Modal choice and optimal congestion*

Quentin David† Renaud Foucart‡

March 25, 2014

Abstract

We study the choice of transportation modes within a city where commuters have heterogeneous preferences for a car. As in standard models of externalities, the market outcome never maximizes aggregate welfare. We show that in the presence of multiple equilibria problems of coordination can worsen this result. We discuss two policy tools: taxation and traffic separation (e.g. exclusive lanes for public transportation). Setting the optimal policy is a necessary but not sufficient condition to maximize aggregate welfare. Even with a social planner maximizing aggregate welfare, a city may find itself stuck in a situation where public transportation remains inefficient and the level of congestion high.

JEL: R4, L5, H2.

Keywords: Modal choice, Coordination, Network effect, Cross-modal congestion.

1 Introduction

The cost of congestion is an increasingly important issue in urban areas. For instance, Duranton and Turner (2011) estimate that a typical American household spends 161 person-minutes in a car every day. Goodwin (2004) expected the annual cost of congestion in the UK to reach £ 30 billion in 2010. De Palma and Lindsey (2011) report congestion costs between 0.5 and 1.5% of GDP

*We would like to thank Rabah Amir, Marco Batarce, Paul Belleflamme, Micael Castanheira, Frederic Dobruzkes, Marjorie Gassner, Marc Ivaldi, Patrick Legros, Moëz Kilani, Georg Kirchsteiger, Martin Koning, Frederic Malherbe, Claire Naiditch, Martin Peitz, Jacques François Thisse, Yves Zenou, two anonymous referees and various conference and seminar participants in Brussels, Luxembourg, Toulouse, Marseille and Istanbul for their helpful comments and suggestions. Quentin David thanks the FNR for financial support.

†EQUIPPE, Université de Lille 1 and iCite, Université libre de Bruxelles. Contact: quentin.max.david@gmail.com

‡Nuffield College and Dept. of Economics, Oxford University. Contact: renaud.foucart@economics.ox.ac.uk

in major United States and European urban areas. Most congestion is due to the use of private cars. On the one hand, cars generate both congestion - on other cars and on public transportation - and pollution. On the other hand, cars are necessary for the economy. Unfortunately, screening commuters to reach the optimal share of cars is a complex policy challenge. Policies must identify tools that affect people's behavior and improve efficiency. In practice, the most widely-used policies addressing traffic issues are taxation,[1] subsidies and traffic separation.[2]

In this paper, we build a theoretical model in which heterogeneous commuters simultaneously decide whether to use a private car or public transportation. Car users generate congestion on all commuters and users of public transportation enjoy a positive network externality. We do not specifically model pollution costs, as this externality affects all commuters regardless of their modal choice, and therefore does not affect this decision. In practice, considering the impact of pollution would lower the socially optimal share of car users obtained with our model.

First, we explain how ex ante similar cities might end up with very different modal shifts. This is a problem of coordination when a large share of commuters have similar preferences. If commuters with preferences close to the mode all take the car, public transportation is not efficient. Therefore, it is indeed a best response for those commuters to take the car. However, if they all use public transportation, it becomes more efficient, and it is a best response for them to use public transportation. In the presence of such multiple equilibria, the one involving the highest share of public transportation always Pareto dominates all others. Second, we show that the market outcome never maximizes aggregate welfare. This is a classic result when externalities are present. Therefore, the market can present two types of inefficiencies. The first one is at the margin: the market provides a too large share of car users in any decentralized equilibrium. The second is more substantial: coordination failures may lead to the presence of inefficient equilibria.

[1] See de Palma and Lindsey (2011) for a survey of the different methods and impacts of congestion tolls.

[2] For instance through the use of exclusive lanes for public transportation. See Cain et al. (2006) and Echeverry et al. (2005) for a larger discussion of the case of Bogota's Transmilenio and its application to other countries. Our definition is more general: any type of infrastructure that increases the marginal cost of congestion for cars and decreases it for public transportation, and whose effect is marginally increasing in the share of car users.

The coordination problem could easily be solved by a benevolent social planner announcing a very high level of taxation if there are too many car users. However, if there are too many car users, this announcement is not credible and the social planner must modify the level of taxation in the interest of the commuters themselves. Thus, the social planner is constrained to announce a credible taxation scheme. To ensure that all announcements are credible, we assume that commuters and the social planner take their decisions simultaneously. The social planner is a player of the game whose objective function is common knowledge. The only way to overcome the commitment problem would be to implement a mechanism where the social planner can credibly commit not to play his best-response.

We consider a social planner using two policies to maximize aggregate welfare: a taxation scheme (which, in our discrete choice setup, is equivalent to a fare subsidy) and traffic separation. Optimal taxation is an increasing function of the share of car users. This is because the marginal congestion externality is also increasing. Under reasonable assumptions, the optimal traffic separation is decreasing in the share of car users. Indeed, it is only a best response for the social planner to implement high levels of traffic separation if she believes enough commuters use public transportation. Due to coordination failures, it is not sufficient to set the optimal policy to reach the optimal outcome. As detailed below, it is well established that policies implemented by a social planner shape the patterns of public transportation use. We show that causality may run both ways: the patterns of transportation also determine the policies taken by a social planner.

The question of optimal congestion has been addressed by many scholars from different fields. Among economists, Pigouvian taxation is generally the preferred way to deal with congestion problems (Beesley and Kemp, 1987, Calfee and Winston, 1998). The idea is that, given both urban layouts and consumers' intrinsic preferences for a car, one should focus on the best way to accommodate traffic flows and make car users pay for the marginal external cost they produce (Anas and Small, 1998). The 'games of congestion' have been largely studied in economic theory (Rosenthal, 1973) and many applied papers deal with congestion costs and car taxation. One of the most famous results is due to Vickrey (1963). He argues that pricing should vary at different times of the day so as to make commuters pay for the marginal cost of congestion. The impact of congestion on public transportation has usually been of minor interest, though some authors (e.g. Mirabel, 1999, Dobruzkes and Fourneau, 2007) addressed the so-called 'crossed modal externalities' (the impact of congestion generated by one mode of transportation on another). Congestion costs have been shown to be convex both in terms of pollution (De Vlieger et al., 2000) and perceived cost (Wardman, 2001).[3]

Another group of papers focuses on urban planning. It emphasizes the fact that the structure of the city is the main driver of commuting patterns. The main idea to improve the performance of urban transportation is to have a shift towards 'transit-oriented development'. Belzer and Aultier (2002) define such a development as follows: 'mixed-use, walkable, location-efficient development that balances the need for sufficient density to support convenient transit service with the scale of the adjacent community'.[4] Some economists indirectly address this dimension by considering a form of traffic separation (see Berglas et al., 1984, Arnott et al., 1992, de Palma and Lindsey, 2002, de Palma et al. 2008). These papers propose various approaches for road pricing and tolls in the presence of alternative roads, modes of transportation and consumer preferences. Their settings differ from ours in various dimensions, but all share the aim to find a unique equilibrium and an optimal policy corresponding to the idea of pricing the marginal externality. In this paper, we show that, in the presence of multiple equilibria, internalizing marginal externalities may not be sufficient.

Multiple equilibria derive from the intersection of congestion, positive externalities from public transportation, and commuters' heterogeneity. A relatively large literature exists on the network

[3]Time is valued 50% higher when spent in congestion. Hence, the cost of congestion is convex, as congestion (i) increases travel time, and (ii) increases the marginal cost of travel time. This principle is applied by Santos and Bhakar (2005) to assess the benefits of the congestion toll in London.

