g may be consistent with the likely reduction in the disutility from congestion entailed by the use of AVs. A necessary and sufficient condition for this to occur is as follows:

Assumption 1. $g < B'(1)$.

To avoid the uninteresting case of some low θ-types never wanting to travel, we posit that the type-0 individual's utility from traveling is nonnegative:

Assumption 2. $U(0) = B(0) \geq 0$.

Assumptions 1 and 2 together imply that all individuals get nonnegative utility from traveling.

Individuals' identity. We assume that the mass of individuals is potentially composed of two different groups:

- *atomistic individuals*: when they travel, they drive AVs that do not belong to a fleet, and do not factor in their travel decisions the external congestion effect they impose on fellow travelers;
- *corporate individuals*: when they travel, they use the services of a fleet of AVs that belong to a monopolist.

We assume that the mass of corporate individuals is equal to μ, with $\mu \in [0, 1]$. The remaining mass $1-\mu$ is composed of atomistic individuals. The proportion μ of corporate travelers is then a measure of the degree of centralization. We let the composition of the mass of individuals be exogenous. This is because belonging to either group may depend on individual preferences or on long-run decisions, such as, for instance, the choice to drive an owned car or not, which are not modeled here. Moreover, we let the distribution of the two groups of individuals be independent of the type θ. That is, in any subinterval $[\theta, \theta + \epsilon]$ of the unit line, with $\epsilon > 0$,there is a fraction μ of corporate individuals and a fraction $1 - \mu$ of atomistic individuals.

Depending on the value of μ, we will analyze three possible compositions of the mass of individuals, meant to illustrate the stages of the transition from full decentralization to full centralization:

- *atomistic individuals only:* $\mu = 0$. This is the fully decentralized regime;[5]

[5]This scenario resembles the current traffic organization with traditional vehicles. The diffusion of navigation systems using real time information provides individuals with an easy solution to the informational and computation problem of choosing the individually optimal route.

- *atomistic and corporate individuals:* $\mu \in (0, 1)$. A mass $1 - \mu$ of individuals is atomistic and the remaining mass μ of individuals use the fleet of AVs managed by the monopolist. This is the partially decentralized regime. The larger is μ, the more advanced this process of centralization is;
- *corporate individuals only:* $\mu = 1$. All individuals use the fleet of AVs managed by the monopolist. This is the final stage of the process of centralization.

The game. We initially look at a two stage game. In the first stage, the monopolist chooses the fares to charge to corporate individuals; in the second stage, all individuals, both corporate and atomistic, simultaneously make their travel decisions. All players have full information on the entire game and the equilibrium concept we use is subgame perfect Nash equilibrium. When all individuals are atomistic, the first stage is irrelevant and we ignore it. On the other hand, in case of a strictly positive mass of corporate individuals, we assume that the monopolist cannot set different fares only on the lane used, but not on the individual's type θ, possibly because of a privacy protection regulation (Montes, Sand-Zantman and Valletti, 2018).

In Section 6, we add an initial stage to this game, in which a tax authority chooses a tax scheme. After this initial stage, the rest of the game unfolds as described before. In this game, we restrict our tax authority to charge unit (per travel) taxes, possibly different by lane but not by individual's identity.

Individuals' incentives. Both atomistic and corporate individuals choose whether or not to travel, and in which lane to do so, based on individual incentives. For a type θ, the decision depends on her utility $U(\theta)$ and on the fares and/or taxes (if any) she pays when traveling in the slow or the fast lane, denoted as $\sigma \geq 0$ and $\phi \geq 0$, respectively. A type-θ individual travels in the slow lane if and only her individual rationality (IR) constraint holds, that is

$$B(\theta) - \theta g s - \sigma \geq 0, \tag{2}$$

and if and only if her incentive compatibility (IC) constraint holds, that is

$$B(\theta) - \theta g s - \sigma \geq B(\theta) - \theta g f - \phi. \tag{3}$$

Similarly, she travels in the fast lane if and only if

$$B(\theta) - \theta g f - \phi \geq 0, \tag{4}$$

$$B(\theta) - \theta g f - \phi \geq B(\theta) - \theta g s - \sigma. \tag{5}$$

When at least one of the IR constraints holds for all individuals, they all travel so the market is fully covered (i.e., $s + f = 1$). Instead, if for some individuals neither IR constraints hold, such individuals prefer not to travel so the market is only partially covered (i.e., $s + f < 1$).

3 First best

We consider an utilitarian welfare maximizing social planner, who is perfectly informed and can decide which individuals travel and directly allocate those traveling to the two lanes. Social welfare is given by[6]

$$W \equiv \int_0^1 U(\theta) d\theta. \tag{6}$$

Welfare maximization requires to partition travelers in (at most) three groups. Some very low θ-types may not travel, while all the other θ's are partitioned into the two lanes, with the higher θ's traveling in the fast lane. Then, (6) may be rewritten as:

$$W' \equiv \int_{1-s-f}^{1-f} (B(\theta) - \theta gs)\, d\theta + \int_{1-f}^{1} (B(\theta) - \theta gf)\, d\theta \tag{7}$$

where the first integral gives the aggregate utility of types traveling in the slow lane, and the second integral gives the aggregate utility of types traveling in the fast lane.[7] The planner's problem may be written as follows:

$$\begin{aligned} &\max_{\substack{s\geq 0,\\ f\geq 0}} \; W' \\ &\text{s.t.} \quad s + f \leq 1. \end{aligned} \tag{8}$$

An interior solution satisfies the following first order conditions:

$$B\left(1 - s_{FB} - f_{FB}\right) - 2gs_{FB}\left(1 - f_{FB} - \frac{3}{4}s_{FB}\right) = 0, \tag{9}$$

$$B\left(1 - s_{FB} - f_{FB}\right) + g\left(s_{FB}^2 + \frac{3}{2}f_{FB}^2 - 2f_{FB}\right) = 0, \tag{10}$$

where the subscript FB is a mnemonic for equilibrium variables referred to First Best. The solution to the planner's problem is characterized in the following Proposition.

Proposition 1. *Let s_{FB} and f_{FB} denote the solutions to problem (8) when the market is partially covered, i.e., $s_{FB} + f_{FB} < 1$. Also, let $\bar{s}_{FB}$ and $\bar{f}_{FB}$ denote the solutions to problem (8) when the market is fully covered, i.e., $\bar{s}_{FB} + \bar{f}_{FB} = 1$. Finally, let*

$$g_{FB} \equiv \frac{36\, B(0)}{4 + \sqrt{7}} \cong 5.4179 \times B(0). \tag{11}$$

[6] Note that fares, as well as taxes, do not appear in the social welfare function even when travelers and/or the monopolist pay them, because they are simply transfers from travelers to the monopolist (in the case of fares) or from the monopolist to the (benevolent) tax authority (in the case of taxes).

[7] See the proof of Proposition 1.

Then, in equilibrium

- *when* $g > g_{FB}$*, the market is partially covered, and travelers with* $\theta \in [1 - s_{FB} - f_{FB}, 1 - f_{FB}]$ *are allocated to the slow lane and those with* $\theta \in [1 - f_{FB}, 1]$ *are allocated to the fast lane, where* s_{FB} *and* f_{FB} *satisfy*

$$f_{FB}(s_{FB}) = \frac{1}{3}\left(2\,(1 + s_{FB}) - \sqrt{7s_{FB}^2 - 4s_{FB} + 4}\right); \tag{12}$$

- *when* $g \leq g_{FB}$*, the market is fully covered, and travelers with* $\theta \in [0, \bar{s}_{FB}]$ *are allocated to the slow lane and those with* $\theta \in [\bar{s}_{FB}, 1]$ *are allocated to the fast lane, where* $\bar{s}_{FB}$ *and* $\bar{f}_{FB}$ *are equal to*

$$\bar{s}_{FB} = \frac{1}{2} + \frac{\sqrt{7} - 2}{6} \cong 0.6076 \quad \textit{and} \quad \bar{f}_{FB} = \frac{1}{2} - \frac{\sqrt{7} - 2}{6} \cong 0.3924. \tag{13}$$

This Proposition characterizes the socially optimal allocation of individuals. The planner may exclude the individuals with the lowest benefit from traveling, so that the market is not fully covered. This occurs when the utility from traveling in the slow lane enjoyed by the type-0 agent, $B(0)$, is lower than the increase in the aggregate congestion costs this individual imposes on all fellow travelers in the slow lane.[8] All the remaining travelers are sorted in the two lanes. A mass of travelers equal to f_{FB} (or $\bar{f}_{FB}$, in the case of full coverage) is allocated to the fast lane and a mass of travelers equal to s_{FB} (or $\bar{s}_{FB}$, in the case of full coverage) to the slow lane. Intuitively, travelers allocated to the fast lane are those with the highest θ.

In the case of partial coverage, we do not fully characterize the solution to the planner's problem, but only provide the optimal choice of f_{FB} as a function of s_{FB}, given in equation (12). This relationship is illustrated in Figure 1 by the green increasing line, and allows us to discuss the main features of the partial coverage solution. First, the planner finds it optimal to differentiate across lanes. Travelers with relatively high θ, who suffer the most from congestion, are assigned to the fast, less congested, lane. As a result, the green line in the Figure always lies below the 45° line. Second, by implicit differentiation of the FOCs in (9) and (10), we find that $\frac{\partial s_{FB}}{\partial g} < 0$ and $\frac{\partial f_{FB}}{\partial g} < 0$. This reflects the intuition that a larger common component of the disutility from congestion, g, is associated to a lower mass of travelers in each lane, and, therefore, to a lower market coverage. Full market coverage is illustrated in Figure 1 by the green solid circle located at the intersection between the green line illustrating equation (12) and the constraint $s + f = 1$.

[8] Condition $g > g_{FB}$ can be rearranged as $B(0) < g\frac{(\bar{s}_{FB})^2}{2}$, where the RHS is indeed the aggregate marginal congestion cost for those traveling in the slow lane.

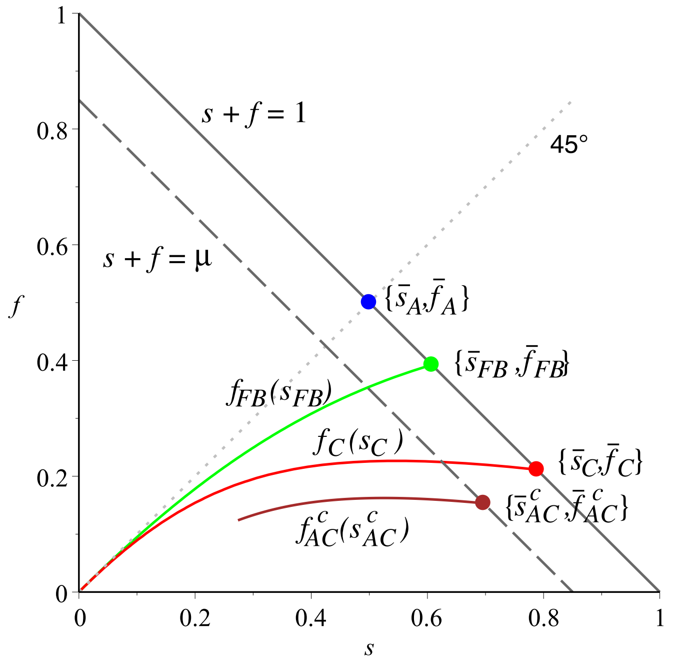

Figure 1: Equilibrium outcomes.
The equilibrium solutions with full decentralization (blue dot, $\{\bar{s}_A, \bar{f}_A\}$); the full coverage solutions in the first best (green dot, $\{\bar{s}_{FB}, \bar{f}_{FB}\}$), with full centralization (red dot, $\{\bar{s}_C, \bar{f}_C\}$) and with partial centralization (brown dot, $\{\bar{s}^c_{AC}, \bar{f}^c_{AC}\}$); and the relationships between optimal s and f in the first best (green line, $f_{FB}(s_{FB})$), and between s^c_{AC} and f^c_{AC} with full centralization (red line, $f_C(s_C)$) and with partial centralization (brown line, $f^c_{AC}(s^c_{AC})$); $f^c_{AC}(s^c_{AC})$ is defined for values of s^c_{AC} above a threshold, since when s^c_{AC} is too low the equilibrium involves no differentiation (mass of atomistic travelers equal to $\mu = 0.85$).

4 Full coverage: equilibrium analysis

In our framework, the optimal management of congestion may require both to optimally ration travel and to optimally sort travelers across lanes. As discussed in the introduction, given the heterogeneity in individuals' disutility from congestion, efficient sorting may have a very significant impact on reducing the aggregate congestion costs. In addition, AVs, along with other recent developments in urban transport, are bound to dramatically reduce the cost of the sorting technology, thereby making sorting a viable and increasingly important tool to manage congestion.

We thus start by focusing on the allocative problem of sorting travelers between the fast and the slow lane - and on the resulting differentiation in the level of congestion across lanes -, when there is no rationing. To this aim, we introduce an assumption that gives a sufficient condition for full coverage to occur in the first best and in equilibrium for all degrees of centralization.[9]

[9] See the appendix for the derivation of this result.

Assumption 3. $\frac{B'(0)}{B(0)} \leq 1$.

Assumption 3 states that at $\theta = 0$ the semielasticity of the gross benefit function $B(\theta)$ is weakly lower than 1. In other words, the gross benefit from traveling for type-0 individual is large relative to its increase as θ rises.

4.1 Full decentralization

We study a fully decentralized regime populated by atomistic individuals only. Each individual uses an AV that does not belong to a fleet, does not pay a fee, and chooses the lane giving her the highest utility, ignoring the effect of this choice on other individuals. Since the monopolist has no role in this regime, the stage of the game in which it sets its fares is muted.

The next Proposition shows that in any equilibrium of this game travelers split equally across the two lanes, so that, differently than the social optimum, there is no differentiation across lanes. Here and in the rest of the paper, equilibrium variables referred to the case of a market with atomistic individuals only have the subscript A, a mnemonic for Atomistic.

Proposition 2. *Assume Assumption 3 holds and all individuals are atomistic. Let $\bar{s}_A$ and $\bar{f}_A$ be the equilibrium masses of travelers in the two lanes in full coverage. Then, any allocation of travelers such that*

$$\bar{s}_A = \bar{f}_A = \frac{1}{2} \tag{14}$$

is an equilibrium.

This Proposition illustrates that atomistic travelers split equally across the two lanes, which therefore feature the same level of congestion. There are infinite payoff equivalent allocations of travelers that satisfy the condition (14), and they all are equilibria of the game.

To get the intuition behind the equal level of congestion across lanes, argue by contradiction and suppose that, for instance, $s_A > f_A$. This cannot be an equilibrium outcome because any traveler in the more congested lane, irrespective of her type θ, would prefer to switch to the less congested lane where she would enjoy a higher net utility.

4.2 Partial centralization

We now analyze the case of partial centralization, where a mass μ of individuals are corporate and use the fleet of AVs managed by the monopolist, and the remaining mass $1-\mu$ are atomistic. Corporate travelers have to pay a lane-specific fare to the monopolist in exchange for the service. The monopolist sets uniform fares within lanes: a fare p

for the slow lane and a fare P for the fast lane. Fares are used by the monopolist as the only instrument to direct corporate individuals to the two lanes. Here and in the rest of the paper, equilibrium variables referred to this market have a subscript AC, a mnemonic for Atomistic and Corporate.

We first establish some conditions on the equilibrium allocations of travelers. In particular, we derive conditions under which the no differentiation equilibrium in the case of full decentralization may extend to the situation of partial centralization. Denote with s^c and f^c the mass of corporate travelers in the slow and fast lane, respectively. Similarly, denote with s^a and f^a the mass of atomistic travelers in the slow and fast lane, respectively. Thus, the total mass of travelers in the two lanes is given by $s = s^a + s^c$ and $f = f^a + f^c$.