[4]Cervero et al. (2002, p.2) emphasize that it does 'involve some combination of intensifying commercial development around stations, inter-mixing land uses, layering in public amenities (e.g., civic spaces, landscaping), and improving the quality of walking and bicycling'. One should also consider the book by Dittmar and Ohland (2003) that summarizes the literature and 'good practices' in transit oriented development.

effect of the number of transit users on the efficiency of public transportation. In a seminal contribution, Mohring (1972, p.591) explains that 'Transportation differs from the typical commodity price theory texts in that travelers and shippers play a producing, not just a consuming role': The underlying idea is the existence of a so-called 'dynamic network externality'. If the demand for bus service doubles, a company is expected to double the number or buses serving the route at the same per capita price. Thus, the waiting time for an individual commuting by bus decreases, which improves the efficiency of public transportation. The combination of network externalities in public transportation and automobile congestion is a feature of several economic models (Tabuchi, 1993; Parry and Small, 2009). To repeat, those models focus on a unique equilibrium. Commuters differ in their preference for the use of a private car (Beirão and Cabral, 2007; Handy et al., 2005; Jensen, 1999; Steg, 2005; Hiscock et al., 2002 and Van Vught et al., 1996). Verhoef and Small (2004) encompass this dimension by considering heterogeneous agents in a model of pricing for car use only. Batarce and Ivaldi (2011) test this feature in a model of modal choice applied to Santiago, Chile. Fosgereau and De Palma (2012) consider a model of bottleneck with heterogeneous traffic distance, leading to different preferences.

The existence of similar cities characterized by different modal shifts was already documented in the late 1980s by Pucher (1988). He observed that, 'urban transportation and traveler behavior vary widely, even among countries with similar per capita income, technology and urbanization.' Kenworthy and Laube (1999) show that the fraction of workers using transit is six times higher in wealthy Asian cities compared to the US.[5] They also find that the commuting time is lower (and cheaper) where the use of public transport is higher and that the cost recovery of transit increases with the share of passengers using it.

The paper is organized as follows. In the next section, we present the model, show that a Nash equilibrium always exists, give conditions for the existence of multiple equilibria and discuss their relative Pareto efficiency. In Section 3, we derive the best response function of the social planner

[5] Similarly, Pucher and Renne (2003) computed that, in the US, public transport accounted for less than 2% of urban travel in 2001.

and study taxation and traffic separation. We show that in the presence of multiple equilibria, the social planner may be unable to credibly commit to the policy associated to the global optimum. We conclude in Section 4.

2 The model

2.1 Basic assumptions

We consider a closed city with a unit mass of commuters who have to make a discrete choice between using a private car or public transportation. In this section, we do not consider the existence of a social planner. The use of a car generates congestion on the other commuters. Space is finite and it is possible to increase neither the number of roads nor the number of the traffic lanes. The traffic separation is the degree of separation of public transportation from the rest of the traffic, denoted by $\alpha \in [0,1]$.[6] Commuters are heterogeneous as they have different intrinsic preferences for the use of a car (relative to public transportation).[7] The outcome of the game is a share z of car users and $(1-z)$ of public transportation users.

The utility[8] of a commuter i, traveling in a private car or with public transportation, is given, respectively, by

$$U_i^c(\alpha,z) = -f_c - t^c(\alpha,z) + \frac{\varepsilon_i}{2} \tag{1}$$

[6] α is exogenous in this section, but we allow the social planner to choose its level in the next section. We use a very general definition of α, one that encompasses many possibilities to protect public transportation from congestion. The condition being that increasing α decreases congestion for public transportation and increases congestion for car users. This excludes the possibility of building an underground. We discuss this possibility in David and Foucart (2012) and find that an underground is likely to worsen coordination problems. We model α as continuous as (i) there are different degrees of separation (drawing a bus line is not the same as building a bus corridor or reserving some streets for public transportation only) and (ii) traffic separation in a city is the sum of many smaller decisions in given areas that affect the commuter trough is journey.

[7] This preference can be negative. One can imagine various alternative ways of modeling heterogeneity: different valuation for time and money, different location within the city, ease of access to the public transportation network, etc. We use the simplest formulation for the tractability of the model, and do not consider specifically the difference in commuting time for different locations.

[8] All components of the utility functions (fixed costs, congestion, individuals' heterogeneity and waiting time) are expressed in monetary terms.

and

$$U_i^{pt}(\alpha,z) = -W(z) - t^{pt}(\alpha,z) - \frac{\varepsilon_i}{2}. \qquad (2)$$

The fixed cost associated with the use of the car is denoted by $f_c > 0$.[9]

The functions $t^c(\alpha,z)$ and $t^{pt}(\alpha,z)$ ($\in IR^+$) represent the congestion faced by cars and public transportation, respectively. They are assumed to be equal if there is no traffic separation between cars and public transportation, and equal to zero if there are no users of cars (i.e. $t^c(0,z) = t^{pt}(0,z)$ and $t^c(\alpha,0) = t^{pt}(\alpha,0) = 0$ respectively). Both functions are increasing and convex in z, and a higher degree of traffic separation (higher α) generates more congestion for cars (because there is less space for them) and less congestion for public transportation. This last effect is assumed to be amplified by z (that is, separation has an impact only if there is actually a problem of congestion). Hence, we have $\forall(\alpha,z) \in [0,1]^2$:

$$\frac{\partial t^c(\alpha,z)}{\partial z} > 0,\ \frac{\partial^2 t^c(\alpha,z)}{\partial z^2} \geqslant 0,\ \frac{\partial t^c(\alpha,z)}{\partial \alpha} > 0,\ \frac{\partial^2 t^c(\alpha,z)}{\partial z \partial \alpha} > 0$$
$$\frac{\partial t^{pt}(\alpha,z)}{\partial z} > 0,\ \frac{\partial^2 t^{pt}(\alpha,z)}{\partial z^2} \geqslant 0,\ \frac{\partial t^{pt}(\alpha,z)}{\partial \alpha} < 0,\ \frac{\partial^2 t^{pt}(\alpha,z)}{\partial z \partial \alpha} < 0.$$

The individual parameter, ε_i, is the preference for using a car compared to public transportation. It comes from a cumulative distribution function $\varepsilon_i \sim F(\varepsilon)$. F is assumed to be strictly increasing, continuous and differentiable over its support $(-\infty,+\infty)$. This support implies that some individuals love public transportation so much that they would never accept not to use it ($\varepsilon_i \to -\infty$), while others will never use public transportation ($\varepsilon_i \to +\infty$). Without loss of generality, we split ε_i equally between the two utility functions. Later in the paper, it will be useful to use the inverse distribution function of $F(x)$. We define $G(x) = -F^{-1}(x)$ where $x \in [0,1]$ is the unique real number x such that $F(\varepsilon) = x$. This allows us to use a function $G(x) = \varepsilon_i$ such that there is a mass x of commuters with $\varepsilon > \varepsilon_i$ (with $G(0) = +\infty$, $G(1) = -\infty$).

[9]This is the additional cost compared to the use of public transportation, which is normalized to 0.

The waiting time for public transportation is $W(z) \in \mathbb{R}_0^+$. It displays a positive network externality for public transportation users. The idea is that, if there are more users, the frequency of public transportation increases and the waiting time decreases.[10] For simplicity, we assume this network externality to be linear. If there are $(1-z)$ users of public transportation, the waiting time of each of them is given by $W(z)$, with

$$W'(z) > 0 \text{ and } W''(z) = 0.$$

Definition 1 $\Delta(\alpha, z)$ *is the additional congestion faced by car users in comparison to the congestion faced by public transportation, i.e.*

$$\Delta(\alpha, z) = t^c(\alpha, z) - t^{pt}(\alpha, z).$$

Using the properties of $t^c(\alpha, z)$ and $t^{pt}(\alpha, z)$, we have:

Lemma 1 *Properties of* $\Delta(\alpha, z)$.
(i) $\frac{\partial \Delta(\alpha,z)}{\partial \alpha} > 0$*,* $\forall (\alpha, z) \in [0,1] \times]0,1]$;
(ii) $\frac{\partial \Delta(\alpha,z)}{\partial z} > 0$*,* $\forall (\alpha, z) \in]0,1] \times [0,1]$;
(iii) Supermodularity of $\Delta(\alpha, z)$***:*** *the effect of separation on the differential of commuting time increases with congestion (with the number of car users), i.e.* $\frac{\partial^2 \Delta(\alpha,z)}{\partial z \partial \alpha} > 0$.

Proof. (i) is straightforward from the properties of $t^c(\alpha, z)$ and $t^{pt}(\alpha, z)$. Property (ii) is derived from the fact that as $t_c(0,z) = t_{pt}(0,z)$, $\frac{\partial \Delta(0,z)}{\partial z} = 0$. And as, $\forall (\alpha, z) \in]0,1]^2$, $\frac{\partial t^c(\alpha,z)}{\partial z} > \frac{\partial t^c(0,z)}{\partial z}$ and $\frac{\partial t^{pt}(\alpha,z)}{\partial z} < \frac{\partial t^{pt}(0,z)}{\partial z}$, it is always true that $\frac{\partial \Delta(\alpha,z)}{\partial z} > 0$. Property (iii), the supermodularity of $\Delta(\alpha, z)$ is obtained using $\frac{\partial^2 t^c(\alpha,z)}{\partial z \partial \alpha} > 0$ and $\frac{\partial^2 t^c(\alpha,z)}{\partial z \partial \alpha} < 0$. The definition of $\Delta(\alpha, z) = t^c(\alpha, z) - t^{pt}(\alpha, z)$ leads to: $\frac{\partial^2 \Delta(\alpha,z)}{\partial z \partial \alpha} = \frac{\partial^2 t^c(\alpha,z)}{\partial z \partial \alpha} - \frac{\partial^2 t^{pt}(\alpha,z)}{\partial z \partial \alpha} > 0$. ■

[10] Assume a public transportation provider collecting fees and providing costly quality given a binding budget constraint. As long as there are scale economies in the production function, an increased number of public transportation users increases the quality for a given individual fee, or decreases the individual fee for a given level of quality. The possible presence of discomfort externalities and/or capacity constraints in public transportation is studied in David and Foucart (2012). We find that this does not solve - and may even worsen - the coordination problems.

2.2 The game

The modal choice is a simultaneous game among a unit mass of commuters. It consists of each commuter choosing the mode of transportation (car or public transportation) that maximizes his utility given her expectation on z. Hence, commuter i commutes by car if $U_i^c(\alpha,z) > U_i^{pt}(\alpha,z)$, i.e.

$$\varepsilon_i > f_c - W(z) + [t^c(\alpha,z) - t^{pt}(\alpha,z)].$$

If it is a best response *ex post* for a commuter j with $\varepsilon_j > \varepsilon_i$ to commute using public transportation, it is also a best response for commuter i to do so.

Using Definition (1), the condition for commuter i to use a car becomes

$$\varepsilon_i > f_c - W(z) + \Delta(\alpha,z). \tag{3}$$

2.3 Decentralized Equilibria

In this section, we first show that a Nash equilibrium always exists. Second, we derive the conditions for the presence of multiple equilibria. Third, we show that one is Pareto dominant.

a. Existence

The existence of at least one Nash equilibrium where all commuters play a pure strategy is relatively easy to show.[11] Stability *ex post* comes from the fact that there always exists an equilibrium where a share of commuters strictly prefers public transportation while the other prefers to use a car.

Lemma 2 *There exists at least one Nash equilibrium in pure strategies.*

Proof. Remember $F(\varepsilon)$ is assumed to be continuous and differentiable over its support $(-\infty,+\infty)$. This implies that there exists at least one commuter k with taste parameter ε_k such that, if all

[11] In general, it as been shown that such an equilibrium always exists in games with a continuum of players (Mas Colell, 1984 and Rath, 1992).

commuters with parameter $\varepsilon_j < \varepsilon_k$ use public transportation, and all commuters with $\varepsilon_k < \varepsilon_i$ take the car,

$$G(z_k) = f_c - W(z_k) + \Delta(\alpha, z_k). \tag{4}$$

Commuter k is indifferent between the private car and public transportation. Sharing the same beliefs, commuters with $\varepsilon_j < \varepsilon_k$ strictly prefer public transportation and $\varepsilon_k < \varepsilon_i$ strictly prefer their car. Thus, it is a Nash equilibrium. ■

b. Multiplicity

The intuition behind the existence of multiple equilibria is the following. Assume that there is a large share of commuters with similar preferences (ε) for the use of a car. When they believe that most of them use public transportation, it is a best response for them to do so. This is a Nash equilibrium with a low share of car users z. If, on the contrary, most of them believe that they will use a car, they expect public transportation not to be efficient and, indeed, it will not be. This is also a Nash equilibrium, involving a high z.

Proposition 1 *There exist multiple equilibria (for a given α) if and only if there exists at least a solution z_k such that*

$$\left.\frac{\partial G(z)}{\partial z}\right|_{z=z_k} > \left.\frac{\partial [f_c - W(z) + \Delta(\alpha, z)]}{\partial z}\right|_{z=z_k}. \tag{5}$$

Proof. The proof is presented in Appendix 1.1. ■

For this condition to be fulfilled, the difference in the costs between the two modes of transportation must be sufficiently low and a sufficiently high mass of commuters must have similar preferences. Consider the particular case of unimodal preferences. When the density around the mode is high, the decision of the large share of commuters with similar preferences has a high impact on the quality of public transportation and the level of congestion. If those commuters all take public transportation, public transportation will be efficient. If they all take the car, public transportation will be of bad quality. This is likely to lead to the presence of three equilibria, as plotted in Figure 1 (with $\alpha = 0$). There are two stable[12] equilibria, one with few users of public transportation (a share z_1 of car users) and one with a large fraction (a share $z_3 < z_1$ of car users). There is also one unstable equilibrium, z_2.

INSERT FIGURE 1 ABOUT HERE

(Illustration with multiple equilibria)

c. Efficiency

Proposition 2 *If there are multiple equilibria, the equilibrium involving the highest use of public transportation Pareto dominates all the other equilibria. The Pareto dominant equilibrium is denoted $\widehat{z}$.*

Proof. The formal proof is provided in Appendix 1.2. ■

Figure 1 illustrates this proposition. Define three groups of people as A, B and C. Group A uses a car in both equilibria, group C uses public transportation in both equilibria, and group B uses public transportation when $z = z_3$ and a car otherwise. As the costs of both public transportation and car use are lower in z_3, groups A and C are strictly better off. By revealed preferences, group B is also better off in that equilibrium: they are better off by using a car in z_3 than in z_1, but they use public transportation instead.