Lemma 1. *Under any degree of market centralization $\mu > 0$, for any equilibrium candidate pair of masses of corporate travelers s^c and f^c, with $s^c \geq f^c$, in equilibrium:*

- *if $1 - \mu \geq s^c - f^c$, there cannot be differentiation across lanes, and*

$$s = f = \frac{s^c + f^c + s^a + f^a}{2}; \tag{15}$$

- *if $1-\mu < s^c - f^c$, there must be differentiation across lanes, all atomistic individuals must travel in the fast lane, and*

$$\begin{aligned} s &= s^c, \\ f &= f^c + f^a = f^c + (1 - \mu) . \end{aligned} \tag{16}$$

When $1 - \mu \geq s^c - f^c$, the mass of atomistic individuals is large relative to the difference in the allocation of corporate travelers across the two lanes. Given the allocation of corporate travelers, if all atomistic travelers chose the same lane, its congestion level would exceed that of the other lane. This clearly cannot be an equilibrium since any atomistic traveler in the more congested lane would then benefit from switching to the less congested one. As a result, in an equilibrium, atomistic travelers must split between the two lanes so as to equalize the total mass of travelers in each lane, $s = f$. When, instead, $1 - \mu \leq s^c - f^c$, the mass of atomistic individuals is small relative to the difference in the masses of corporate travelers allocated to the two lanes. Given the allocation of corporate travelers, when all atomistic travelers choose the fast lane, then congestion is still less than in the slow lane. Hence, in any equilibrium, there must be differentiation across lanes, and $f = f^c + (1 - \mu) < s$.

We now characterize the equilibrium, starting from the case where μ is relatively low. The equilibrium allocation of travelers must satisfy (15). Since $s = f$, the monopolist must charge the same fares for the two lanes, i.e., $p = P$. The IR constraints (2) and (4) collapse to a single one, which, because of full coverage, can be written as

$$p \leq B(0) . \tag{17}$$

Since all corporate travelers are charged the same fare p, the monopolist profits are affected by their total mass, and not by their allocation across lanes. Hence, the monopolist problem may be written as

$$\max_{\substack{s^c \geq 0, \\ f^c \geq 0}} B(0)(s^c + f^c), \quad (18)$$
$$\text{s.t.} \quad s^c + f^c = \mu,$$

where we take constraint (17) to be binding, as it must be in equilibrium.

Next, focus on the case where μ is relatively large, so that the equilibrium allocation of travelers must satisfy (16). In setting the fares to allocate corporate individuals - p for the slow lane and P for the fast lane - the monopolist faces constraints from (2) to (5). By a standard argument, only the IR constraint (2) for the marginal corporate traveler in the slow lane (type-0, given our focus on full coverage), and the IC constraint (5) for the corporate traveler indifferent between the slow and the fast lane (with type θ equal to $\theta = 1 - \frac{f^c}{\mu}$) are binding. These constraints may be written as follows,

$$p = B(0),$$
$$P = p + g\left(1 - \frac{f^c}{\mu}\right)(s - f).$$

The fare p is set to make the corporate individual with type-0 just willing to travel in the slow lane. Instead, P clearly illustrates the trade-off the monopolist faces in choosing the profit maximizing degree of differentiation between lanes. On the one hand, a large difference in the mass of travelers across the two lanes, $s - f$, entails a large extra-fee paid by travelers in the fast lane, hence a high mark-up. On the other hand, a large differentiation implies that the mass of travelers in the fast lane is small.

Incorporating the two constraints in the problem faced by the monopolist, this is given by

$$\max_{\substack{s^c \geq 0, \\ f^c \geq 0}} B(0)(s^c + f^c) + g\left(1 - \frac{f^c}{\mu}\right)(s - f) f^c, \quad (19)$$
$$\text{s.t.} \quad s^c + f^c = \mu.$$

The next Proposition characterizes the subgame perfect Nash equilibrium of the game.

Proposition 3. *Assume Assumption 3 holds and partial centralization. Let $\bar{s}^c_{AC}$ and $\bar{f}^c_{AC}$ denote the equilibrium mass of corporate travelers in full coverage. Similarly, let $\bar{s}^a_{AC}$ and $\bar{f}^a_{AC}$ denote the equilibrium mass of atomistic travelers in full coverage. Then,*

- *when* $\mu \in \left(0, \frac{1}{2}\right]$, *an equilibrium is any allocation of travelers such that*

$$\begin{gathered}\bar{s}^c_{AC} + \bar{s}^a_{AC} = \bar{f}^c_{AC} + \bar{f}^a_{AC} = \tfrac{1}{2}, \\ \bar{s}^c_{AC} + \bar{f}^c_{AC} = \mu, \\ \bar{s}^a_{AC} + \bar{f}^a_{AC} = 1 - \mu; \end{gathered} \tag{20}$$

- *when* $\mu \in \left(\frac{1}{2}, 1\right)$, *the equilibrium is such that all* $1 - \mu$ *atomistic travelers together with corporate travelers with* $\theta \in \left[\frac{\bar{s}^c_{AC}}{\mu}, 1\right]$ *are allocated to the fast lane, and corporate travelers with* $\theta \in \left[0, \frac{\bar{s}^c_{AC}}{\mu}\right]$ *are allocated to the slow lane, so that*

$$\begin{gathered}\bar{s}^c_{AC} = \tfrac{1}{2} + \tfrac{\sqrt{4\mu^2 - 2\mu + 1} - 2(1-\mu)}{6}, \\ \bar{f}^c_{AC} = \tfrac{1}{2} - \tfrac{\sqrt{4\mu^2 - 2\mu + 1} + 4(1-\mu)}{6}, \\ \bar{s}^a_{AC} = 0, \\ \bar{f}^a_{AC} = 1 - \mu. \end{gathered} \tag{21}$$

When the mass of corporate individuals is small enough, multiple equilibrium allocations exist. In all of these, the two lanes are equally congested. No matter how the monopolist sorts corporate individuals, atomistic travelers allocate themselves across the two lanes, so as to make them equally congested. This induces the monopolist to charge the same fare across lanes. The equilibrium price is uniquely determined given by the IR of the lowest θ, i.e. $p = B(0)$, because of full coverage.

When instead the mass μ of corporate individuals is sufficiently large, the equilibrium allocation of travelers is unique. The mass of atomistic travelers is not sufficiently large to bridge the congestion gap across lanes as determined by the prices set by the monopolist to sort corporate travelers. This means that the monopolist can price discriminate across lanes. Taking into account that all atomistic individuals will use the fast lane, the monopolist allocates a large enough mass of corporate travelers to the slow lane and a relatively small mass to the fast lane. This solution is illustrated by the brown dot in Figure 1.

In conclusion, notice that the threshold level of μ that distinguishes between the two types of equilibria, with and without lane differentiation, is equal to $\frac{1}{2}$.[10] Whenever the mass of corporate individuals is below this threshold, there are no differential allocations of corporate travelers in the two lanes that atomistic travelers cannot undo. When instead μ is above $\frac{1}{2}$, not only it is feasible for the monopolist to differentiate, but is also strictly preferable. Indeed, when there is no differentiation across lanes, the monopolist charges all corporate travelers the same price $p = B(0)$. If, instead, differentiation arises, the price for corporate travelers in the slow lane remains unaffected, while the price for corporate travelers in the fast lane is higher.

[10] To see this, we plug $\bar{s}^c_{AC}$ and $\bar{f}^c_{AC}$ as in (21) into conditions $1 - \mu > (<)\, s^c - f^c$ and solve by μ; this yields $\mu < (>)\, \frac{1}{2}$.

4.3 Full centralization

We now consider a fully centralized market, with $\mu = 1$, where all AVs are part of a fleet managed by the monopolist. As in the previous case, in setting the two fares, p for the slow lane and P for the fast lane, the monopolist faces constraints from (2) to (5). By a standard argument, only the IR constraint (2) for the marginal traveler in the slow lane and the IC constraint (5) for the traveler indifferent between the slow and the fast lane are binding. Under full coverage, these constraints read as

$$p = B(0), \tag{22}$$

$$P = p + g(1-f)(s-f). \tag{23}$$

The monopolist problem can therefore be written as

$$\max_{\substack{s \geq 0, \\ f \geq 0}} B(0)(s+f) + [g(1-f)(s-f)]\, f, \qquad \text{s.t.} \quad s + f = 1. \tag{24}$$

In the next Proposition, we show how the monopolist allocates travelers when it fully covers the market. Here and in the rest of the paper, equilibrium variables referred to this market have a subscript C, a mnemonic for Corporate.

Proposition 4. *Assume Assumption 3 holds and full centralization. Let $\bar{s}_C$ and $\bar{f}_C$ denote the solutions to the monopolist problem (24) under full coverage. Then, the equilibrium is such that corporate travelers with $\theta \in [0, \bar{s}_C]$ are allocated to the slow lane and those with $\theta \in [\bar{s}_C, 1]$ are allocated to the fast lane, where*

$$\begin{aligned} \bar{s}_C &= \tfrac{1}{2} + \tfrac{\sqrt{3}}{6} \cong 0.7887, \\ \bar{f}_C &= \tfrac{1}{2} - \tfrac{\sqrt{3}}{6} \cong 0.2113. \end{aligned} \tag{25}$$

The Proposition characterizes the profit maximizing allocation of travelers. High-θ travelers, in a mass equal to $\overline{f}_C$, are sorted into the fast lane and low-θ travelers, in a mass equal to $\bar{s}_C$, into the slow lane. This solution is illustrated in Figure 1 by the red solid circle. The figure shows that the outcome of a centralized market is overdifferentiation across lanes. Too few travelers travel in the fast lane as compared to the socially optimal level, $\bar{f}_C < \bar{f}_{FB}$, and too many in the slow lane, $\bar{s}_C > \bar{s}_{FB}$. This result, due to the IC constraint (23) that allows to charge an increasingly high fare in the fast lane the larger the congestion differential across lanes, is reminiscent of Mussa and Rosen (1978).

4.4 Welfare analysis

In this section, we investigate the welfare effects of the transition from a decentralized to a centralized regime. At full coverage, where $s = 1 - f$, social welfare as in (6) may be rewritten as

$$\bar{W} = \int_0^1 B(\theta)d\theta - \mu \left[\int_0^{1-\frac{f^c}{\mu}} \theta g\left(s^c + s^a\right) d\theta + \int_{1-\frac{f^c}{\mu}}^1 \theta g\left(f^c + f^a\right) d\theta \right] + \qquad (26)$$
$$- (1-\mu) \left[\int_0^{1-\frac{f^a}{1-\mu}} \theta g\left(s^a + s^c\right) d\theta + \int_{1-\frac{f^a}{1-\mu}}^1 \theta g\left(f^a + f^c\right) d\theta \right]$$

where we use the “bar” notation as in the rest of this section to denote full coverage. The first term illustrates the aggregate benefit from traveling, enjoyed by all travelers thanks to full coverage. The expressions in square brackets for each group of travelers (corporate and atomistic) are the sum of the congestion disutility suffered by those traveling in the slow (first term) and fast (second term) lane.

We evaluate (26) at the equilibrium for different degrees of centralization. It may seem intuitive that, as μ increases, welfare is positively affected by the increasing ability of the monopolist to internalize the congestion externality. We will show that this intuition is incomplete, as it ignores two welfare-reducing distortions that emerge with centralization. Our results are summarized in the following Proposition.

Proposition 5. *Assume Assumption 3 holds. Then, in equilibrium,*

- *$\bar{W}$ does not vary with μ when $\mu \in \left[0, \frac{1}{2}\right]$ and is strictly decreasing in μ when $\mu \in \left(\frac{1}{2}, 1\right]$;*
- *$\bar{W}$ is strictly below first best for all values of μ.*

The Proposition, also illustrated in Figure 2, shows that welfare always lies below the welfare level achieved by the social planner (which does not depend on μ). It also illustrates that the increasing ability of the monopolist to internalize the congestion externality when the level of market centralization μ increases does not reflect into a welfare improvement. To the contrary, welfare is decreasing in μ (strictly decreasing for μ above $\frac{1}{2}$).

The negative welfare effect of centralization derives from two distortions that emerge as the share of corporate individuals increases above $\frac{1}{2}$. The first distortion involves the level of differentiation across lanes. Under full decentralization, the two lanes have exactly the same number of travelers, leading to suboptimal differentiation vis-à-vis the socially optimal level. Instead, under full decentralization, there is, to the contrary, excess differentiation in congestion across lanes with respect to social optimum. The excess differentiation in monopoly turns out to reduce welfare more than the suboptimal differentiation under full decentralization for two reasons. It involves more travelers (i.e., $\bar{s}_C - \bar{s}_{FB} > \bar{s}_{FB} - \bar{s}_A$) and travelers with higher θ's, who are more bothered by congestion.

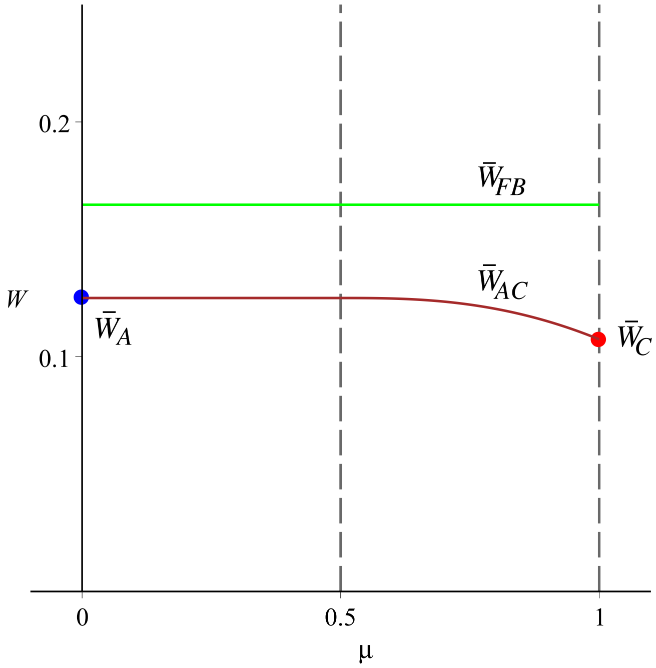

Figure 2: Equilibrium welfare.
Equilibrium welfare under full coverage in the first best ($\bar{W}_{FB}$, green line), with full decentralization ($\bar{W}_A$, blue dot), with partial centralization ($\bar{W}_{AC}$, brown line) and with full centralization ($\bar{W}_C$, red dot); $g = \frac{1}{2}$ and $\int_0^1 B(\theta)d\theta = \frac{1}{4}$.

The second distortion relates to the identity of individuals traveling in each lane, rather than to their mass. When $\mu \in \left(\frac{1}{2}, 1\right)$, Proposition 7 states that all the atomistic individuals travel in the fast lane, including those with a relative low value of traveling and disutility from congestion. By contrast, only high-θ corporate individuals travel in the fast lane. It follows that there are corporate travelers with a relatively high θ, say θ'', traveling in the slow lane, while some atomistic travelers a relatively low θ, say $\theta' < \theta''$, travel in the fast lane. In particular, there exists a level of μ (i.e., $\mu = \frac{2\sqrt{7}+1}{9} \cong 0.6991$), such that the mass of travelers in the slow and in the fast lane exactly replicate the mass of travelers in the two lanes chosen by the social planner under full coverage. That notwithstanding, welfare under $\mu = \frac{2\sqrt{7}+1}{9}$ is below that achieved by the social planner, because of the misallocation of travelers described above. This distortion emphasizes that, given travelers' heterogeneity, welfare is maximized not only by generating the appropriate level of congestion in the two lanes, but also by making sure that each traveler is correctly allocated.

We conclude this discussion with some remarks on the distributional effects in terms of individuals' utility net of prices across different regimes. As compared to the social optimum, a fully decentralized regime makes travelers with relatively low θ (i.e., $\theta \in [0, \bar{s}_{FB}]$) better off and travelers with relatively high θ (i.e., $\theta \in [\bar{s}_{FB}, 1]$) worse off. On the one hand, travelers with low θ's, who would travel in the slow lane under a central planner, travel in a less congested lane under full decentralization. On the other hand, travelers with high θ's, who would travel in the fast lane under a central planner, end up traveling in a more congested lane under full decentralization.

It is also possible to assess the distributional effects (net of prices) of an increase in the degree of centralization. When $\mu \leq \frac{1}{2}$, equilibrium allocation does not change and therefore there is no distributional effects. When $\mu \geq \frac{1}{2}$, an increase in μ increases the level of differentiation across lanes. Therefore, a larger μ makes all atomistic travelers and corporate travelers traveling in the fast lane better off. To the contrary, corporate travelers traveling in the slow lane are worse off.

5 Partial coverage: equilibrium analysis

In Section 4, we focused on sorting. We showed the welfare effects of moving from centralization to decentralization when the aggregate amount of travelers does not change, but their allocation across lanes does, with an effect on aggregate congestion costs, and, as a result, on welfare. In this section, we abandon Assumption 3 and allow the aggregate amount of travelers to vary. As a result, we can also consider the perhaps more intuitive strategy to deal with congestion, which consists in restricting aggregate output.

5.1 Full decentralization

Full coverage in the fully decentralized regime is implied by Assumptions 1 and 2 only and does not depend on Assumption 3 being met. Hence, the analysis contained in Section 4.1 applies here too. All atomistic individuals travel and the split equally across the two lanes, so that $s_A = f_A = \bar{s}_A = \bar{f}_A = \frac{1}{2}$, where s_A and f_A denote the equilibrium mass of atomistic travelers when Assumption 3 is not met.

5.2 Partial centralization

We now look at the case of partial centralization. We start our analysis by noting that Lemma 1 does not depend on Assumption 3. Therefore, this Lemma applies also here, when Assumption 3 is relaxed.

Assume first that μ is small enough, so that the equilibrium allocation of travelers must satisfy (15), with no differentiation across lanes. This implies that the monopolist charges the same price across lanes, so that $p = P$. Using the same argument as in Section 4.2, the only constraint faced by the monopolist is the IR constraint for the lowest corporate θ-type traveling in either lane. When binding, this may be written as

$$p = B\left(1 - \frac{s^c + f^c}{\mu}\right) - \left(1 - \frac{s^c + f^c}{\mu}\right) g\, \frac{s^c + f^c + s^a + f^a}{2}, \tag{27}$$

where $\theta = 1 - \frac{s^c+f^c}{\mu}$ is the type of the corporate traveler that is indifferent between not traveling and traveling, and $\frac{s^c+f^c+s^a+f^a}{2}$ is the mass of (atomistic and corporate) travelers traveling in either lanes.

Since the monopolist charges the same fare to all corporate travelers, its profit is affected by their total mass only. This implies that $\pi = p \times (s^c + f^c)$. Denote by c the total mass of traveling corporate individuals, so that $c \equiv s^c + f^c$. Using (27), for any s^c and f^c such that $1 - \mu < s^c - f^c$, the monopolist problem can then be written as

$$\max_{c \geq 0} \left[B\left(1 - \frac{c}{\mu}\right) - \left(1 - \frac{c}{\mu}\right) g\, \frac{c + s^a + f^a}{2} \right] c, \tag{28}$$
$$\text{s.t.} \quad c \leq \mu.$$

We denote by c_{AC} the solution to (28). When it is interior, this solution is implicitly defined by the following FOC:

$$\frac{3g\left(c_{AC}\right)^2}{2\mu} + c_{AC}\frac{g\left(1-2\mu\right) - B'\left(1 - \frac{c_{AC}}{\mu}\right)}{\mu} + \frac{2B\left(1 - \frac{c_{AC}}{\mu}\right) - g\left(1-\mu\right)}{2} = 0. \tag{29}$$

We now turn to the analysis of the case of sufficiently large μ, so that the equilibrium allocation of travelers must satisfy (16), with differentiation across lanes. Using the same argument as in Section 4.2, for any s^c and f^c such that $1 - \mu < s^c - f^c$, the constraints faced by the monopolist may be written as

$$p = B\left(1 - \frac{s^c + f^c}{\mu}\right) - \left(1 - \frac{s^c + f^c}{\mu}\right) g\left(s^c + s^a\right), \tag{30}$$

$$P = p + g\left(1 - \frac{f^c}{\mu}\right)\left[s^c + s^a - \left(f^c + f^a\right)\right], \tag{31}$$

where $\theta = 1 - \frac{f^c}{\mu}$ is the type of the corporate traveler that is indifferent between traveling in the fast and the slow lane, $s^c + s^a$ is the mass of (atomistic and corporate) travelers traveling in the slow lane, and $f^c + f^a$ that in the fast lane. The monopolist problem may then be written as

$$\max_{\substack{s^c \geq 0,\\ f^c \geq 0}} \left[B\left(1 - \frac{s^c + f^c}{\mu}\right) - \left(1 - \frac{s^c + f^c}{\mu}\right) g\left(s^c + s^a\right)\right]\left(s^c + f^c\right) +$$
$$+ g\left(1 - \frac{f^c}{\mu}\right)\left[s^c + s^a - \left(f^c + f^a\right)\right] f^c, \tag{32}$$
$$\text{s.t.} \quad s^c + f^c \leq \mu.$$

Denote by s^c_{AC} and f^c_{AC} the solutions to this problem when they are interior. Exploiting $s^a = 0$ and $f^a = 1 - \mu$, they are implicitly defined by the following FOCs:

$$B\left(1 - \frac{s^c_{AC} + f^c_{AC}}{\mu}\right) - \frac{B'\left(1 - \frac{s^c_{AC}+f^c_{AC}}{\mu}\right) \times \left(s^c_{AC} + f^c_{AC}\right)}{\mu} + \\ + \frac{g s^c_{AC}\left[3 s^c_{AC} + 4 f^c_{AC} - 2\mu\right]}{\mu} = 0, \tag{33}$$

$$B\left(1 - \frac{s^c_{AC} + f^c_{AC}}{\mu}\right) - \frac{B'\left(1 - \frac{s^c_{AC}+f^c_{AC}}{\mu}\right) \times \left(s^c_{AC} + f^c_{AC}\right)}{\mu} + \\ + \frac{g\left[2\left(s^c_{AC}\right)^2 + f^c_{AC}\left(4\left(1-\mu\right) + 3 f^c_{AC} - 2\right) - \left(1-\mu\right)\mu\right]}{\mu} = 0. \tag{34}$$

Before characterizing the equilibrium, we provide the following Lemma:

Lemma 2. *Assume partial centralization. Let s^c_{AC} and f^c_{AC} denote the equilibrium mass of corporate travelers under partial coverage. Let μ' be the value of μ such that $1 - \mu = s^c_{AC} - f^c_{AC}$. Then, there is differentiation across lanes if and only if $\mu \in (\mu', 1]$, where $\mu' \in \left(\frac{1}{2}, 1\right)$.*

This Lemma characterizes the threshold level μ', such that for any $\mu > \mu'$, lanes are differentiated in equilibrium. This value is defined in terms of the equilibrium mass of corporate travelers in the slow and in the fast lane when differentiation arises, s^c_{AC} and f^c_{AC}. While the threshold level is $\frac{1}{2}$ under full coverage, it is strictly above $\frac{1}{2}$ in case of partial coverage. The intuition is pretty straightforward. When the monopolist prefers to dispatch only a portion of its customers rather than all of them, it takes a smaller mass of atomistic travelers to equalize the total mass of travelers in each lane, $s = f$.

We are now in the position to illustrate the subgame perfect Nash equilibrium.

Proposition 6. *Assume partial centralization. Let s^a_{AC} and f^a_{AC} denote the equilibrium mass of atomistic travelers under partial coverage. Then,*

- *full coverage occurs when*

$$g \geq g_{AC} \equiv K\left(\mu\right)\left[B'\left(0\right) - B\left(0\right)\right], \tag{35}$$

where

$$K\left(\mu\right) \equiv \begin{cases} 2 & \text{when } \mu \in \left(0, \mu'\right), \\ \frac{18\mu}{8\mu^2 + 5\mu - 1 + \left(4\mu - 1\right)\sqrt{4\mu^2 - 2\mu + 1}} & \text{when } \mu \in [\mu', 1); \end{cases} \tag{36}$$

- *when $\mu \in (0, \mu']$ and $g < g_{AC}$, an equilibrium is any allocation of travelers such that*

$$\begin{aligned} s^c_{AC} + s^a_{AC} = f^c_{AC} + f^a_{AC} &= \tfrac{c_{AC} + 1 - \mu}{2}, \\ s^c_{AC} + f^c_{AC} &= c_{AC}, \\ s^a_{AC} + f^a_{AC} &= 1 - \mu, \end{aligned} \tag{37}$$

where $c_{AC} < \mu$ is implicitly defined in (29);

- *when $\mu \in (\mu', 1)$ and $g < g_{AC}$, the equilibrium is such that corporate individuals with $\theta \in \left[0, 1 - \frac{s^c_{AC} + f^c_{AC}}{\mu}\right]$ do not travel, corporate individuals with $\theta \in \left[1 - \frac{s^c_{AC} + f^c_{AC}}{\mu}, 1 - \frac{f^c_{AC}}{\mu}\right]$ are allocated in the slow lane, and all atomistic individuals together with corporate individuals with $\theta \in \left[1 - \frac{f^c_{AC}}{\mu}, 1\right]$ are allocated to the fast lane, where*

$$\begin{aligned} s^a_{AC} &= 0, \\ f^a_{AC} &= 1 - \mu, \end{aligned} \tag{38}$$

and f^c_{AC} and s^c_{AC} are such that $f^c_{AC} + s^c_{AC} < \mu$ and satisfy

$$f^c_{AC}(s^c_{AC}) = \frac{1}{3}\left(2\mu - 1 + 2s^c_{AC} - \sqrt{7\left(s^c_{AC}\right)^2 - 2s^c_{AC}(2-\mu) + 1 - \mu + \mu^2}\right). \tag{39}$$

The Proposition first illustrates the condition under which full/partial coverage occurs. Equation (35) shows a stark contrast with the first best full coverage condition in (11). Differently than a social planner, the monopolist dispatches all travelers when the congestion cost g is sufficiently high. To get the intuition for this perhaps counterintuitive result, consider that, as g gets larger, the travelers' utility function $U(\theta)$ gets flatter in θ, because the higher willingness to pay of higher θ types is increasingly compensated by their congestion disutility. This affects the traditional price/quantity tradeoff by providing the monopolist with a greater incentive to increase market coverage. The monopolist's full coverage condition depends also on $B'(0)$, with the same logic as above; when $B'(0)$ is small, $U(\theta)$ gets flatter in θ, which makes quantity more sensitive to price.

The Proposition also characterizes the equilibrium allocation of travelers, which depends on the relative masses of atomistic and corporate travelers. These allocations mirror closely those in the case of full coverage. When the mass of corporate travelers is relatively small, there are multiple equilibria. Corporate travelers are indifferent between the two lanes, because they are charged the same price due to the fact that atomistic travelers distribute between the two lanes to equalize the total mass of travelers in each lane. Any allocation of the two types of travelers which results in identical lanes is an equilibrium and is also payoff-equivalent. When instead the mass of corporate travelers is relatively large, the monopolist allocates travelers in the two lanes with masses sufficiently far apart so that atomistic travelers cannot equalize them. Atomistic travelers then all travel in the less congested lane. At this equilibrium with no differentiation, the relationship between f^c_{AC} and s^c_{AC} in (39) is depicted by the brown line in Figure 1.

5.3 Full centralization

We now consider a fully centralized regime. In setting the two fares, p for the slow lane and P for the fast lane, the monopolist, as in Section 4.3, faces the IR constraint for the lowest type traveling and the IC constraint for the type indifferent between traveling in the slow or fast lane. When binding, these constraints now read as

$$p = B(1 - s - f) - (1 - s - f)\, gs, \tag{40}$$

$$P = p + g(1 - f)(s - f). \tag{41}$$

The monopolist problem can therefore be written as

$$\max_{\substack{s \geq 0, \\ f \geq 0}} \; (B(1 - s - f) - (1 - s - f)\, gs)(s + f) + [g(1 - f)(s - f)]\, f, \tag{42}$$
$$\text{s.t.} \;\; s + f \leq 1.$$

Let s_C and f_C be the solutions to this problem when they are interior. These are implicitly defined by the following FOCs

$$B(1 - s_C - f_C) - B'(1 - s_C - f_C)(s_C + f_C) + \tag{43}$$
$$- gs_C(2 - 4f_C - 3s_C) = 0,$$
$$B(1 - s_C - f_C) - B'(1 - s_C - f_C)(s_C + f_C) + \tag{44}$$
$$g\left[3(f_C)^2 - 2f_C + 2(s_C)^2\right] = 0.$$

The subgame perfect Nash equilibrium is illustrated in the following Proposition.

Proposition 7. *Assume full centralization. Let s_C and f_C denote the solutions to the monopolist problem (42) under partial coverage. Then,*

- *full coverage occurs when*

$$g \geq g_C \equiv \frac{6}{4 + \sqrt{3}}\left[B'(0) - B(0)\right]; \tag{45}$$

- *when $g < g_C$, the equilibrium is such that corporate travelers with $\theta \in [0, 1 - s_C - f_C]$ do not travel, those with $\theta \in [1 - s_C - f_C, 1 - f_C]$ are allocated in the slow lane and those with $\theta \in [1 - f_C, 1]$ are allocated in the fast lane, where s_C and f_C satisfy $s_C + f_C < 1$ and*

$$f_C(s_C) = \frac{1}{3}\left(1 + 2s_C - \sqrt{7s_C^2 - 2s_C + 1}\right); \tag{46}$$

In equilibrium, the masses of travelers in the fast and slow lanes are given by the relationship in (46). This relationship is illustrated in Figure 1 by the solid red line curve.

An interesting feature of this equilibrium is that full coverage may occur under full centralization but not at the social optimum. This implies that the monopolist may dispatch more travelers than the social planner. Intuitively, this may happen when g is relatively large. This result is clearly at odds with the standard outcome that a monopolist reduces total output. However, our result echoes the possibility that a monopolist underprovides quality relatively to the social optimum (Spence, 1975). In our model, a high level of congestion may be interpreted as a low level of quality.

5.4 Welfare analysis

In Section 4.4, we considered the welfare effects when the market is fully covered. In that setting, welfare changed only as a result of the different sorting of travelers across lanes. We observed that two distortions may arise as we move from decentralization to centralization, one related to the masses of travelers and the other one to the identity of travelers. Overall, we showed that welfare (weakly) decreases as centralization increases.

That analysis ruled out another, fundamental, potential source of distortion, namely the size of travelers, which may depart from social optimality. While in a fully decentralized regime the market is always covered, we indeed proved that both a social planner and a monopoly may want to ration travel, albeit in a different manner. In this section, we show that the interplay between rationing and the two distortions under sorting produces a welfare comparison that is not clear cut.

To derive our welfare results, we make use of a specific (linear) functional form for the gross benefit function, $B(\theta) = b_0 + b\theta$.[11] For the specific parametric conditions under which we are not able to obtain analytical results, we run numerical simulations spanning the entire parameter space.

Our analysis aims at comparing the welfare under partial or full centralization with the welfare under full decentralization. We recall that, under full decentralization, the welfare is simply given by $W_A = \int_0^1 \left(B(\theta) - \theta g \frac{1}{2}\right) d\theta$. To avoid duplicating the analysis in Section 4.4, we restrict to cases in which partial coverage occurs under partial or full centralization.