3 Social planner

In this section, we identify the optimal outcomes of this game in terms of traffic separation (α) and share of car users (z). To reach an optimum, the social planner has two policy tools at her disposal: traffic separation and taxation. To ensure the credibility of the announced policy, we consider that

[12]Those equilibria are locally stable in the sense that agents' best response to any small perturbation to the equilibrium z would bring this share back to equilibrium.

the policy choice is simultaneous to the modal choice.[13] The timing of the game reflects the well-known fact that governments face commitment problems in taxation. Consequently, as formalized by Kydland and Prescott (1977, p. 473),[14] a government maximizing aggregate welfare may lead to a wrong outcome.

While traffic separation is directly chosen by the social planner, the share of car users is the outcome of the modal choice of commuters. To be as general as possible, rather than setting a tax level, the social planner announces a taxation scheme. This scheme can be made conditional on the realized level of congestion $T(z)$.[15]

The game of modal choice suffers from two types of inefficiencies. The first one, *sub-utilization of public transportation* implies that, at the margin, the first best equilibrium requires more users of public transportation at any initial Nash equilibrium. This would be easily solved by setting the optimal policy in absence of the second inefficiency: *coordination failure.* When the optimal policy leads to multiple equilibria, the right policy can lead to a wrong outcome.

3.1 Policy tools

We assume that the government has two policy tools at her disposal: the taxation of car users ($T(z)$), and the possibility to change the traffic separation between cars and public transportation (α). Due to the discrete choice nature of the model, a tax is equivalent to a fare subsidy.[16] We do not consider variations of taxation schemes that can have differential effects among car users or time of the day.[17] An illustration of the two policies is presented in Figure 2. On both sides of the figure, the benchmark is the game without intervention. On the left-hand side, we present

[13] A sequential game where first, the social planner announces a policy, second, commuters make their modal choice and third, the social planner can revise the policy would lead to the same results.

[14] "Even if there is an agreed-upon, fixed social objective function and policymakers know the timing and magnitude of the effects of their actions, discretionary policy, namely, the selection of that decision which is best, given the current situation and a correct evaluation of the end-of-period position, does not result in the social objective function being maximized."

[15] As we identify coordination issues with the most general taxation function, this also holds for simpler and more realistic ones.

[16] An alternative policy would be to allow the social planner to invest in lower W for a given value of z.

[17] One can refer to Parry (2002) for a comparison between a single lane toll, a uniform congestion tax across freeway lanes, a gasoline tax, and a transit fare subsidy for the reduction of congestion.

the effect of a fixed tax level $T(z) = T > 0$. The right-hand side represents the effect of traffic separation $\alpha > 0$.

We assume that a taxation policy would levy a tax $T(z)$ on every car user and that this tax is redistributed lump-sum among all commuters.[18] Therefore, every commuter receives a transfer $zT(z)$. The utility functions become:

$$\begin{aligned} U_i^c(\alpha, T, z) &= -f_c - t^c(\alpha, z) - (1-z)T(z) + \frac{\varepsilon_i}{2}, \\ U_i^{pt}(\alpha, T, z) &= -W(z) - t^{pt}(\alpha, z) + zT(z) - \frac{\varepsilon_i}{2}. \end{aligned}$$

After the introduction of a taxation policy, a commuter i uses public transportation if and only if

$$\varepsilon_i < \Delta(\alpha, z) - W(z) + T(z) + f_c.$$

INSERT FIGURE 2 ABOUT HERE

(Effect of taxation and traffic separation)

[18] Assuming alternative uses of the tax, ranging from throwing away its product to investing in public transportation infrastructure, would not qualitatively affect the results.

3.2 Maximizing welfare

The objective function of the social planner is to maximize the sum of all commuters' utilities,[19] and is common knowledge, i.e.

$$
\begin{aligned}
& \underset{\alpha,z}{Max} \int_0^1 \left[\phi U^c\left(\alpha,z,G(x)\right) + (1-\phi)\, U_i^{pt}\left(\alpha,z,G(x)\right)\right] dx \\
\text{such that } \phi &= 1 \text{ if } G(x) \geq G(z) \\
&= 0 \text{ otherwise} \\
(\alpha,z) &\in [0,1]^2
\end{aligned}
$$

As $T(z)$ is a lump-sum transfer among commuters, the taxes paid and received cancel out, and the aggregate utility maximization problem rewrites:

$$
\max_{\alpha,z}\left[\int_0^z \left(-f_c - t^c(\alpha,z) + \frac{G(x)}{2}\right) dx + \int_z^1 \left(-W(z) - t^{PT}(\alpha,z) - \frac{G(x)}{2}\right) dx\right].
$$

We first derive the optimality condition in terms of outcomes α^* and z^*. Then, we consider the corresponding optimal policy tools α^* and $T^*(z)$. The first order conditions are[20]

$$
G(z^*) = f_c - W(z^*) + \Delta(\alpha,z^*) + zt_z^c(\alpha,z^*) + (1-z^*)\left[W'(z^*) + t_z^{PT}(\alpha,z^*)\right] \quad (6)
$$

and

$$
zt_\alpha^c(\alpha^*,z) + (1-z)\, t_\alpha^{PT}(\alpha^*,z) = 0, \quad (7)
$$

if there is an interior solution for α^*. Otherwise, if either $zt_\alpha^c(\alpha,z) + (1-z)\, t_\alpha^{PT}(\alpha,z) > 0$ or $zt_\alpha^c(\alpha,z) + (1-z)\, t_\alpha^{PT}(\alpha,z) < 0\ \forall \alpha \in (0,1)$, there is a corner solution. This case is discussed just below lemma 3. As G is continuous, $G(0) = +\infty$ and $G(1) = -\infty$, equation (6) has only interior

[19] We present in David and Foucart (2012) the conditions for policies to be Pareto improving instead of maximizing aggregate welfare. The main intuitions behind our results hold, although a strictly positive level of taxation is not always Pareto improving.

[20] The partial derivatives with respect to z and α are denoted by t_z and t_α.

solutions. Rearranging the terms to compare the private costs and the public benefits for commuter $i : G(z^*) = \varepsilon_i$, equation (6) leads to the following condition:

$$W(z^*) - \Delta(\alpha, z^*) - f_c + G(z^*) = z^* t_z^c(\alpha, z^*) + (1 - z^*) \left[W'(z^*) + t_z^{PT}(\alpha, z^*)\right]. \tag{8}$$

Thus, we have:

Proposition 3 *In any Nash equilibrium, the share of car users is too high. For the share of car users to be socially optimal, there must exist public transportation users that strictly prefer the car in the decentralized equilibrium.*

Proof. The right-hand side of equation (8) corresponds to the marginal social cost of increasing the share of car users. This is clearly positive, as all negative externalities of a car are increasing with z. On the left hand side is the individual preference for the car of the swing commuter z^*, such that all commuters with $\varepsilon < G(z^*)$ take public transportation and the rest use a car. For the equality to hold, this must be positive. This implies that the socially optimal swing commuter strictly prefers to use a car rather than public transportation. Indeed, by equation (4), for the consumer to be indifferent, it must be true that $G(z) = f_c + \Delta(\alpha, z) - W(z)$. As here, $G(z)$ is higher, it means that the commuter prefers the car. ■

This result is classic in the presence of externalities and recalls the literature on Pigouvian taxation. The condition in (8) gives an insight into what can be socially optimal to solve the *sub-utilization of public transportation* at the margin: in any decentralized state of the world, there are not enough users of public transportation.