We first focus on partially centralized regimes with $\mu \in (0, \mu']$, so that differentiation does not occur in equilibrium. We denote as W_{AC} the welfare under partial coverage. This is equal to

$$\begin{aligned} W_{AC} =& \mu \int_{1-\frac{c_{AC}}{\mu}}^{1} \left(B(\theta) - \theta g \frac{c_{AC} + 1 - \mu}{2}\right) d\theta \\ &+ (1-\mu) \int_0^1 \left(B(\theta) - \theta g \frac{c_{AC} + 1 - \mu}{2}\right) d\theta. \end{aligned} \tag{47}$$

The first (second) term of this expression is the aggregate net utility of corporate (atomistic) travelers. The level of congestion is identical across all travelers, since both lanes

[11] When $B(\theta) = b_0 + b\theta$, Assumption 1 becomes $g < b$, Assumption 2 becomes $b_0 \geq 0$, and relaxing Assumption 3 yields $b_0 < b$. Also, it is easy to check the second-order conditions are fully satisfied in all the problems we analyze.

feature the same mass of travelers, $\frac{c_{AC}+1-\mu}{2}$. However, while the entire mass $(1-\mu)$ of atomistic agents travels, only a fraction $c_{AC} \in (0,\mu)$ of corporate agents does so. We denote as ΔW the difference between W_{AC} and W_A, which may be rewritten as follows

$$\begin{aligned}\Delta W \equiv & W_{AC} - W_A \\ = & (1-\mu)\left[\int_0^1 \left(\theta g \frac{1}{2} - \theta g \frac{c_{AC}+1-\mu}{2}\right) d\theta\right] + \\ & + \mu \left[\int_{1-\frac{c_{AC}}{\mu}}^1 \left(\theta g \frac{1}{2} - \theta g \frac{c_{AC}+1-\mu}{2}\right) d\theta - \int_0^{1-\frac{c_{AC}}{\mu}} \left(B(\theta) - \theta g \frac{1}{2}\right) d\theta\right].\end{aligned} \tag{48}$$

This expression clearly illustrates the welfare pros and cons of partial centralization, when the monopoly excludes some individuals. The first line illustrates the positive effect on the aggregate net utility of atomistic travelers. They all travel – as under full centralization –, but they now face a lower congestion level because of the rationing imposed by the monopolist on corporate travelers. The second line expresses the effect on the aggregate net utility of corporate travelers. The first term is the gain for those with $\theta \in \left[1-\frac{c_{AC}}{\mu}, 1\right]$ who are still traveling, but now face lower congestion. The other term illustrates the loss for those with $\theta \in \left[0, 1-\frac{c_{AC}}{\mu}\right]$ who are no longer traveling.

We state the following result:

Proposition 8. *Assume $g < g_{AC}$. Let*

$$\underline{c} \equiv \frac{2b + g(2\mu - 1) - \sqrt{(2b-g)^2 - 4g(4b_0 - g)\mu}}{2g}.$$

For any $\mu \in (0, \mu']$, welfare under partial centralization is larger than welfare under full decentralization (i.e., $\Delta W > 0$) if and only if both following conditions hold:

i) $b_0 < \frac{g}{4}$,

ii) $c_{AC} \in (\underline{c}, \mu)$,

where i) and ii) are satisfied at least for $\mu \to 0$ and $b_0 < \frac{g}{2} - \frac{1}{3}b$.

The Proposition illustrates that a partially centralized regime enhances welfare provided that two rather intuitive conditions apply. First, condition *i)* shows that partial centralization under partial coverage increases welfare only when it is socially optimal to ration travel. Indeed, the benefit of the type-0 traveler, b_0, must be smaller than the increase in the aggregate congestion costs, equal to $\frac{g}{4}$, this individual would impose on all fellow travelers in the same lane if she traveled. Second, condition *ii)* shows that not too many corporate individuals have to be excluded. Recall indeed that corporate individuals may, on aggregate, lose from partial centralization, since, as a result, some of them are excluded (see the second line in (48)). Hence, condition *ii.)* requires this rationing to be limited.

We extend our analysis to partially centralized regimes with relatively high $\mu \in (\mu', 1]$ so that differentiation occurs in equilibrium. We resort to numerical methods, due to the difficulty in characterizing the analytical solutions to the monopolist problem. A graphical illustration of our results is given in Figure 3. In each panel of this Figure, we illustrate, for a given value of μ, the sign of the difference between the welfare under partial or full centralization and the welfare under full decentralization for a grid of admissible values of the parameters b_0 and g, having normalized b to 1. Figure 3 shows that the results in Proposition 8 extend nicely also to the case of a larger μ. We include in the Figure the results of numerical simulations for small values of μ, for which we have full analytical results, to emphasize that our analytical results carry through to larger values of μ.

A further interesting feature that can be observed from comparing the results in the different panels of Figure 3 is that the set of parameters under which partial centralization has a positive welfare effect shrinks as μ increase. An intuition behind this result is that, as the mass of corporate individuals rises, there is more room for the negative excessive rationing effect clearly identified in Proposition 8 in the case of a small μ. On top of this, as μ increases, the two distortions due to sorting gains prominence, as proved in Section 4.4.

6 Equilibrium analysis with taxes

Up to now, we have analyzed an economy under the different degrees of centralization excluding government intervention. This laissez-faire approach is often implemented in practice. While economists advocate road pricing and congestion taxes as tools to improve upon market outcomes, these are rarely implemented in practice (notable exceptions include London, Stockholm and Singapore), likely for political economy reasons (Oberholzer-Gee and Weck-Hannemann, 2002). This section characterizes the tax/subsidy schemes that restore optimality under the two polar cases of a fully decentralized and of a fully centralized regime. We show that the structure of the optimal scheme under full centralization differs sharply from that under decentralization. We then discuss the political feasibility and distributional effects of the first-best restoring tax/subsidy schemes in the two regimes.

To this end, we move to a three-stage game in which the tax authority sets the taxes in the first stage, while the two subsequent stages are identical to those analyzed in the previous sections. We restrict to per-traveler unit taxes, differentiated by lane but not by the identity of the traveler. We denote by t and T the unit tax levied on AVs traveling in the slow and fast lane, respectively.

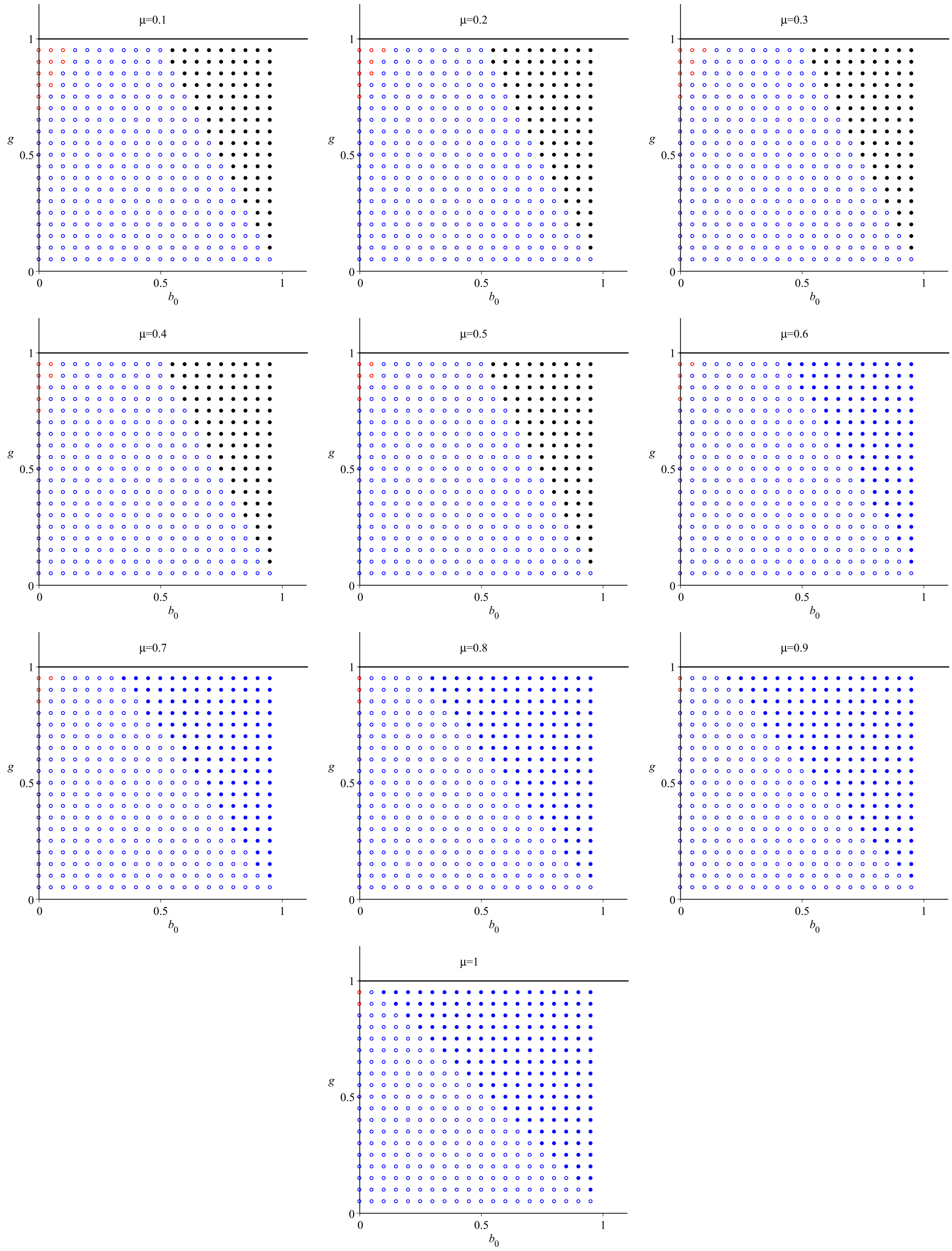

Figure 3: Welfare comparisons across different market equilibria when $b = 1$.
From left to right and top to bottom, μ varies from 0.1 to 1 by 0.1 increases. Blue dots illustrate combinations of parameters such a fully decentralized regime delivers higher welfare (i.e., $W_{AC} < W_A$, or $W_C < W_A$ when $\mu = 1$), red dots illustrate combinations of parameters such that a centralized regime delivers higher welfare (i.e., $W_A < W_{AC}$, or $W_A < W_C$ when $\mu = 1$); black dots illustrate combinations of parameters such that welfare is identical in the two regimes (i.e., $W_A = W_{AC}$). Full (empty) dots denote full (partial) coverage under centralization.

6.1 Full decentralization

In an economy with atomistic individuals and externalities, it is well known that optimality is restored by Pigouvian taxes, which impose on each individual the non internalized social cost. In our framework, optimal road taxes have two goals: i) they need to induce to travel those and only those individuals whose private benefit from a trip is larger than the social cost they impose on fellow travelers; ii) they need to restore the appropriate allocation of travelers across lanes, to minimize the aggregate cost of congestion.

Let t_A and T_A denote the unit taxes levied on atomistic individuals traveling in the slow and fast lane, respectively, that restore social optimum. These are described in the following:

Proposition 9. *Assume full decentralization. The pairs of taxes that replicate the social optimum are as follows*

- *when* $g \leq g_{FB} \cong 5.4179 \times B(0)$, *then*

$$\begin{aligned} t_A &\leq B\left(0\right), \\ T_A &= t_A + g\tfrac{5-\sqrt{7}}{18}; \end{aligned} \tag{49}$$

- *when* $g > g_{FB} \cong 5.4179 \times B(0)$, *then*

$$\begin{aligned} t_A &= g\, s_{FB}\left(1 - f_{FB} - \tfrac{s_{FB}}{2}\right), \\ T_A &= t_A + g\left(1 - f_{FB}\right)\left(s_{FB} - f_{FB}\right). \end{aligned} \tag{50}$$

Taxes modify the after-tax net utility atomistic individuals from traveling, thereby modifying their choices as to whether or not to travel and in which lane. In particular, when g is small relatively to $B(0)$, misallocation of travelers across lanes is the only distortion to be solved, since full coverage would occur not only with atomistic travelers (as always) both also in the first best. Hence, the optimal pair of taxes - as in (50) - should not restrict market coverage. A multiplicity of low enough t, including $t_A = 0$, delivers this. On the other hand, optimal differentiation across lanes is obtained through an appropriate difference $T_A - t_A$, which ensures that the location of the traveler indifferent across the two lanes is identical to that of the social planner. Notice that the indeterminacy of t_A provides a flexible set of alternatives to the tax authority, ranging from solutions that minimize tax burden (when only those who travel in the fast lanes are taxed), to others associated to a larger tax burden, paid by travelers in both lanes.

When instead g is large relatively to $B(0)$, a social planner would exclude individuals with a low value for the travel, while these travelers would travel when atomistic. Hence, both the total number of travelers and their allocation across lanes have to be corrected. The tax in the slow lane, t_A, in (49) is then uniquely determined and ensures that the market coverage replicates the first best. On the other hand, the tax in the fast lane, T_A, induces an optimal degree of differentiation across lanes.

Taxes as in (49) and (50) are congestion charges, based on the external cost imposed by the marginal travelers on fellow riders. They align the incentives of the marginal travelers to social optimality. The distributional consequences, hence the political feasibility, of such congestion charges depend on how the revenue from them is used (Small, 1992). Since congestion charges are welfare-improving, an appropriate redistribution scheme that fully compensates losers could be Pareto-improving. However, probably as a result of imperfect or unclear compensations, congestion charges are typically unpopular, and therefore rarely implemented (Oberholzer-Gee and Weck-Hannemann, 2002). Net of the tax payment, without compensation low θ atomistic individuals stand to lose from the tax scheme. If the market remains fully covered after the tax scheme is implemented, all θ's traveling in the slow lane face more congestion, thus enjoying a lower level of utility. If, instead, the tax scheme excludes the lower portion of θ's, those excluded are worse-off.

6.2 Full centralization

We now look at the welfare maximizing tax/subsidy scheme to be imposed to the monopolist managing the entire AVs fleet.[12] We show that it is remarkably different than the tax scheme to be applied to atomistic travelers. Indeed, a monopolist already perfectly internalizes the congestion externality, so congestion charges are not appropriate. The tax/subsidy scheme has instead to correct the alternative distortions in terms of sorting and rationing, illustrated in the previous sections.

We consider a per-traveler tax/subsidy, potentially differing by lane, imposed on the monopolist. We allow taxes to depend on the mass of travelers in the two lanes, so that $t = t(s, f)$ and $T = T(s, f)$. The maximization problem the monopolist faces is:

$$\begin{aligned} \max_{\substack{p \geq 0, \\ P \geq 0}} \; & [p - t(s, f)]\, n + [P - T(s, f)]\, N \\ s.t. \; & (40) - (41) \end{aligned} \tag{51}$$

where (40) and (41) are the same individual rationality and incentive compatibility constraints faced by the monopolist in the absence of taxes and already discussed in Section 5.3. Solving these two constraints w.r.to the monopolist's fares p and P and plugging them into the monopolist's problem in (51) allows us to rewrite it as follows

$$\begin{aligned} \max_{\substack{s \geq 0, \\ f \geq 0}} \; & [B\,(1 - s - f) - (1 - s - f)\, gs - t(s, f)]\, s + \\ & [B\,(1 - s - f) - (1 - s - f)\, gs + g\,(1 - f)\,(s - f) - T(s, f)]\, f \\ \text{s.t.} \; & s + f \leq 1. \end{aligned} \tag{52}$$

[12] Our analysis would be identical if the same taxes were imposed directly on travelers.