Lemma 3 *For any level of z, the optimal level of traffic separation (α^*) is the one that minimizes total congestion.*

Proof. Straightforward from equation (7). ■

One can go further on the interpretation of the relationship between z and optimal traffic separation α^*. In the presence of an interior solution, the second first order condition gives:

$$\frac{z}{(1-z)} = -\frac{t_{\alpha}^{PT}(\alpha^*,z)}{t_{\alpha}^{c}(\alpha^*,z)}. \tag{9}$$

The right-hand side of equation (9) is positive and represents a measure of the relative efficiency of a traffic separation policy. As $t_{\alpha}^{c} > 0$ and $t_{\alpha}^{PT} < 0$, when z increases, an intuitive interpretation is that the policy should be "accommodating": α^* has to decrease for the equality to hold.[21] When the share of car users is high, the best response of a social planner is to set a low level of traffic separation. However, with a low level of traffic separation, commuters are incentivized to use the car.

If there is no interior solution and if the former intuition holds, the social planner chooses $\alpha^* = 0$ if z is sufficiently high, and $\alpha^* = 1$ if z is sufficiently small. If this intuition does not hold, the result is the exact opposite.

3.3 Optimal policies and outcome

We define (α^*, z^*) as an outcome simultaneously satisfying the two first order conditions (equations 6 and 7). As for the decentralized equilibrium, several pairs (α^*, z^*) may satisfy these conditions. If this is the case, the global maximum is denoted by $(\bar{\alpha}^*, \bar{z}^*)$.

Proposition 4 *The optimal outcome $(\bar{\alpha}^*, \bar{z}^*)$ can be sustained by an optimal taxation scheme $\bar{T}^*(z) = T^*(\bar{\alpha}^*, z)$ corresponding to the social marginal impact of the use of a car when $\alpha = \bar{\alpha}^*$:*

$$T^*(\bar{\alpha}^*, \bar{z}^*) = \bar{z}^* t_{\bar{z}}^{c}(\bar{\alpha}^*, \bar{z}^*) + (1-\bar{z}^*)\left[W'(\bar{z}^*) + t_{z}^{PT}(\bar{\alpha}^*, \bar{z}^*)\right]. \tag{10}$$

[21] Formally, this is not always the case, as we do not have specific assumption on the sign of $\beta_{\alpha}(\alpha,z)$, with $\beta(\alpha,z) = -\frac{t_{\alpha}^{PT}(\alpha,z)}{t_{\alpha}^{c}(\alpha,z)}$. It is reasonable to believe that $\beta_{\alpha}(\alpha,z) \geq 0 \ \forall \ z \in [0,1]$. In line with Tabuchi (1993), this implies that the impact of bottlenecks is not marginally increasing. Indeed, on the one hand, by increasing the share of roads dedicated to public transportation, the incidence of congestion on public transportation is reduced proportionally. On the other hand, the creation of dissociated traffic lanes for public transportation generates bottlenecks for cars. The creation of these bottlenecks is likely to increase congestion but at a marginally decreasing rate (by increasing the number of bottlenecks, the impact of each one is reduced).

Proof. The proof derives directly from the result obtained in equation (8). The optimal level of taxation $\bar{T}^*(\bar{z}^*)$ is such that the optimal share of car users holds in equilibrium. This equilibrium is locally stable, as from equation (10), for any $z' < \bar{z}^*$, $\bar{T}^*(z') < \bar{T}^*(\bar{z}^*)$ and for any $z'' > \bar{z}^*$, $\bar{T}^*(z'') > \bar{T}^*(\bar{z}^*)$ ■

The marginal social impact of car use in $\bar{z}^*$ (the right hand side of equation 10) is the sum of the social marginal congestion for cars and public transportation and of the social marginal opportunity cost in terms of the network externality of car users not using public transportation.

The optimal tax defined in equation (10) is increasing in z (i.e. $\frac{\partial T^*(\bar{\alpha}^*,z)}{\partial z} \geq 0$). This means that, comparing two similar cities, if the optimal share of car users is higher in one of the two cities, the level of taxation in that city must also be higher. This result is due to the marginal cost of car use (both in terms of congestion and in terms of network externalities) which is increasing in the share of car users.

Proposition 5 *If there exists z_k such that*

$$G(z_k) = f_c + \bar{T}^*(z_k) - W(z_k) + \Delta(\bar{\alpha}^*, z_k) \tag{11}$$

and

$$\left.\frac{\partial G(z)}{\partial z}\right|_{z=z_k} > \left.\frac{\partial [f_c + \bar{T}^*(\bar{\alpha}^*, z) - W(z) + \Delta(\bar{\alpha}^*, z)]}{\partial z}\right|_{z=z_k}, \tag{12}$$

the optimal policy ($\bar{\alpha}^*, \bar{T}^*(z)$) is associated with multiple equilibria. Thus no credible policy ensures to achieve the optimal outcome $(\bar{\alpha}^*, \bar{z}^*)$.

Proof. The conditions (11) and (12) follow directly from equations (4) and (5), in presence of the optimal policy. Consider the example in Figure 3. Assume that the optimal outcome $(\bar{\alpha}^*, \bar{z}^*)$ is associated with an optimal policy ($\bar{\alpha}^*, \bar{T}^*(z)$). Consider a social planner publicly announcing that she will set the globally optimal policy ($\bar{\alpha}^*, \bar{T}^*(z)$). This policy may still lead to multiple equilibria. The social planner may as well announce another taxation function, with higher levels of T for highest levels of z to deter commuters to take the car. She can also announce a higher level of α.

However, this announcement is not credible: if the beliefs are such that most commuters take the car, it is not a best response for the social planner to implement a too high level of taxation and/or traffic separation. Hence, if the beliefs of the social planner are such that commuters coordinate towards the bad equilibrium, her best response is to set the policy (α^*,$T^*(z)$) which is locally optimal at this equilibrium and distinct from ($\bar{\alpha}^*, \bar{T}^*(z)$).[22]

INSERT FIGURE 3 HERE

(An illustration of an optimal policy leading to a wrong outcome) ■

A first problem is thus that, even if the social planner could credibly commit to $T^*(z)$ regardless of the actual value of z, this may not be enough to coordinate commuters to the good equilibrium. The second problem, more general, is that the social planner cannot credibly commit to a policy that is not her best response given her (common knowledge) objective function. Indeed, to force commuters to coordinate towards the best equilibrium, the social planner could announce an extremely high tax for any $z > z^*$. However, if the beliefs of commuters are such that z would, indeed, be higher than z^*, it is not a best response for the social planner to implement it. This is precisely because the social planner is known to be willing to maximize aggregate welfare[23] that he may be unable to do so. So, the fact that some cities may end up in different equilibria does not need to derive from social planners with different objective functions. Ex ante identical social planners in cities with ex ante identical commuters may end up taking very different policies. In particular, a social planner maximizing aggregate welfare may end up in a city where aggregate welfare is not maximal.