Let the solutions to this problem be denoted by s_{CT} and f_{CT}. When solutions are interior, these are implicitly defined by the following set of FOCs

$$\begin{aligned}&B(1-s_{CT}-f_{CT})-B'(1-s_{CT}-f_{CT})(s_{CT}+f_{CT})+\\&-gs_{CT}(2-4f_{CT}-3s_{CT})-t(s_{CT},f_{CT})+\\&-\frac{\partial t(s_{CT},f_{CT})}{\partial s_{CT}}s_{CT}-\frac{\partial T(s_{CT},f_{CT})}{\partial s_{CT}}f_{CT}=0;\end{aligned} \tag{53}$$

$$\begin{aligned}&B(1-s_{CT}-f_{CT})-B'(1-s_{CT}-f_{CT})(s_{CT}+f_{CT})+\\&+g\left(3(f_{CT})^2+2(s_{CT})^2-2f_{CT}\right)-T(s_{CT},f_{CT})+\\&-\frac{\partial T(s_{CT},f_{CT})}{\partial f_{CT}}f_{CT}-\frac{\partial t(s_{CT},f_{CT})}{\partial f_{CT}}s_{CT}=0.\end{aligned} \tag{54}$$

Let t_C and T_C denote the per-traveler tax/subsidy in the slow and fast lane respectively that restore the social optimum. We establish the following result:

Proposition 10. *Assume full centralization. The social optimum is restored by a system of per-traveler taxes/subsidies of the following form*

$$\begin{aligned}t_C&=g\,s-z_C;\\T_C&=g\,f-z_C.\end{aligned} \tag{55}$$

where

$$z_C\equiv\begin{cases}0 & \text{if } g\le\frac{18\,(B(0)-B'(0))}{4+\sqrt{7}}\\ & \cong 2.7085\times(B(0)-B'(0))\,;\\ B'(0)-B(0)+g\,\frac{4+\sqrt{7}}{18} & \text{if } \frac{18\,(B(0)-B'(0))}{4+\sqrt{7}}\le g\le\frac{36\,B(0)}{4+\sqrt{7}};\\ B'(1-s_{FB}-f_{FB})(s_{FB}+f_{FB})+ & \\ \quad +g\left(2f_{FB}-(s_{FB}^2-\frac{3}{2}f_{FB}^2\right) & \text{if } g\ge\frac{36\,B(0)}{4+\sqrt{7}}\quad\cong 5.4179\times B(0).\end{cases} \tag{56}$$

The Proposition illustrates that social optimality is restored by imposing on the monopolist a per-traveler tax/subsidy scheme based on the mass of travelers, differentiated by lane. Hence, t_C and T_C consist of a tax component (i.e., gs and gf), which increases in the mass of travelers in that lane, and a subsidy component (i.e., z_C in both lanes), exogenously determined by the tax authority, based on the socially optimal number of travelers and equal across the two lanes.

The tax components of t_C and T_C differs dramatically vis-à-vis those in case of full decentralization as they are not congestion charges. Instead, they induce the monopolist to mitigate the misallocation typical of a fully centralized regime, by shifting passengers from the slow to the fast lane, as long as $s>f$. The logic of this tax is similar, for instance, to that of the tax on quality in Lambertini and Mosca (1999), and Cremer and Thisse (1994) in the context of a vertically differentiated oligopoly, when in our model we interpret congestion as a quality level.

An additional instrument is however needed to eliminate the distortion caused by the monopolist when it excludes travelers in a socially inefficient manner. Without taxes, as discussed in Section 5.3, the total number of travelers may be below or above the social optimum. After introducing the tax components of t_C and T_C, however, the monopolist always reduces total travel below the efficient level. To ensure a socially efficient number of travelers, the monopolist must be granted a subsidy on the total mass of travelers, as in (56).

The two components of the taxes serve very different purposes, and also have remarkably different features. The levels of the tax components only depend on the monopolist's choices, being contingent on s and f. In this respect, it is a very simple tax to set, since it does not require any specific knowledge by the tax authority, if not the value of g. On the other hand, the subsidy component requires a deeper knowledge of the market, being based on a perfect knowledge of the travelers' benefit function and of the solution to the first best problem.

Overall, in equilibrium, the subsidy component may exceed the tax component, so that the monopolist receives a net subsidy from the tax authority. This situation always occurs when g is sufficiently high so that efficiency commands to partially cover the market. The tax/subsidy scheme therefore requires to absorb some funding from general taxation. Pels and Verhoef (2004) emphasize the political difficulties in implementing a negative tax scheme on airline companies. The same logic might well apply to the AVs market. The political feasibility of this tax/subsidy scheme appears very dubious, even more so than the congestion charges. A potential solution to improve political feasibility is to associate the scheme to an upfront fixed license that preserves the budget neutrality.

7 Conclusions

The transition to AVs will open a variety of important issues in several domains, including technology, law and ethics.

The technological trajectory of AVs is linked to the progress of artificial intelligence. The ability of artificial intelligence to learn quickly, and, in particular, to adapt switfly to new circumstances will determine how fast the level of automation will progress. The standards set by the Society of Automotive Engineers International identify five different levels of automation, ranging from level 0 of no automation, to level 5 of a fully automated vehicle, able to move autonomously in all terrains and under all circumstances in which an experienced human driver would drive. The most advanced currently manufactured vehicles, produced by Waymo, stand at level 4, defined as highly automatized vehicles that can run without a driver in selected (usually urban) areas. There is some debate on the time horizon for the emergence and commercialization of level 5 vehicles.[13]

[13] https://www.economist.com/leaders/2019/10/10/driverless-cars-are-stuck-in-a-jam (last accessed January 29, 2020).

However, level 4 vehicles are enough for urban traffic to be organized around robotaxis and fleets, with all the resulting welfare effects illustrated in our analysis.

From the legal standpoint, the introduction of AVs, both at level 4 and at level 5, will likely shift the responsibility for accidents from drivers to manufacturers, which will require the design of a new liability regime in the urban transport context (Abraham and Rabin, 2019).

From the moral perspective, Awad et al. (2018) discuss the tradeoffs involved in distributing the wellbeing created by machines, as well as the harm they cannot eliminate. An important moral question has to do, for instance, with how to solve the problem of the division of road risk between the different parties, including the occupants of the cars and the other stakeholders, for example pedestrians.

The transition to AVs, however, will also raise many important economic questions. Key for our analysis, a crucial effect on reducing congestion will come from the process of traffic centralization, resulting from the organization of urban traffic around fleets, which AVs will make cheaper thanks to the reduction in drivers' costs.

Our paper considers a world of AVs, with the resulting potential benefits from coordination. We analyze the welfare effects of moving from a decentralized system, with atomistic vehicles only, to a centralized market, in which all travelers use vehicles that are part of a fleet managed by a monopolist, through a transition period where some atomistic vehicles share the road with others that are part of a fleet. We analyze an environment with heterogenous travelers who are sorted in one of two lanes, with potentially different levels of congestion. In this setting, a reduction in aggregate congestion costs potentially arises not only as a result of a reduction in overall travel, but also as a result of an optimal sorting of travelers across different lanes with different speeds. AVs, along with other recent developments in urban transport will dramatically decrease the cost of the sorting technology, thereby making it a viable and important alternative to manage congestion. We show that, while centralization internalizes the congestion externality, it introduces additional distortions. We analyze sorting and rationing as tools to manage congestion. We find that, when sorting is the most relevant, welfare decreases with centralization. When, instead, rationing is also required, centralization may be welfare-superior.

The self driving technology will likely affect both travel demand, and the welfare-maximizing level of congestion. With AVs, consumers may spend more productively their time on vehicles. This will arguably increase the welfare-maximizing level of congestion. At the same time, however, travel demand is expected to increase as well as a result (Gucwa, 2014). Whether the combination of the two effects will induce more or less travel rationing under the social planner is debatable. Additional factors could contribute to changing both travel demand, and the welfare-maximizing level (with an unclear effect on required rationing), including the improved vehicles coordination, as well as changes in the cost of infrastructural expansions. As a result, it may well possible that sorting will become, in several urban contexts, more important than reducing the aggregate amount of cars circulating, to curb congestion cost. Our model shows that, in such environments, the centralization process brought about by the emergence of fleets of AVs is likely to have adverse welfare effects. Notice, finally, that a monopolistic company that manages the infrastructure (rather than a fleet of AVs) and can charge different prices across lanes behaves exactly like our monopolist in a fully centralized setting. Our results suggest that, when congestion is dealt with by sorting, and not by rationing, a monopolist that manages the infrastructure reduces welfare with respect to the fully decentralized case, with atomistic drivers. This should raise a word of caution on infrastructural projects for AVs managed by private unregulated entities.

We then analyze how to restore first best with road taxes. The familiar congestion charges are optimal with decentralized travel. However, they fail to restore optimal welfare when vehicles are part of a fleet. In this case, the welfare-maximizing tax/subsidy scheme is very different, and may require subsidizing the company – something likely to be politically very unappealing.

A couple of extensions of our model seem natural. One would analyze welfare when multiple companies, each managing a fleet of AVs, compete for passengers. While we believe the qualitative results would not be altered, the analysis might provide some additional insights. A second possible extension would involve the analysis of the interaction between a welfare-oriented public transit company and a profit-maximizing service provider. This could be illustrative of the effects of government direct involvement in the urban transit business in a world of fleets.

References

Abraham, K. and R. Rabin, 2019. Automated Vehicles and Manufacturer Responsibility for Accidents: a New Legal Regime for a New Era. Virginia Law Review, 105(1), 127-171.

Arnott, R., A. de Palma and R. Lindsey, 1990. Economics of a bottleneck. Journal of Urban Economics, 27, 111-130.

Awad, E., S. Dsouza, R. Kim, J. Schulz, J. Henrich, A. Shariff, J.-F. Bonnefon and I. Rahwan, 2018. The Moral Machine experiment. Nature, 563, 59-64.

Basso, L., 2008. Airport deregulation: Effects on prices and capacity. International Journal of Industrial Organization, 26, 1015-1031.

Basso, L. and A. Zhang, 2007. Congestible facility rivalry in vertical structures. Journal of Urban Economics, 61, 218-237.

Börjesson, M., M. Fosgerau and S. Algers, 2012. On the income elasticity of the value of travel time. Transportation Research Part A: Policy and Practice, 46, 368-377.

Brueckner, J. K., 2002. Airport congestion when carriers have market power. American Economic Review, 92(5), 1357-1375.

Brueckner, J. K., 2005. Internalization of airport congestion: A network analysis. International Journal of Industrial Organization, 23(7-8), 599-614.

Cremer, H. and J.-F. Thisse, 1994. Commodity taxation in a differentiated oligopoly. International Economic Review, 35(3), 613-633.

Czerny, A. and A. Zhang, 2015. Third-degree price discrimination in the presence of congestion externality. Canadian Journal of Economics, 48(4), 1430-1455.

Daniel, J. I., 1995. Congestion Pricing and Capacity of Large Hub Airports: A Bottleneck Model with Stochastic Queues. Econometrica, 63(2), 327-370.

Daniel, J. I. and K. T. Harback, 2008. (When) Do hub airlines internalize their self-imposed congestion delays? Journal of Urban Economics, 63(2), 583-612.

Downs, A., 1962. The law of peak-hour express-way congestion. Traffic Quarterly, 16(3), 393-409.

Duranton, G. and M. Turner, 2011. The Fundamental Law of Road Congestion: Evidence from US Cities. American Economic Review, 101(6), 2616-2652.

Fagnant, D. and K. Kockelman, 2015. Preparing a nation for autonomous vehicles. Transportation Research Part A: Policy and Practice, 77(4), 167-181.

Gucwa, M., 2014. Mobility and Energy Impacts of Automated Cars. Presentation to the Automated Vehicles Symposium. Burlingame, CA, July 16.

Lambertini, L. and M. Mosca, 1999. On the Regulation of a Vertically Differentiated Market. Australian Economic Papers, 38(4), 354-366.

Lamotte, R., A. de Palma and N. Geroliminis, 2016. Sharing the road: the economics of autonomous vehicles. Elitisme Working Paper Series 2016-002.

Lindsey, R., A. de Palma and H. Silva, 2019. Equilibrium in the Bottleneck Model with Atomic and non Atomic Users. Transportation Research Part B: Methodological, 124, 82-107.

Mayer, C. and T. Sinai, 2003. Network Effects, Congestion Externalities, and Air Traffic Delays: Or Why Not All Delays Are Evil. American Economic Review, 93(4), 1194-1215.

Molnar, A., 2013. Congesting the commons: A test for strategic congestion externalities in the airline industry, mimeo.

Montes, R., W. Sand-Zantman and T. Valletti, 2018. The Value of Personal Information in Online Markets with Endogenous Privacy. Management Science, 65(3), 1342-1362.

Mussa, M. and S. Rosen, 1978. Monopoly and Product Quality. Journal of Economic Theory, 18, 301-317.

Oberholzer-Gee, F. and H. Weck-Hannemann, 2002. Pricing road use: politico-economic and fairness considerations. Transportation Research Part D: Transport and Environment, 7(2), 357-371.

Ostrovsky, M. and M. Schwarz, 2018. Carpooling and the Economics of Self-Driving Cars. NBER Working Paper No. 24349.

Pels, E. and E. Verhoef, 2004. The economics of airport congestion pricing. Journal of Urban Economics, 55(2), 257-277.

Rupp, N. G., 2009. Do carriers internalize congestion costs? Empirical evidence on the internalization question. Journal of Urban Economics, 65(1), 24-37.

Schrank, D., T. Lomax and B. Eisele, 2011. TTI's 2011 Urban Mobility Report. Texas Transportation Institute, The Texas A&M University System.

Silva, H. and E. Verhoef, 2013. Optimal pricing of flights and passengers at congested airports and the efficiency of atomistic charges. Journal of Public Economics, 106, 1-13.

Silva H., R. Lindsey, A. de Palma and V. van den Berg, 2016. On the Existence and Uniqueness of Equilibrium in the Bottleneck Model with Atomic Users. Transportation Science, 51 (3), 863-881.

Simoni M., K. Kockelmana, K. Gurumurthya and J. Bischoff, 2019. Congestion pricing in a world of self-driving vehicles: An analysis of different strategies in alternative future scenarios. Transportation Research Part C: Emerging technologies, 98, 167-185.

Small, K., 1992. Using the revenues from congestion pricing. Transportation, 19, 359-381.

Small, K., 2012. Valuation of travel time. Economics of Transportation, 1, 2-14.

Small, K., C. Winston and J. Yan, 2005. Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability. Econometrica, 73 (4), 1367-1382.

Small, K., C. Winston and J. Yan, 2006. Differentiated Road Pricing, Express Lanes, and Carpools: Exploiting Heterogeneous Preferences in Policy Design. Brookings-Wharton Papers on Urban Affairs, 53-96.

Spence, A., 1975. Monopoly, quality, and regulation. Bell Journal of Economics. 6(2), 417-429.

Van den Berg, V. and E. Verhoef, 2016. Autonomous cars and dynamic bottleneck congestion: The effects on capacity, value of time and preference heterogeneity. Transportation Research Part B: Methodological, 94, 43-60.

Vickrey, W. S., 1969. Congestion theory and transport investment. American Economic Review, 59(2), 251-260.

Appendix A: derivations and proofs

This Appendix contains the proofs of all Lemmata and Propositions of the paper, together with the derivations of additional results contained in the paper.

Proof of Proposition 1. First, we show that, if the planner optimally excludes some travelers, it would exclude travelers with the lowest θ's. By contradiction, suppose that it dispatches a θ'-type, and it does not dispatch a θ''-type, with $\theta' < \theta''$. Then, a switch between the θ''-type not traveling and the θ'-type traveling would leave congestion unaffected while increasing the aggregate net benefit from traveling because $B'(\theta) > 0$, thereby increasing social welfare.

Second, we show that it is never optimal to assign to the same lane travelers in non-contiguous partitions of the unit line. Let types θ' and θ'' be traveling in a lane with a mass of travelers equal to $l' + l''$. Assume, by contradiction, that $\theta' \in [\bar{\theta}', \bar{\theta}' + l']$ and $\theta'' \in [\bar{\theta}'', \bar{\theta}'' + l'']$, with $\bar{\theta}' + l' < \bar{\theta}''$. Let also type θ''' be traveling in a lane with a mass of travelers equal to l''' and that $\theta''' \in [\bar{\theta}' + l', \bar{\theta}'']$. Next, assume that $l' + l'' < l'''$ so that the lane where types θ' and θ'' travel is less congested than the one where type θ''' travels. A switch between types θ' and θ''' would leave congestion in both lanes unaltered, while increasing the aggregate net benefit from traveling because $\theta' < \theta'''$ and $B'(\theta) > 0$. Similarly, assume that $l''' < l' + l''$ so that the lane where types θ' and θ'' travel is more congested than the one where type θ''' travels. A switch between types θ'' and θ''' would leave congestion in both lanes unaltered, while increasing the aggregate net benefit from traveling because $\theta'' > \theta'''$ and $B'(\theta) > 0$.

As a result, the welfare function can be rewritten as follows

$$W = \int_{1-l_1-l_2}^{1-l_2} (B(\theta) - \theta g l_1)\, d\theta + \int_{1-l_2}^{1} (B(\theta) - \theta g l_2)\, d\theta, \qquad \text{(A-1)}$$

where l_1 is the mass of travelers in lane 1 and l_2 is the mass of travelers in lane 2. It is easy to prove that $l_1 > l_2$, so that travelers with lower θ's are placed in lane 1, the more congested lane. Consider two travelers, θ' and θ'', with $\theta' < \theta''$ and first suppose θ' uses lane 1, while θ'' uses lane 2. In equilibrium, the aggregate net benefit for the two travelers is $w \equiv B(\theta') - \theta' g l_1 + B(\theta'') - \theta'' g l_2$. Suppose now that θ'' uses lane 1 and θ' uses lane 2. The aggregate net benefit for the two travelers is $w' \equiv B(\theta'') - \theta'' g l_1 + B(\theta') - \theta' g l_2$. Then, $w - w' = B(\theta') - \theta' g l_1 + B(\theta'') - \theta'' g l_2 - (B(\theta'') - \theta'' g l_1 + B(\theta') - \theta' g l_2) = g(\theta'' - \theta')(l_1 + l_2) > 0$, which shows that social welfare is higher in the first case. The planner problem may then be written as in (7).

The Lagrangean of problem (8) is

$$\mathcal{L}_{FB} \equiv \int_{1-s-f}^{1-f} (B(\theta) - \theta g s)\, d\theta + \int_{1-f}^{1} (B(\theta) - \theta g f)\, d\theta - \lambda(s + f - 1). \qquad \text{(A-2)}$$

At the solutions to this problem, denoted by s_{FB}, f_{FB} and λ_{FB}, Kuhn-Tucker conditions require

$$\frac{\partial \mathcal{L}_{FB}}{\partial s} = B\left(1 - s_{FB} - f_{FB}\right) - 2g s_{FB}\left(1 - f_{FB} - \frac{3}{4}s_{FB}\right) - \lambda_{FB} = 0, \quad \text{(A-3)}$$

$$\frac{\partial \mathcal{L}_{FB}}{\partial f} = B\left(1 - s_{FB} - f_{FB}\right) + g\left(s_{FB}^2 + \frac{3}{2}f_{FB}^2 - 2f_{FB}\right) - \lambda_{FB} = 0, \quad \text{(A-4)}$$

$$\frac{\partial \mathcal{L}_{FB}}{\partial \lambda} = s_{FB} + f_{FB} - 1 \leq 0, \qquad \lambda_{FB} \geq 0 \text{ and } \qquad \frac{\partial \mathcal{L}_{FB}}{\partial \lambda}\lambda_{FB} = 0. \quad \text{(A-5)}$$

Assume that $\lambda_{FB} = 0$ and $s_{FB} + f_{FB} - 1 < 0$, so that the solution is interior. Substitute $\lambda_{FB} = 0$ in (A-3) and (A-4), equate them and solve w.r.to f_{FB} to obtain (12).

Assume instead that $\lambda_{FB} \geq 0$ and $s_{FB} + f_{FB} = 1$. Substitute $s_{FB} = 1 - f_{FB}$ in (A-3) and (A-4), equate them and solve w.r.to f_{FB} to obtain $f_{FB} = \frac{5-\sqrt{7}}{6}$. Use again $s_{FB} = 1 - f_{FB}$ to get $s_{FB} = \frac{1+\sqrt{7}}{6}$. Plug f_{FB} and s_{FB} thus obtained into (A-3) or (A-4) and solve w.r.to λ_{FB} to obtain $\lambda_{FB} = B(0) - g\frac{4+\sqrt{7}}{36}$. Solve $\lambda_{FB} \geq 0$ w.r.to g to get $g \leq g_{FB}$. ■

Derivation of the comparative static results in the first best. We derive here the comparative statics results mentioned in Section 3, i.e., $\frac{\partial s_{FB}}{\partial g} < 0$ and $\frac{\partial f_{FB}}{\partial g} < 0$.

Denote the FOCs of the maximization problem for the planner in (8), given in (9) and (10), as $h_s(s_{FB}, f_{FB}, g) = 0$ and $h_f(s_{FB}, f_{FB}, g) = 0$, respectively. Implicit differentiation of the FOCs w.r.to g gives

$$\frac{ds_{FB}}{dg} = \frac{\frac{\partial h_f}{\partial g}\frac{\partial h_s}{\partial f_{FB}} - \frac{\partial h_f}{\partial f_{FB}}\frac{\partial h_s}{\partial g}}{\frac{\partial h_f}{\partial f_{FB}}\frac{\partial h_s}{\partial s_{FB}} - \frac{\partial h_f}{\partial s_{FB}}\frac{\partial h_s}{\partial f_{FB}}}; \quad \text{(A-6)}$$

$$\frac{df_{FB}}{dg} = \frac{\frac{\partial h_f}{\partial g}\frac{\partial h_s}{\partial s_{FB}} - \frac{\partial h_f}{\partial s_{FB}}\frac{\partial h_s}{\partial g}}{\frac{\partial h_f}{\partial f_{FB}}\frac{\partial h_s}{\partial s_{FB}} - \frac{\partial h_f}{\partial s_{FB}}\frac{\partial h_s}{\partial f_{FB}}}. \quad \text{(A-7)}$$

In problem (8), SOCs require $\frac{\partial h_f}{\partial f_{FB}}\frac{\partial h_s}{\partial s_{FB}} - \frac{\partial h_f}{\partial s_{FB}}\frac{\partial h_s}{\partial f_{FB}} > 0$. As a result,

$$sign\frac{ds_{FB}}{dg} = sign\left(\frac{\partial h_f}{\partial g}\frac{\partial h_s}{\partial f_{FB}} - \frac{\partial h_f}{\partial f_{FB}}\frac{\partial h_s}{\partial g}\right) \quad \text{(A-8)}$$

and

$$sign\frac{df_{FB}}{dg} = sign\left(\frac{\partial h_f}{\partial g}\frac{\partial h_s}{\partial s_{FB}} - \frac{\partial h_f}{\partial s_{FB}}\frac{\partial h_s}{\partial g}\right). \quad \text{(A-9)}$$

Using (9) and (10) yields

$$\frac{\partial h_s}{\partial s_{FB}} = -B'(1 - s_{FB} - f_{FB}) - 2g\left(1 - f_{FB} - \frac{3}{2}s_{FB}\right), \quad \text{(A-10)}$$

$$\frac{\partial h_s}{\partial f_{FB}} = -B'(1 - s_{FB} - f_{FB}) + 2gs_{FB}, \tag{A-11}$$

$$\frac{\partial h_f}{\partial s_{FB}} = -B'(1 - s_{FB} - f_{FB}) + 2gs_{FB}, \tag{A-12}$$

$$\frac{\partial h_f}{\partial f_{FB}} = -B'(1 - s_{FB} - f_{FB}) + g\left(3f_{FB} - 2\right), \tag{A-13}$$

$$\frac{\partial h_s}{\partial g} = -2s_{FB}\left(1 - f_{FB} - \frac{3}{4}s_{FB}\right), \tag{A-14}$$

$$\frac{\partial h_f}{\partial g} = (s_{FB})^2 + \frac{3}{2}(f_{FB})^2 - 2f_{FB}. \tag{A-15}$$

When evaluated at the equilibrium relationship (12), the numerator of $\frac{ds_{FB}}{dg}$ is negative for any $s_{FB} \leq \bar{s}_{FB}$, where $\bar{s}_{FB}$ is the full coverage and maximum value of s_{FB}, computed in Proposition 1. Hence, $\frac{ds_{FB}}{dg} < 0$. Also, from (12), $\frac{df_{FB}}{ds_{FB}} > 0$, hence $\frac{df_{FB}}{dg} = \frac{df_{FB}}{ds_{FB}}\frac{ds_{FB}}{dg} < 0$.

Proof of Proposition 2. We first prove that any equilibrium must be such that $s = f$. Assume, by contradiction, that $s > f$; any traveler in the lane S has an incentive to switch to the other lane, proving that this cannot be an equilibrium. When $s = f$ and in the absence of fares, the two IC constraints (3) and (5) are always trivially satisfied and the two IR constraints (2) and (4) become identical and equal to $B(\theta) - \theta g s = B(\theta) - \theta g f$. This value is nonnegative for any θ, showing that all atomistic travelers travel. ■

Proof of Lemma 1. We proved in Proposition 2 that all atomistic travelers travel. It follows that

$$s^a + f^a = 1 - \mu. \tag{A-16}$$

Next, consider separately the two conditions on μ in the Lemma. First, assume

$$1 - \mu \geq s^c - f^c. \tag{A-17}$$

We want to prove that s^a and f^a are such that $s = s^a + s^c = f^a + f^c = f$. Assume, by contradiction, s^a and f^a are such that $s > f$. For all θ-type atomistic travelers in lane S, it must be the case that

$$B(\theta) - \theta g(s^a + s^c) \geq B(\theta) - \theta g(f^a + f^c),$$

which reduces to $s^a + s^c \leq f^a + f^c$ or, equivalently, using $s = s^a + s^c$ and $f = f^a + f^c$, to $s < f$. This contradicts our initial hypothesis that $s > f$. A similar argument rules out the possibility that $s < f$. Hence, when (A-17) holds, s^a and f^a are such that $s = f$.

Next, assume

$$1 - \mu < s^c - f^c. \tag{A-18}$$

We want to prove that $s^a = 0$. Assume, by contradiction, that $s^a > 0$. For all θ-type atomistic travelers in lane S, it must be the case that

$$B(\theta) - \theta g(s^a + s^c) \geq B(\theta) - \theta g(f^a + f^c),$$

which reduces to $s^a + s^c \leq f^a + f^c$ or, equivalently, using (A-16), to $1-\mu \geq s^c - f^c + 2s^a$. This contradicts (A-18). Hence, when (A-18) holds, $s^a = 0$ and, substituting $s^a = 0$ into (A-16), $f^a = 1 - \mu$ as in (16). ■

Proof of Proposition 3. The proof is part of the proof of Proposition 6. ■

Proof of Proposition 4. The proof is part of the proof of Proposition 7. ■

Proof of Proposition 5. We evaluate and compare the welfare expression in (26) at the equilibrium masses of travelers in the different market situations.

Let $\mathcal{B} \equiv \int_0^1 B(\theta)d\theta$. Denote by $\bar{W}_{subscript}$ the equilibrium welfare, where the *subscript* is the ones used in the different subsections of Section 4, that is, FB for first best, A for atomistic travelers only, AC for atomistic and corporate travelers coexisting, and C for corporate travelers only; in case of atomistic and corporate travelers coexisting, we use subscript AC when $\mu \in \left(0, \frac{1}{2}\right]$ and AC' otherwise. Then

$$\bar{W}_{FB} = \mathcal{B} - \frac{44 - 7\sqrt{7}}{108} g, \tag{A-19}$$

$$\bar{W}_A = \bar{W}_{AC} = \mathcal{B} - \frac{1}{4} g, \tag{A-20}$$

$$\bar{W}''_{AC'} = \mathcal{B} - \frac{16\mu^3 - 24\mu^2 + 45\mu - 1 + (8\mu^2 - 10\mu - 1)\sqrt{4\mu^2 - 2\mu + 1}}{108\mu} g, \tag{A-21}$$

and

$$\bar{W}_C = \mathcal{B} - \frac{12 - \sqrt{3}}{36} g; \tag{A-22}$$

the comparison follows through immediately. ■

Proof of Lemma 2. The proof is part of the proof of Proposition **6.** ■

Proof of Proposition 6. Assume μ is sufficiently large so that $1 - \mu < s^c - f^c$ holds at equilibrium s^c and f^c (something that will be checked later on), in which case Proposition 2 and Lemma 1 ensure that $s^a = 0$ and $f^a = 1 - \mu$.

The Lagrangean of the monopolist problem in (32) is given by

$$\mathcal{G}_{AC} \equiv \left(B\left(1-\frac{s^c+f^c}{\mu}\right) - \left(1-\frac{s^c+f^c}{\mu}\right) g s^c \right)(s^c+f^c) + \tag{A-23}$$
$$+ g\left(1-\frac{f^c}{\mu}\right)(s^c - f^c - f^a) f^c - \gamma(s^c+f^c-\mu).$$

At the solutions to this problem, denoted by s^c_{AC},f^c_{AC} and γ_{AC}, exploiting $f^a = 1-\mu$, Kuhn-Tucker conditions require

$$\frac{\partial \mathcal{G}_{AC}}{\partial s^c} = B\left(1-\frac{s^c_{AC}+f^c_{AC}}{\mu}\right) - \frac{B'\left(1-\frac{s^c_{AC}+f^c_{AC}}{\mu}\right)\times(s^c_{AC}+f^c_{AC})}{\mu} + \tag{A-24}$$
$$+ \frac{g s^c_{AC}\left[3s^c_{AC}+4f^c_{AC}-2\mu\right]}{\mu} - \gamma_{AC} = 0,$$

$$\frac{\partial \mathcal{G}_{AC}}{\partial f^c} = B\left(1-\frac{s^c_{AC}+f^c_{AC}}{\mu}\right) - \frac{B'\left(1-\frac{s^c_{AC}+f^c_{AC}}{\mu}\right)\times(s^c_{AC}+f^c_{AC})}{\mu} + \tag{A-25}$$
$$+ \frac{g\left[2\left(s^c_{AC}\right)^2 + f^c_{AC}\left(4(1-\mu)+3f^c_{AC}-2\right)-(1-\mu)\mu\right]}{\mu} - \gamma_{AC} = 0,$$

$$\frac{\partial \mathcal{G}_{AC}}{\partial \gamma} = s^c_{AC}+f^c_{AC}-\mu \leq 0, \qquad \gamma_{AC} \geq 0 \text{ and } \qquad \frac{\partial \mathcal{G}_{AC}}{\partial \gamma}\gamma_{AC} = 0. \tag{A-26}$$

Assume that $\gamma_{AC} > 0$ and $s^c_{AC} + f^c_{AC} = \mu$. Substitute $s^c_{AC} = \mu - f^c_{AC}$ in (A-24) and (A-25), equate them and solve w.r.to f^c_{AC} to obtain f^c_{AC} as in (21). Use again $s^c_{AC} + f^c_{AC} = \mu$ to obtain s^c_{AC} as in (21). Plug s^c_{AC} and f^c_{AC} thus obtained into (A-24) or (A-25) and solve w.r.to γ_{AC} to obtain $\gamma_{AC} = B(0) - B'(0) + g\frac{(4\mu-1)\sqrt{4\mu^2-2\mu+1}+8\mu^2+5\mu-1}{18\mu}$. Solve $\gamma_{AC} \geq 0$ w.r.to g to obtain $g \leq g_{AC}$ as in (35). Plug s^c_{AC} and f^c_{AC} back into the inequality $1-\mu < s^c - f^c$, which becomes $\mu > \frac{1}{2}$. Hence, when the monopolist chooses s^c_{AC} and f^c_{AC}, atomistic travelers place themselves as in (16).