[22]To repeat, while $T^*(\alpha,z)$ is unique, $T(\alpha,z) \neq T(\alpha',z)$ when $\alpha \neq \alpha'$. Hence, $T^*(z)$ is not unique.

[23]One could consider a politician seeking reelection as an example of such a common knowledge objective function. If everyone believes that the commuters coordinate towards the bad equilibrium, the politician does not implement a policy that is likely to hurt a majority of voters. Therefore, the commuters actually coordinate towards the bad equilibrium.

3.4 Discussion

Just like central banks for monetary policies, an intuitive solution to the commitment problem is to delegate policy tools. We show in Appendix 2 that delegation to a fully independent agency allows reaching the best outcome in our specification. However, its practical implementation presents important limitations. First, the agency must be credible in being independent from the social planner and its objective function must be common knowledge. Second, it must be credible in implementing a taxation function that deters multiple equilibria. Taxation must be high enough to convince most car users to switch to public transportation even when W is very high. A social planner, in particular if she is a politician seeking reelection, must be convinced that voters make the difference between her and the agency. Also, the transition may be very costly. Not only because there may be some switching costs in practice such that the socially too high level of taxation may have to be implemented for some time. But also because if commuters take their commuting decisions based on the history of what have been the past equilibrium (as in Young (1993, 1996)), it may take several periods of socially too high taxation before the optimal equilibrium is reached. Therefore, such an agency must not only be independent, but also have long mandates. In addition, the social planner delegating the policy tools must not discount the future too much.

As shown in Proposition (5), the right policy mix can lead to the wrong equilibrium. In addition, real world constraints may make the setting of an individual tax whose level is determined by the realized aggregate outcome difficult to implement. In practice, congestion tolls may vary throughout the day or evolve over time, but we are not aware of any taxation scheme made conditional to the aggregate outcome. Such a tax whose amount is unknown ex ante would create too much uncertainty for commuters. Considering a more realistic taxation scheme such as a fixed tax whose level would be announced before the game of modal choice takes place can only worsen the coordination problems we identify. Such a tax constitutes a subset of the general taxation function considered throughout the paper.

For practical reasons, it can be difficult to simultaneously implement a policy of both taxation and traffic separation. Considering a fixed tax level, it is possible to evaluate the respective merits of taxation and traffic separation using a very general criterion: their relative ability to lead to Pareto improvements. We define $\beta(\alpha,z) = -\frac{t_{\alpha}^{PT}(\alpha,z)}{t_{\alpha}^{c}(\alpha,z)}$ in Section 3.2 as the marginal effect of α on the relative commuting time ratio. Assume that $\beta_{\alpha}(\alpha,z) \geq 0$. As discussed above, this corresponds to the intuition that the impact of bottlenecks on the relative marginal congestion is not increasing.

Proposition 6 *Assume there exist two distinct sets of policies, a policy of traffic separation* $(\alpha_1, 0)$ *and a policy of taxation* $(0, T_1)$*, that yield the same equilibrium,* z_1*, and that car users enjoy the same utility under either of these two policies. Then, for any other two distinct sets of policies,* $(\alpha_2, 0)$ *and* $(0, T_2)$*, yielding another equilibrium,* z_2*, associated to a lower share of car user (*$z_2 < z_1$*), the policy of traffic separation Pareto dominates the policy of taxation as car users are better off.*

Proof. *See Appendix 1.3.* ■

As α and T are higher and z is lower, public transportation users are better off in z_2. So, for Pareto efficiency, we need to focus on car users only. It follows from this Proposition that even though we cannot theoretically exclude the possibility that one of the two policies is always better than the other, if this is not the case, taxation should be preferred for small changes in z, while separation should be preferred for larger changes. This relates to the two schools of thought we presented in the literature review. If a social planner is convinced that the city is car-dependent, and that any policy can only have a marginal impact on the modal split, then a policy of taxation may be the best policy. But, if one believes that a large shift can take place, traffic separation is a better choice.

4 Conclusion

We show that the combination of cross modal congestion and network externalities with heterogeneous commuters can lead to multiple equilibria. This explains why a priori similar cities might end up with very different patterns of car use. Some are characterized by an efficient public transportation network and a low degree of congestion, others face inefficient public transportation and high congestion.

We consider a social planner seeking to maximize the total welfare. She has two policy tools at her disposal: traffic separation - which consists of separating cars and public transportation to reduce the congestion generated by the former on the latter - and taxation. We show that both of them have to be used to reach the first best. In addition, even considering the possibility for the central planner to define the taxation as a function of the share of car users, setting the optimal policies is a necessary but not sufficient condition to reach the optimal level of congestion. In some circumstances, a city can be stuck in a situation where the use of public transportation remains inefficient even if the optimal policies are implemented.

We identify two sources of inefficiencies. The *sub-utilization of public transportation* implies that, at the margin, more commuters should use public transportation. This inefficiency can be easily solved. The *coordination failure* suggests that commuters coordinate on a bad, Pareto dominated equilibrium. In the presence of these multiple equilibria, the coordination issue may be impossible to overcome by the social planner. Her incapacity comes from the very fact that she is known to be maximizing total welfare. If the beliefs are such that commuters are stuck in a bad equilibrium, her best response is to take the policy decision that maximizes total welfare given the decisions of commuters. Therefore, it is not only the decisions of a social planner that influence the way people commute in a city. It is also the way people commute, and more importantly how they expect others to commute, that determines the policy decisions.

Bibliography

Anas, A., Arnott, R. and K. Small (1998). "Urban Spatial Structure" ; *Journal of Economic Literature ;* 36 ; 3 ; pp. 1426-1464

Arnott, R & de Palma, A and R Lindsey (1992). "Route choice with heterogeneous drivers and group-specific congestion costs," Regional Science and Urban Economics, Elsevier, vol. 22(1), pages 71-102, March.

Batarce, M and M Ivaldi (2011). "Travel Demand Model with Heterogeneous Users and Endogenous Congestion: An application to optimal pricing of bus services" ; *IDEI Working Papers,* Toulouse

Beesley, M and M Kemp (1987). "Urban Transportation" ; *Handbook of Regional and Urban Economics;* vol. 2 ; chpt. 26 ; pp. 1023-1052

Beirão, G and Cabral, S (2007). "Understanding attitudes towards public transportation and private car: A qualitative study" ; *Transport Policy ;* 14 ; pp. 478-489

Belzer, D and G Aultier (2002). "Transit oriented development: moving from rhetoric to reality" ; *Brookings Institutions Center on Urban Metropolitan Policy*; Discussion paper ; 55pp.

Benhabib, J. and Rustichini, A. (1997). Optimal taxes without commitment.Journal of Economic Theory, 77(2), 231-259.

Berglas, E, Fresko, D and D Pines (1984). “Right of Way and Congestion Toll”, Journal os Transport Economics and Policy, Vol. 18, N°2, pp. 165-187.

Calfee, J and C Winston (1998). "The value of automobile travel time: implications for congestion policy"; *Journal of Public Economics ;* 69 ; pp.83-102

Cain A, Darido G, Baltes M, Rodriguez P and J Barrios (2006). "Applicability of Bogotá’s TransMilenio BRT System to the United States", *Federal Transit Administration Working Paper*

Cervero R., Ferrell C and S Murphy. (2002). "Transit-Oriented Development and Joint Development in the United States: A Literature Review" ; *TCRP Research Results* Digest Number 52 ; National Research Council: Washington, D.C.