Assume that $\gamma_{AC} = 0$ and $s^c_{AC} + f^c_{AC} \leq \mu$, so that the solution is interior. Substitute $\gamma_{AC} = 0$ in (A-24) and (A-25), equate them and solve w.r.to f^c_{AC} to obtain $f^c_{AC}(s^c_{AC})$ as in (39).

Assume now μ is sufficiently small so that $1-\mu \geq s^c - f^c$ in equilibrium (something that will be checked later on) and (16) holds.

The Lagrangean of the monopolist problem in (28) is given by

$$\mathcal{L}_{AC} \equiv \left[B\left(1-\frac{c}{\mu}\right) - \left(1-\frac{c}{\mu}\right) g\frac{c+s^a+f^a}{2}\right] c - \lambda(c-\mu). \tag{A-27}$$

At the solutions to this problem, denoted by c_{AC} and λ_{AC}, Kuhn-Tucker conditions require

$$\begin{aligned}\frac{\partial \mathcal{L}_{AC}}{\partial c} =& \frac{3g\left(c_{AC}\right)^{2}}{2\mu} + c_{AC}\frac{g\left(1-2\mu\right) - B'\left(1-\frac{c_{AC}}{\mu}\right)}{\mu} + \\ &+ \frac{2B\left(1-\frac{c_{AC}}{\mu}\right) - g\left(1-\mu\right)}{2} - \lambda_{AC} = 0, \end{aligned} \tag{A-28}$$

$$\frac{\partial \mathcal{L}_{AC}}{\partial \lambda} = c_{AC} - \mu \leq 0, \qquad \lambda_{AC} \geq 0 \text{ and } \qquad \frac{\partial \mathcal{L}_{AC}}{\partial \lambda}\lambda_{AC} = 0. \tag{A-29}$$

Assume that $\lambda_{AC} > 0$ and $c_{AC} = \mu$. Substitute $c_{AC} = \mu$ in (A-28) and solve w.r.to λ_{AC} to obtain $\lambda_{AC} = B(0) - B'(0) + \frac{1}{2}g$. Solve $\lambda_{AC} > 0$ w.r.to g to obtain g_{AC} as in (35).

Assume that $\lambda_{AC} = 0$ and $c_{AC} < \mu$, so that the solution is interior. Substitute $\lambda_{AC} = 0$ in (A-28) to obtain (29).

We now check under which conditions $1-\mu < (\geq)\, s^{c} - f^{c}$ in equilibrium. First note that $1-\mu > \frac{1}{2} > \mu \geq s^{c}_{AC} - f^{c}_{AC}$ for any s^{c}_{AC}, f^{c}_{AC}; this means that $\mu \geq \frac{1}{2}$ is a necessary condition for $1-\mu < s^{c}_{AC} - f^{c}_{AC}$ to be fulfilled, hence for differentiation to arise. Therefore, we focus on the interval $\mu \in \left[\frac{1}{2}, 1\right)$ to calculate the threshold value μ' such that $1-\mu = s^{c}_{AC} - f^{c}_{AC}$.

The difference $s^{c}_{AC} - f^{c}_{AC}$, computed using (39), is

$$\frac{1}{3}\left(s^{c}_{AC} - 2\mu + \sqrt{-4s^{c}_{AC} - \mu + 2s\mu + \mu^{2} + 7\left(s^{c}_{AC}\right)^{2} + 1} + 1\right).$$

We find that: $1-\mu > s^{c}_{AC} - f^{c}_{AC}$ at $\mu = \frac{1}{2}$ for any admissible s^{c}_{AC} and f^{c}_{AC}; $1-\mu < s^{c}_{AC} - f^{c}_{AC}$ at $\mu = 1$ for any s^{c}_{AC} and f^{c}_{AC} such that $s^{c}_{AC} + f^{c}_{AC} \neq 0$. Since both $1-\mu$ and $s^{c}_{AC} - f^{c}_{AC}$ are continuous in μ, it follows that $s^{c}_{AC} - f^{c}_{AC}$ has $2n+1$ intersections with $1-\mu$ in $\mu \in \left(\frac{1}{2}, 1\right)$, $n \in N = \{0, 1, 2, 3...\}$. If $n = 0$, the result that $\mu' \in \left(\frac{1}{2}, 1\right)$ follows because $1-\mu > s^{c}_{AC} - f^{c}_{AC}$ for any $\mu \in \left[\frac{1}{2}, \mu'\right)$ and $1-\mu < s^{c}_{AC} - f^{c}_{AC}$ for any $\mu \in (\mu', 1]$. To rule out that $n > 0$, we proceed as follows. If n were higher than 0, there would be n intersections between $s^{c}_{AC} - f^{c}_{AC}$ and $1-\mu$ - denote μ^{*} the values of μ at these intersections - such that $s^{c}_{AC} - f^{c}_{AC} > (<)\, 1-\mu$ when $\mu < (>)\, \mu^{*}$. This would imply that in the n neighborhoods of μ^{*}, the firm finds it profitable to switch from differentiation to no differentiation as μ increases. Yet, this cannot be the case. Indeed, the only partial coverage solution that solve problem (32) is a pair (s^{c}_{AC}, f^{c}_{AC}) satisfying FOCs (33) and (34) such that $s^{c}_{AC} - f^{c}_{AC} > 1-\mu$ (and $P > p$); any alternative pair (s^{c}_{AC}, f^{c}_{AC}) such that $s^{c}_{AC} - f^{c}_{AC} \leq 1-\mu$ (and $P = p$) could have been chosen by the firm but it is not because it does not maximize profits. As μ increases, the partial coverage profit-maximizing solutions are still feasible in that $s^{c}_{AC} - f^{c}_{AC} > 1-\mu$ is fulfilled a fortiori, hence switching to no differentiation cannot be a profit-maximizing strategy. ■

Proof of Proposition 7. The Lagrangean of the monopolist problem in (42) is given by

$$\mathcal{L}_C \equiv (B(1-s-f) - (1-s-f)gs)(s+f) + (g(1-f)(s-f))f - \lambda(s+f-1). \tag{A-30}$$

At the solutions to this problem, denoted by s_C, f_C and λ_C, Kuhn-Tucker conditions require

$$\begin{aligned}
\frac{\partial \mathcal{L}_C}{\partial s} =& B(1-s_C-f_C) - B'(1-s_C-f_C)(s_C+f_C) + \\
& - g s_C (2 - 4f_C - 3s_C) - \lambda_C = 0,
\end{aligned} \tag{A-31}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}_C}{\partial f} =& B(1-s_C-f_C) - B'(1-s_C-f_C)(s_C+f_C) + \\
& g\left(3(f_C)^2 - 2f_C + 2(s_C)^2\right) - \lambda_C = 0,
\end{aligned} \tag{A-32}$$

$$\frac{\partial \mathcal{L}_C}{\partial \lambda} = s_C + f_C - 1 \leq 0, \qquad \lambda_C \geq 0 \text{ and } \qquad \frac{\partial \mathcal{L}_C}{\partial \lambda}\lambda_C = 0. \tag{A-33}$$

Assume that $\lambda_C = 0$ and $s_C + f_C - 1 < 0$. Substitute $\lambda_C = 0$ in (A-31) and (A-32), equate them and solve w.r.to f_C to obtain (46).

Assume instead that $\lambda_C \geq 0$ and $s_C + f_C = 1$. Substitute $s_C = 1 - f_C$ in (A-31) and (A-32), equate them and solve w.r.to f_C to obtain $f_C = \frac{1}{2} - \frac{\sqrt{3}}{6}$ as in (B-14). Use again $s_C = 1 - f_C$ to get $s_C = \frac{1}{2} + \frac{\sqrt{3}}{6}$ as in (B-14). Plug f_C and s_C thus obtained into (A-31) or (A-32) and solve w.r.to λ_C to obtain $\lambda_C = B(0) - B'(0) + g\frac{4+\sqrt{3}}{6}$. Solve $\lambda_C \geq 0$ w.r.to g to obtain $g \leq g_C$ as in (45). ■

Derivation of the sufficiency of Assumption 3 for full coverage to occur in first best and in all regimes. We show that, when Assumption 3 holds, full coverage always occurs.

First best. In the first best, full coverage occurs when $g \leq g_{FB}$, where g_{FB} is given in (11). From the concavity of $B(.)$ and Assumption 3, we may write $B'(1) \leq B'(0) \leq B(0)$. Using (11) and Assumption 1, then $g < B'(1) < g_{FB}$, which proves our claim.

Full decentralization. Full coverage is the result of Assumptions 1 and 2 only, which ensure that all travelers enjoy a nonnegative utility when traveling and paying no fee.

Partial centralization. Note that g_{AC} in (35) is nonpositive when $\frac{B'(0)}{B(0)} \leq 1$, which proves our claim.

Full centralization. An argument identical to the one used in the case of partial centralization proves our claim. □

Derivation of the conditions for the existence of the Spence distortion. A sufficient condition for the mass of travelers to be larger under full centralization than in the first best is whenever the monopolist fully covers the market but the social planner does not, i.e., $\max\{g_{FB}, g_{AC}\} \leq g$. This occurs when $\max\left\{\frac{6}{4+\sqrt{3}}[B'(0) - B(0)]; \frac{36}{4+\sqrt{7}}B(0)\right\} \leq g < B'(1)$. This interval is not empty if and only if $B'(0) < \frac{28+\sqrt{7}+6\sqrt{3}}{36}B'(1) \cong$

1.1399 $B'(1)$, thus requiring the gross benefit function $B(.)$ to be "not too" concave. $\square$

Proof of Proposition 8. With the linear specification $B(\theta) = b_0 + b\theta$, the welfare gap $\Delta W \equiv W_{AC} - W_A$ becomes

$$\Delta W = \frac{(\mu - c_{AC})(-gc_{AC}^2 - 2b\mu + 2g\mu - g\mu^2 - 4\mu b_0 + 2bc_{AC} - gc_{AC} + 2gc_{AC}\mu)}{4\mu} \tag{A-34}$$

with

$$c_{AC} = \frac{2b - g + 2g\mu - \sqrt{g^2\mu^2 - g^2\mu - 4bg + 4b^2 + g^2 + 2bg\mu - 6g\mu b_0}}{3g}.$$

To study the sign of ΔW, we first observe that plugging $c_{AC} = 0$ into ΔW yields

$$\Delta W(0) = \frac{\mu}{4}(-2b + 2g - 4b_0 - g\mu),$$

which is negative under Assumption 1. We then solve $\Delta W = 0$ for c_{AC} and get three solutions: μ,

$$\underline{c} = \frac{2b + g(2\mu - 1) - \sqrt{(2b - g)^2 - 4g(4b_0 - g)\mu}}{2g}$$

and

$$\overline{c} = \frac{2b + g(2\mu - 1) + \sqrt{(2b - g)^2 - 4g(4b_0 - g)\mu}}{2g}.$$

We prove our result in three steps. (i) First observe that $(2b - g)^2 - 4g(4b_0 - g)\mu < 0$ iff $\mu > \frac{(2b-g)^2}{4g(4b_0-g)}$ and $4b_0 - g > 0$. In this case, the quadratic expression in (A-34) is negative, hence $\Delta W < 0$. (ii) Second, when $\mu < \frac{(2b-g)^2}{4g(4b_0-g)}$ and $4b_0 - g > 0$, we find that $\mu < \underline{c} < \overline{c}$ and that $\frac{\partial \Delta W}{\partial c_{AC}} > 0$ at $c_{AC} = \mu$. Since $\Delta W < 0$ at $c_{AC} = 0$ and ΔW is a continuous function in $c_{AC} \in (0, \mu)$, then $\Delta W < 0$ at any $c_{AC} \in (0, \mu)$. (iii) Finally, when $4b_0 - g < 0$, we find that $\underline{c} < \mu < \overline{c}$ and $\frac{\partial \Delta W}{\partial c_{AC}} < 0$ at $c_{AC} = \mu$. This implies that $\Delta W \leq 0$ when $c_{AC} \leq \underline{c}$ and $\Delta W > 0$ when $\underline{c} < c_{AC} < \mu$.

To prove that $c_{AC} > \underline{c}$ when $\mu \to 0$ and $\max\left\{\frac{2}{3}b + 2b_0, 4b_0\right\} < g$, we proceed as follows. Note that $c_{AC} = \underline{c} = 0$ at $\mu = 0$. Moreover,

$$\lim_{\mu \to 0} \frac{\partial c_{AC}}{\partial \mu} = \frac{6b_0 + 3(2b - g)}{6(2b - g)} > 0 \tag{A-35}$$

and

$$\lim_{\mu \to 0} \frac{\partial \underline{c}}{\partial \mu} = \frac{4b_0 + 2(b - g)}{2b - g} > 0. \tag{A-36}$$

Note that $(A - 35) > (A - 36)$ iff $g > \frac{2}{3}b + 2b_0$, which proves our result. $\blacksquare$

Derivation of the marginal aggregate congestion cost in a fully decentralized

regime. We calculate the marginal aggregate congestion cost imposed by type-0 traveler on fellow travelers in the same lane in a fully decentralized regime when there is no differentiation across lanes (i.e. when $\mu \in (0, \mu']$).

From (37), all travelers face the same congestion level $s_{AC} = f_{AC} = \frac{c_{AC}+1-\mu}{2}$. Aggregate congestion costs are given by $\Gamma \equiv \int_0^1 \theta g \frac{c_{AC}+1-\mu}{2} d\theta$. The marginal congestion cost is then simply given by $\frac{\partial \Gamma}{\partial c_{AC}} = \frac{g}{4}$. When choosing whether to exclude the type-0 from traveling, the planner compares this with the benefit this individual derives from traveling, i.e. $U(0) = b_0$. □

Proof of Proposition 9. When $g \leq g_{FB}$, full coverage occurs not only with atomistic travelers but also in the first best (see (11) and Proposition 2). Since the type-0 traveler gets utility $B(0) - t$ from traveling in the slow lane, Assumptions 1 and 2 implies that full coverage occurs as in the social optimum when $t = t_A$. Given t_A, substitute $\bar{s}_{FB}$ as in (13) into the IC constraint (5) to write

$$B\left(\frac{1}{2} + \frac{\sqrt{7}-2}{6}\right) - \left(\frac{1}{2} + \frac{\sqrt{7}-2}{6}\right) g \left(\frac{1}{2} - \frac{\sqrt{7}-2}{6}\right) - T \geq \\ B\left(\frac{1}{2} + \frac{\sqrt{7}-2}{6}\right) - g\left(\frac{1}{2} + \frac{\sqrt{7}-2}{6}\right)^2 - t_A \qquad \text{(A-37)}$$

and solve it w.r. to T when holding as an equality to obtain T_A.

When $g \geq g_{FB}$, the market is fully covered under atomistic travelers, but not in the social optimum. To obtain t_A, consider that the marginal net effect on social welfare of a θ-type traveler deciding to travel in the slow lane (as opposed to not traveling) is given by the LHS of (9), while his private benefit is given by (1). The difference between (1) and the LHS of (9) is positive and therefore corresponds to the non-internalized component of the marginal social cost, which we set equal to t_A. Given t_A, use s_{FB} and f_{FB} in the IC constraint (5) to write

$$B(1 - f_{FB}) - (1 - f_{FB}) g f_{FB} - T = B(1 - f_{FB}) - (1 - f_{FB} - s_{FB}) g s_{FB} - t_A.$$

and solve it w.r. to T when holding as an equality to obtain T_A. ■

Proof of Proposition 10. Using (55), the Lagrangean of the monopolist problem in (52) is given by

$$\begin{aligned} \mathcal{L}_{CT} \equiv & \left(B(1-s-f) - (1-s-f) gs\right) s + && \text{(A-38)} \\ & + \left(B(1-s-f) - (1-s-f) gs + g(1-f)(s-f)\right) f + \\ & - (gs - z_C) s - (gf - z) f && \text{(A-39)} \\ & - \lambda (s + f - 1). \end{aligned}$$

At the solutions to this problem, denoted by s_{CT}, f_{CT} and γ_{CT}, Kuhn-Tucker conditions require

$$\frac{\partial \mathcal{L}_{CT}}{\partial s} = B\left(1 - s_{CT} - f_{CT}\right) - B'\left(1 - s_{CT} - f_{CT}\right)\left(s_{CT} + f_{CT}\right) + \\ - g s_{CT}\left(2 - 4f_{CT} - 3s_{CT}\right) + z - \lambda_{CT} = 0; \tag{A-40}$$

$$\frac{\partial \mathcal{L}_{CT}}{\partial f} = B\left(1 - s_{CT} - f_{CT}\right) - B'\left(1 - s_{CT} - f_{CT}\right)\left(s_{CT} + f_{CT}\right) + \\ + g\left(3(f_{CT})^2 + 2(s_{CT})^2 - 2f_{CT}\right) + z - \lambda_{CT} = 0; \tag{A-41}$$

$$\frac{\partial \mathcal{L}_{CT}}{\partial \gamma} = s_{CT} + f_{CT} - 1 \leq 0, \qquad \gamma_{CT} \geq 0 \text{ and } \quad \frac{\partial \mathcal{L}_{CT}}{\partial \lambda}\gamma_{CT} = 0. \tag{A-42}$$

Assume that $\lambda_{CT} = 0$ and $s_{CT} + f_{CT} - 1 < 0$. Substituting $\lambda_{CT} = 0$ in (A-40) and (A-41), equalize them and solve w.r.to f to get

$$f_{CT}(s_{CT}) = \frac{1}{3}\left(2\left(1 + s_{CT}\right) - \sqrt{7s_{CT}^2 - 4s_{CT} + 4}\right); \tag{A-43}$$

which is identical to (12). Notice that the level of the subsidy z does not affect $f_{CT}(s_{CT})$ in (A-43).

Assume instead that $\lambda_{CT} \geq 0$ and $s_{CT} + f_{CT} = 1$. Substituting $s_{CT} = 1 - f_{CT}$ in (A-40) and (A-41), equating them and solving w.r.to f_{CT} gives $f_{CT} = \frac{1}{2} - \frac{\sqrt{7}-2}{6}$ as in (13). Using again $s_C = 1 - f_C$ gives $s_C = \frac{1}{2} - \frac{\sqrt{7}-2}{6}$ as in (13). Plugging f_{CT} and s_{CT} thus obtained into (A-40) or (A-41) and solving w.r.to λ_{CT} it gives $\lambda_{CT} = \frac{g\left(4+\sqrt{7}\right)}{B(0)-B'(0)+z}$. Solving $\lambda_{CT} \geq 0$ w.r.to g gives

$$g \leq g_{CT} \equiv \frac{18\left[B(0) - B'(0) + z\right]}{4 + \sqrt{7}}. \tag{A-44}$$

Notice that, when $z = 0$, $g_{CT} < g_{FB}$, so that $g \leq g_{CT}$ implies $g \leq g_{FB}$; that is, whenever the monopolist subject to a system of tax/subsidy as in (55) with $z = 0$ covers the market, a social planner would do it as well. This implies that, when $g \leq g_{CT}$, the subsidy that restores social optimality is equal to zero.

When instead $g > \frac{18[B(0)-B'(0)]}{4+\sqrt{7}}$, a positive subsidy is required for (A-44) to hold. The smallest subsidy is obtained by solving (A-44) w.r.to z when this holds as an equality. This gives $z = B'\left(0\right) - B\left(0\right) + g\,\frac{4+\sqrt{7}}{18}$, which is optimal as long as full coverage is socially optimal, i.e. $g \leq \frac{36\,B(0)}{4+\sqrt{7}}$.

Focus now on $g > \frac{36\,B(0)}{4+\sqrt{7}}$. In this case, the social planner partially covers the market, hence we do not want the monopolist to fully cover the market. The optimal subsidy is then computed as follows. Equalize the FOC w.r.to s or f in the monopolist problem in (53) to the FOC w.r.to s or f in the social planner problem in (9), and solve with respect to z. This gives $z_C = B'(1 - s_{FB} - f_{FB})(s_{FB} + f_{FB}) + g\left(2f_{FB} - (s_{FB})^2 - \frac{3}{2}(f_{FB})^2\right)$. ■

Appendix B: numerical simulations

In this Appendix, we sketch the methodology we use for the numerical welfare analysis in Section 5.4. First, we assume

$$B(\theta) = b_0 + b\theta, \tag{B-1}$$

with $b_0 \geq 0$ and $b > 0$. As mentioned, Assumption 1 and 2 become $g < b$ and $b_0 \geq 0$, respectively. Also, Assumption 3 becomes $b \leq b_0$, which we take never to hold to avoid duplicating the analysis contained in Section 4.4. Without further loss of generality, we normalize $b = 1$. As a result, recalling that, by construction, $\mu \in [0, 1]$, the space of parameters of interest of our analysis is $\{\mu, b_0, g\} \in [0, 1] \times [0, 1) \times (0, 1)$.

We use Maple to perform our numerical analysis. We evaluate and compare welfare under full decentralization ($\mu = 0$) with welfare under partial or full centralization ($0 < \mu \leq 1$).

When $\mu = 0$, we know from Proposition 2 that any allocation of travelers such that $s_A = f_A = \frac{1}{2}$ is an equilibrium and that all equilibria are payoff equivalent. Hence, we allocate travelers with $\theta \in \left[0, \frac{1}{2}\right)$ to line S and travelers with $\theta \in \left[\frac{1}{2}, 1\right]$ to line F and plug these values in (26), which now, because of (B-1), takes the following form

$$W_A = \int_0^{\frac{1}{2}} \left(b_0 + b\theta - \frac{1}{2}\theta g\right) d\theta + \int_{\frac{1}{2}}^{1} \left(b_0 + b\theta - \frac{1}{2}\theta g\right) d\theta. \tag{B-2}$$

When $0 < \mu < 1$, we create a grid of parameters combinations, letting μ vary by 0.1 and letting the other parameters vary by 0.05. For each combination of the triplet $\{\mu, b_0, g\}$, we calculate with numerical methods the equilibrium values of s^c_{AC}, f^c_{AC}, s^a_{AC} and f^a_{AC} and then use these values to calculate social welfare. Details of these calculations are given below:

i) When $\mu \in \left(0, \frac{1}{2}\right]$, we know from Lemma 2 that there is never differentiation across lanes. Full coverage occurs when $g \geq g_{AC}$, where, because of (B-1), $g_{AC} = 2(b - b_0)$.

Full coverage. When $g \geq g_{AC}$, we know from Proposition 7 that an equilibrium is any allocation of travelers such that

$$\begin{gathered} \bar{s}^c_{AC} + \bar{s}^a_{AC} = \bar{f}^c_{AC} + \bar{f}^a_{AC} = \tfrac{1}{2}, \\ \bar{s}^c_{AC} + \bar{f}^c_{AC} = \mu, \\ \bar{s}^a_{AC} + \bar{f}^a_{AC} = 1 - \mu, \end{gathered} \tag{B-3}$$

and that all these allocations are payoff equivalent. Hence, we allocate an equal mass of both atomistic and corporate travelers to each lane and let (both types of) travelers with $\theta \in \left[0, \frac{1}{2}\right)$ travel in lane S and those with $\theta \in \left[\frac{1}{2}, 1\right]$ in lane F. We use these values in (26), which now, because of full coverage, no differentiation and (B-1), takes the following simple form

$$W_{AC} = \int_0^1 \left(b_0 + b\theta - \frac{1}{2}\theta g \right) d\theta. \tag{B-4}$$

Partial coverage. We know from Proposition 6 that

$$\begin{aligned} s^c_{AC} + s^a_{AC} = f^c_{AC} + f^a_{AC} = \tfrac{c_{AC}+1-\mu}{2}, \\ s^c_{AC} + f^c_{AC} = c_{AC}, \\ s^a_{AC} + f^a_{AC} = 1-\mu, \end{aligned} \tag{B-5}$$

We solve numerically for c_{AC} the FOC of the monopolist' problem in (29), which now, because of (B-1), becomes

$$\frac{3g\left(c_{AC}\right)^2}{2\mu} + c_{AC}\frac{g\left(1-2\mu\right)-b}{\mu} + \frac{2\left(b_0 + b\left(1-\frac{c_{AC}}{\mu}\right)\right) - g\left(1-\mu\right)}{2} = 0. \tag{B-6}$$

Then, we allocate an equal mass of both atomistic and corporate travelers to each lane and let (both types of) travelers with $\theta \in \left[1-\left(c_{AC}+\left(1-\mu\right)\right), \frac{1-(c_{AC}+(1-\mu))}{2}\right)$ travel in lane S and those with $\theta \in \left[\frac{1-(c_{AC}+(1-\mu))}{2}, 1\right]$ to lane F. We use these values in (47), which now, because of partial coverage, no differentiation and (B-1), takes the following form

$$\begin{aligned} W_{AC} =& \mu \int_{1-\frac{c_{AC}}{\mu}}^1 \left(b_0 + b\theta - \theta g \frac{c_{AC}+1-\mu}{2} \right) d\theta + \\ &+ \left(1-\mu\right) \int_0^1 \left(b_0 + b\theta - \theta g \frac{c_{AC}+1-\mu}{2} \right) d\theta. \end{aligned} \tag{B-7}$$

ii) When $\mu \in \left(\frac{1}{2}, 1\right)$, we know from Lemma 2 that differentiation may occur depending on the value of μ'. We assume that differentiation occurs, calculate the equilibrium solutions for the monopolist (see below) and check ex-post whether $1-\mu < s^c_{AC} - f^c_{AC}$. If this inequality does not hold, no differentiation occurs and we proceed as detailed in the case of $\mu \in \left(0, \frac{1}{2}\right]$. If it does, we proceed as described below.

Full coverage occurs when $g \geq g_{AC}$ as in (36), that, because of (B-1), becomes $g_{AC} = K(\mu)\left(b - b_0\right)$.

Full coverage. We know from Proposition that in equilibrium $\bar{s}^a_{AC} = 0$ and $\bar{f}^a_{AC} = 1-\mu$. Also, from (21), $\bar{s}^c_{AC}$ and $\bar{f}^c_{AC}$ are given by

$$\begin{aligned} \bar{s}^c_{AC} = \tfrac{1}{2} + \tfrac{\sqrt{4\mu^2-2\mu+1}-2(1-\mu)}{6}, \\ \bar{f}^c_{AC} = \tfrac{1}{2} - \tfrac{\sqrt{4\mu^2-2\mu+1}+4(1-\mu)}{6}. \end{aligned} \tag{B-8}$$

We use these values in (47), which now, because of full coverage, differentiation and

(B-1), takes the following form

$$W_{AC} = \mu \left[\int_0^{1-\frac{\bar{f}^c_{AC}}{\mu}} \left(b_0 + b\theta - g\theta \bar{s}_{AC}\right) d\theta + \int_{1-\frac{\bar{f}^c_{AC}}{\mu}}^1 \left(b_0 + b\theta - g\theta \bar{f}_{AC}\right) d\theta \right] + \\ + (1-\mu) \int_0^1 \left(b_0 + b\theta - g\theta \bar{f}_{AC}\right) d\theta \tag{B-9}$$

Partial coverage. We know from Proposition 6 that

$$s^a_{AC} = 0, \\ f^a_{AC} = 1 - \mu, \tag{B-10}$$

We solve numerically for s^c_{AC} and f^c_{AC} the FOCs of the monopolist' problem in (33) and (34), which now, because of (B-1), become

$$b_0 + b\left(1 - \frac{s^c_{AC} + f^c_{AC}}{\mu}\right) - \frac{b\left(s^c_{AC} + f^c_{AC}\right)}{\mu} + \\ + \frac{g s^c_{AC}\left[3 s^c_{AC} + 4 f^c_{AC} - 2\mu\right]}{\mu} = 0, \tag{B-11}$$

$$b_0 + b\left(1 - \frac{s^c_{AC} + f^c_{AC}}{\mu}\right) - \frac{b\left(s^c_{AC} + f^c_{AC}\right)}{\mu} + \\ + \frac{g\left[2\left(s^c_{AC}\right)^2 + f^c_{AC}\left(4(1-\mu) + 3 f^c_{AC} - 2\right) - (1-\mu)\mu\right]}{\mu} = 0. \tag{B-12}$$

Then, we set $s_{AC} = s^c_{AC}$ and $f_{AC} = f^c_{AC} + f^a_{AC}$. We let corporate travelers with $\theta \in \left[1 - \frac{s^c_{AC}+f^c_{AC}}{\mu}, 1 - \frac{f^c_{AC}}{\mu}\right)$ travel in line S and those with $\theta \in \left[1 - \frac{f^c_{AC}}{\mu}, 1\right]$ travel in line F, together with all atomistic travelers. We use these results in (47), which now, because of partial coverage, no differentiation and (B-1), takes the following form

$$W_{AC} = \mu \left[\int_{1-\frac{f^c_{AC}+f^c_{AC}}{\mu}}^{1-\frac{f^c_{AC}}{\mu}} \left(b_0 + b\theta - g\theta s_{AC}\right) d\theta + \int_{1-\frac{f^c_{AC}}{\mu}}^1 \left(b_0 + b\theta - g\theta f_{AC}\right) d\theta \right] + \\ + (1-\mu) \int_0^1 \left(b_0 + b\theta - g\theta f_{AC}\right) d\theta. \tag{B-13}$$

iii) When $\mu = 1$, full coverage occurs when $g \geq g_C$ as in (45), that, because of B-1), becomes $g_C = \frac{6}{4+\sqrt{3}}\left(b - b_0\right)$.

Full coverage. We know from Proposition 4 that

$$\bar{s}_C = \tfrac{1}{2} + \tfrac{\sqrt{3}}{6} \cong 0.7887, \\ \bar{f}_C = \tfrac{1}{2} - \tfrac{\sqrt{3}}{6} \cong 0.2113. \tag{B-14}$$

We use these values in (47), which now, because of full coverage, differentiation and (B-1), takes the following form

$$W_C = \left[\int_0^{1-\bar{f}_C} (b_0 + b\theta - g\theta \bar{s}_C)\, d\theta + \int_{1-\bar{f}_{AC}}^{1} \left(b_0 + b\theta - g\theta \bar{f}_C\right) d\theta \right]. \tag{B-15}$$

Partial coverage. We solve numerically for s_C and f_C the FOCs of the monopolist' problem in () and (), which now, because of (B-1), become

$$b_0 + b(1 - s_C - f_C) - b(s_C + f_C) + \\ - g s_C (2 - 4f_C - 3s_C) = 0, \tag{B-16}$$

$$b_0 + b(1 - s_C - f_C) - b(s_C + f_C) + \\ g\left[3(f_C)^2 - 2f_C + 2(s_C)^2\right] = 0. \tag{B-17}$$

We let corporate travelers with $\theta \in [1 - s_C - f_C, 1 - f_C)$ travel in line S and those with $\theta \in [1 - f_C, 1]$ travel in line F, together with all atomistic travelers. We use these results in (47), which now, because of partial coverage, no differentiation and (B-1), takes the following form

$$W_C = \int_{1-s_C-f_C}^{1-f_C} (b_0 + b\theta - g\theta s_C)\, d\theta + \int_{1-s_C+f_C}^{1} (b_0 + b\theta - g\theta f_C)\, d\theta.$$

The results of the numerical simulations thus obtained are available here.