David, Q. and Foucart, R. (2012). Modal choice and optimal congestion (No. 12-03). Center for Research in Economic Analysis, University of Luxembourg.

de Palma, A and R Lindsey (2002). "Private roads, competition, and incentives to adopt time-based congestion tolling," Journal of Urban Economics, Elsevier, vol. 52(2), pages 217-241, September.

de Palma, A, Kilani, M and R Lindsey (2008). "The merits of separating cars and trucks," Journal of Urban Economics, Elsevier, vol. 64(2), pages 340-361, September.

De Palma A and R Lindsey (2011). "Traffic congestion pricing methodologies and technologies", *Transportation Research Part C*

De Vlieger I, De Keukeleere D and JG Kretzschmar. (2000). "Environmental effects of driving behaviour and congestion related to passenger cars" ; *Atmospheric Environment ;* 34 ; pp.4469-4655

Dittmar H and G Ohland (2003). "The new transit town: best practices in transit-oriented development", *Island Press*.

Dobruszkes F and Y Fourneau (2007). "Coûts directs et géographie des ralentissements subis par les transports publics bruxellois" ; *Brussels Studies* ; N°7.

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

Echeverry JC, Ibáñez AM, Moya A, Hillón LC, Cárdenas M and A Gómez-Lobo (2005). "The Economics of TransMilenio, a Mass Transit System for Bogotá", *Economía*, Vol. 5, No. 2 (Spring, 2005), pp. 151-196

Fosgereau, M and A de Palma (2012). “Congestion in a city with a central bottleneck”, *Journal of Urban Economics*, 71, 3, 269-277

Goodwin, P (2004). "The Economic Costs of Road Traffic Congestion", ESRC Transport Studies Unit, University College London Working Paper, 26pp

Handy S, Weston L and P Mokhtarian (2005). "Driving by choice or necessity?", *Transportation Research Part A: Policy and Practice* ; pp. 183-203

Hiscock R, Macintyre S, Kearns A and A Ellaway (2002). "Means of transport and ontological security: Do cars provide psycho-social benefits for their users ?" ; *Transportation Research part D* ; 7 ; pp. 1119-1135

Jensen M (1999). "Passion and Earth in transport: a sociological analysis on travel behaviour" ; *Transport Policy* ; 6 ; pp. 19-33

Kenworthy J and F Laube (1999). "Patterns of automobile dependance in cities: an international overview of key physical and economic dimensions with some implications for urban policy" ; *Transportation Research part A* ; 33 ; pp. 691-723

Kydland, F. E., and E.C. Prescott (1977). “Rules rather than discretion: The inconsistency of optimal plans”; *The Journal of Political Economy*, 473-491.

Mas-Colell A (1984). "On a Theorem of Schmeidler" ; *Journal of Mathematical Economics*, 13, 201-206

Mirabel F (1999). "Répartitions modales urbaines, externalités et instauration de péages: Le cas des externalités de congestion et des 'externalités modales croisées" ; *Revue économique* ; 50 ; 5 ; pp. 1007-1027

Mohring H (1972). "Optimization and Scale Economies in Urban Bus Transportation" ; *The American Economic Review* ; 62 ; 4 ; pp. 591-604

Parry I. (2002). "Comparing the efficiency of alternative policies for reducing traffic congestion" ; *Journal of Public Economics* ; 85 ; 3 ; pp. 333-362

Parry I and K Small (2009). "Should Urban Transit Subsidies Be Reduced ?" ; *The American Economic Review* ; 99 ; 3 ; pp. 700-724

Pucher J (1988). "Urban Travel Behavior as the Outcome of Public Policy: The Example of Modal-Split in Western Europe and North America", *Journal of the American Planning Association*, Vol. 54, No. 4, pp. 509-520

Pucher J and J Renne (2003). "Socioeconomics of Urban Travel: Evidence from the 2001 NHTS" ; *Transportation Quarterly* ; 57 ; 3 ; pp. 49-77

Rath KP (1992). "'A direct proof of the existence of pure strategy equilibria in games with a continuum of players"' ; *Economic Theory*, 2:427-433.

Rosenthal R (1973). "A class of Games Possessing Pure-Strategy Nash Equilibria" ; *International Journal of Game Theory* ; 2 ; 1 ; pp. 65-67

Santos G and J Bhakar (2005). "The Impact of the London congestion charging scheme on the generalized cost of car commuters to the city of London from a value of travel time savings perspective" ; *Transport Policy ;* 13 ; pp. 22-23

Steg L (2005). "Car use: lust and must. Instrumental, symbolic and affective motives for car use" ; *Transportation Research part. A* ; 39 ; pp. 147-162

Tabuchi, T (1993). "Bottleneck Congestion and Modal Split" ; *Journal of Urban Economics* ; 34 ; pp. 414-431

Van Vugt M, Van Lange P and Meertens R (1996). "Commuting by car or public transportation ? A social dilemna analysis of travel mode judgements" ; *European Journal of Social Psychology* ; 26 ; pp. 373-395

Verhoef E and K Small (2004). "Product Differentiation on Road. Constrained Congestion Pricing with Heterogeneous Users" ; *Journal of Transport Economics and Policy* ; 38 ; 1 ; pp. 127-156

Vickrey W (1963). "Pricing in Urban and Suburban Transport" ; *The American Economic Review* ; 53 ; 2 ; pp. 452-465

Wardman, M (2001). "A review of British evidence on time and service quality valuations" ; *Transportation Research Part E: Logistics and Transportation Review*, Volume 37, Issues 2-3, April-July 2001, pp 107-128

Young P (1993). "The Evolution of Conventions" *Econometrica*, 61, 1, 57-84

Young P (1996). "The Economics of Convention"; *The Journal of Economic Perspectives*, 10, 2, 105-122

1 Proofs

1.1 Proof of Proposition 1

Proof. We know from Lemma (2) that an equilibrium is a solution to

$$G(z_k) = f_c - W(z_k) + \Delta(\alpha, z_k).$$

First, we show that the existence of a Nash equilibrium, z_k, satisfying

$$\left.\frac{\partial G(z)}{\partial z}\right|_{z=z_k} > \left.\frac{\partial[f_c - W(z) + \Delta(\alpha, z)]}{\partial z}\right|_{z=z_k} \tag{13}$$

is a sufficient condition for the existence of multiple equilibria. Second, we show that this is a necessary condition.

(i) If there exist such a z_k, then for any $\eta > 0$ arbitrarily small, we have

$$\begin{aligned} G(z_k) &= f_c - W(z_k) + \Delta(\alpha, z_k) \\ G(z_k + \eta) &> f_c - W(z_k + \eta) + \Delta(\alpha, z_k + \eta) \\ G(z_k - \eta) &< f_c - W(z_k - \eta) + \Delta(\alpha, z_k - \eta). \end{aligned}$$

Since the support of F is $(-\infty, \infty)$ and therefore $G(1) < f_c - W(1) + \Delta(\alpha, 1)$ and $G(0) > f_c - W(0) + \Delta(\alpha, 0)$. It implies that the functions must cross at least three times and there exist at least three equilibria. So, condition (13) is a sufficient condition for the existence of multiple equilibria.

(ii) Assume that, at any Nash equilibrium z_k, we have

$$\left.\frac{\partial G(z)}{\partial z}\right|_{z=z_k} < \left.\frac{\partial[f_c - W(z) + \Delta(\alpha, z)]}{\partial z}\right|_{z=z_k}$$

then for any $\eta > 0$, we have

$$\begin{aligned} G(z_k) &= f_c - W(z_k) + \Delta(\alpha, z_k) \\ G(z_k + \eta) &< f_c - W(z_k + \eta) + \Delta(\alpha, z_k + \eta) \\ G(z_k - \eta) &> f_c - W(z_k - \eta) + \Delta(\alpha, z_k - \eta). \end{aligned}$$

Since the support of F is $(-\infty, \infty)$, for any $z', z'' : z' < z_k < z''$, $G(z') > f_c - W(z') + \Delta(\alpha, z')$ and $G(z'') < f_c - W(z'') + \Delta(\alpha, z'')$. This implies that the functions cross only once and condition (13) is necessary.

From (i) and (ii), condition (13) is, indeed, a necessary and sufficient condition. ■

1.2 Proof of Proposition 2

Proof. Assume there exist T equilibria $z_1 > z_2 > ... > z_T$

(1) We want to show that z_T Pareto dominates any equilibrium z_j, $j = \{1,...,T-1\}$

(2) For any pair z_j, z_1 with $z_j > z_T$, there are three categories of commuters:

(a) Commuters with ε_i such that $F(\varepsilon) < 1 - z_j$. Their best response is to use public transportation in both equilibria. Those users are better off in equilibrium z_T as

$$t^{pt}(\alpha, z_T) < t^{pt}(\alpha, z_j) \text{ and } W(z_T) < W(z_j),$$

then

$$t^{pt}(\alpha, z_T) + W(z_T) + \frac{\varepsilon_i}{2} < t^{pt}(\alpha, z_j) + W(z_j) + \frac{\varepsilon_i}{2}$$

(b) Commuters with ε_i such that $1 - z_j < F(\varepsilon) < 1 - z_T$. Their best response is using public transportation in equilibrium z_T and using a car in equilibrium z_j. Those users are better off in equilibrium z_T. Indeed, as commuters reveal their preferences by choosing their mode, then for any $F(\varepsilon_j) \in [1 - z_j, 1 - z_T]$:

$$W(z_j) + t^{pt}(\alpha, z_j) + \varepsilon_j > f_c + t^c(\alpha, z_j) \tag{14}$$

and

$$W(z_T) + t^{pt}(\alpha, z_T) + \varepsilon_j < f_c + t^c(\alpha, z_T). \tag{15}$$

As congestion increases in z,

$$f_c + t^c(\alpha, z_j) > f_c + t^c(\alpha, z_T).$$

Hence, it is straightforward that

$$f_c + t^c(\alpha, z_j) > W(z_T) + t^{pt}(\alpha, z_T) + \varepsilon_j$$

(c) Commuters with ε_i such that $1 - z_T < F(\varepsilon)$. Their best response in both equilibria is to take the car. Those users are better off in equilibrium z_T as:

$$f_c + t^c(\alpha, z_j) > f_c + t^c(\alpha, z_T).$$

■

1.3 Proof of Proposition 6

Proof. Starting from $\alpha_0 \geq 0$ and $T_0 = 0$ and given the definition and the properties of $\beta(\alpha, z)$ (note that $\beta(\alpha, z) = -\frac{t^{pt}_{\alpha}(\alpha,z)}{t^{c}_{\alpha}(\alpha,z)}$ with $\beta_{\alpha}(\alpha, z) \geq 0 \; \forall z \in [0, 1]$), it is possible to define $\gamma(\alpha)$ such that

$$t^{pt}(\alpha_1, z_1) - t^{pt}(\alpha_0, z_1) = -\gamma(\alpha_1)\left[t^c(\alpha_1, z_1) - t^c(\alpha_0, z_1)\right], \tag{16}$$

with $\gamma'(\alpha) > 0$.

(i) Consider two policies: α_1 ($\alpha_1 > \alpha_0$ is associated with $T = 0$) and T_1 (associated with α_0) yielding the same equilibrium z_1. By definition, a commuter indifferent between the two policies has an ε_j such that

$$f_c + T_1 - W(z_1) + \Delta(\alpha_0, z_1) = \varepsilon_j = f_c - W(z_1) + \Delta(\alpha_1, z_1).$$

This simplifies to

$$T_1 = \Delta(\alpha_1, z_1) - \Delta(\alpha_0, z_1),$$

which can be conveniently rewritten as

$$T_1 = \left[t^c\left(\alpha_1, z_1\right) - t^c \alpha_0, z_1\right] - \left[t^{pt}\left(\alpha_1, z_1\right) - t^{pt}\left(\alpha_0, z_1\right)\right].$$

By assumption, this leads to

$$T_1 = \left(1 + \gamma(\alpha_1)\right)\left[t^c\left(\alpha_1, z_1\right) - t^c\left(\alpha_0, z_1\right)\right]. \tag{17}$$

(ii) Car users are indifferent between these two policies if $\exists\ \alpha_1$, T_1, z_1 such that

$$\left(1 - z_1\right) T_1 = t^c\left(\alpha_1, z_1\right) - t^c\left(\alpha_0, z_1\right).$$

These conditions imply

$$\begin{aligned}
\left(1 - z_1\right)\left[1 + \gamma(\alpha_1)\right]\left[t^c\left(\alpha_1, z_1\right) - t^c\left(\alpha_0, z_1\right)\right] &= t^c\left(\alpha_1, z_1\right) - t^c\left(\alpha_0, z_1\right) \\
\left(1 - z_1\right) &= \frac{1}{1 + \gamma(\alpha_1)}.
\end{aligned}$$

Now consider two alternative policies (α_2 and T_2) associated with a higher use of public transportation (z_2) with $z_2 < z_1$. Car users are now better off with traffic separation than with taxation iff

$$\left(1 - z_2\right) T_2 > t^c\left(\alpha_2, z_2\right) - t^c\left(\alpha_0, z_2\right).$$

From the expression of T in equation (17),

$$\begin{aligned}
\left(1 - z_2\right)\left[1 + \gamma(\alpha_2)\right]\left[t^c\left(\alpha_2, z_2\right) - t^c\left(\alpha_0, z_2\right)\right] &> t^c\left(\alpha_2, z_2\right) - t^c\left(\alpha_0, z_2\right) \\
\left(1 - z_2\right) &> \frac{1}{1 + \gamma(\alpha_2)}.
\end{aligned}$$

This is always true as $z_2 < z_1$, $\gamma(\alpha_2) \geq \gamma(\alpha_1)$ and given that $\left(1 - z_1\right) = \frac{1}{1+\gamma(\alpha_1)}$ ■

2 A delegation mechanism

Consider that the policy tools are delegated to an agency. The agency is fully independent of the social planner. It is entirely credible in implementing traffic separation α^* and committing to a tax scheme such that the only equilibrium share of commuters is z^*. An optimal tax scheme always exists and is any function $\hat{T}(z)$ meeting the following conditions, $\forall z' > z^*$ and $z'' < z^*$:

$$
\begin{aligned}
G(z') &< f_c + \hat{T}(z') - W(z') + \Delta(\alpha^*, z'), \\
G(z'') &> f_c + \hat{T}(z'') - W(z'') + \Delta(\alpha^*, z''), \\
G(z^*) &= f_c + \hat{T}(z^*) - W(z^*) + \Delta(\alpha^*, z^*). \qquad (18)
\end{aligned}
$$