Congestion and departure time choice equilibrium in urban road networks

THÈSE N^{O} 8862 (2018)

PRÉSENTÉE LE 23 NOVEMBRE 2018

À LA FACULTÉ DE L'ENVIRONNEMENT NATUREL, ARCHITECTURAL ET CONSTRUIT
LABORATOIRE DE SYSTÈMES DE TRANSPORTS URBAINS
PROGRAMME DOCTORAL EN GÉNIE CIVIL ET ENVIRONNEMENT

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Raphael Ali Francis LAMOTTE

acceptée sur proposition du jury:

Prof. A. M. Alahi, président du jury
Prof. N. Geroliminis, directeur de thèse
Prof. M. Fosgerau, rapporteur
Prof. L. Leclercq, rapporteur
Prof. K. W. Axhausen, rapporteur

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Suisse
2018

Acknowledgements

This PhD has been a very rich and enjoyable period of my life. I was given the incredible chance to spend four years to think about interesting problems and try to solve them. This chance, I largely owe it to my advisor, Nikolas Geroliminis. Not only he secured funds so that I could focus on the scientific challenges, but he also guided me towards promising questions, listened to whatever I had to say without ever getting tired or offended and most importantly, he trusted me. Thank you Nikolas.

Besides this great freedom, I also had the chance to exchange with many extraordinary persons that helped me become a little bit closer to a true Scientist. I had the unique chance to spend many Thursdays with André de Palma at EPFL and six weeks with Richard Arnott in Riverside. These interactions have been very precious for me, and with hindsight, I think their knowledge, enthusiasm, rigor and viewpoints really had a great and positive influence on me. So thanks to both of you!

I am also very grateful to the members of my thesis committee (Mogens Fosgerau, Ludovic Leclercq, Kay Axhausen and Alexandre Alahi), and to those of my candidacy exam (Matthieu de Lapparent, Michel Bierlaire). Matthieu and Michel actually provided feedback to me throughout my PhD via our internal seminars. Thank you!

Four years however is a rather long period, and it is not always easy to remain enthusiastic. Fortunately I was well surrounded both at EPFL and outside. Thanks to all the LUTS and TRANSP-OR friends, and in particular to Mor, Tasos, Reza, Martin, Mikhail, Semin, Leo, Işık, Claudia, Dimitris, Mohsen, Mehmet, Nan, Gael, Nicholas and Virginie. Special thanks go to Mikhail, for all the attention he gave to my mathematical problems! I also had the chance to work with many nice and smart students. I especially appreciated all the informal discussions I had with Romain, David and Martí. Having such students really gives hope for the future!

I have also been well supported outside EPFL. I cannot thank everybody here but thanks to all the LUCiens, the Kléberiens, thanks to the wonderful O'Kane family, to my own family and to Julie, who followed me across the Atlantic and back, and still support(e/s) me everyday!

Finally, thanks to the researchers who have stayed true to the academic dream and that do what they can to make the world a better place. Dreams are the best fuel for research!

Lausanne, October 16, 2018 R. L.

Abstract

When drivers are regularly faced with congestion, they try to optimize their departure time. If the demand and the road network evolve slowly enough, the entire system may approach an equilibrium, i.e. a state such that no one can be better off by unilaterally changing departure time. The transportation community has devoted a significant effort to identify such equilibria. The ultimate goal is to be able to predict the consequences of large infrastructure projects and to design smart policies to alleviate congestion. Yet, several issues still limit the applicability of the existing literature. This thesis identifies three complementary challenges and attempts to address them.

We first investigate whether, from a theoretical viewpoint, real world unidirectional flows are likely to be in a near-equilibrium state. Our analytical findings reveal the influence of schedule preferences on stability, and explain why morning commutes are more likely to be unstable than evening ones. Fortunately, user heterogeneity or socially optimal pricing can soften the effects of instability. Residual oscillations result in a congestion cost decomposition that differs from the one observed at equilibrium, but the overall average congestion cost at equilibrium is remarkably accurate.

We then characterize departure time choice equilibria in isotropic regions, representing multi-directional road networks. In this context, users can slow down others who started their trips earlier and trip length is an important determinant of departure time choice. We show that with a widely used type of schedule preferences, users with long trips tend to avoid the peak period. Although the First-In, First-Out (FIFO) property does not hold in general with heterogeneous trip lengths, such a pattern emerges among families of early and late users having the same preferences. Simulations suggest that the social cost and its decomposition greatly differ from those observed in unidirectional settings.

We finally propose an alternative way of reducing congestion, based on an optional booking service with dedicated right-of-way. Such a service would be advantageously implemented with shared and/or autonomous vehicles, due to similar requirements and complementarity. We recognize that participation to such a program would entail an alternative-specific inconvenience and evaluate the consequences on welfare depending on the way it is administrated and on the capacity split.

Overall, this thesis advances the state of knowledge regarding the prevailing traffic conditions and suggests ways to improve them. The adopted approach is largely analytical to provide insight and generality. Complementary simulations add realism and push back the limits of tractability.

Keywords: Departure time, Scheduling, Congestion, Equilibrium, Stability, Macroscopic Fundamental Diagram, Bottleneck, Demand management, Directional Flows.

Résumé

Les automobilistes régulièrement confrontés à la congestion essaient d'optimiser l'heure de leur départ. Si la demande et le réseau routier évoluent lentement, le système peut s'approcher d'un équilibre, i.e. un état dans lequel personne ne peut améliorer sa situation en changeant de décision unilatéralement. De nombreuses recherches ont été effectuées pour identifier ces équilibres, dans le but de prédire les conséquences des grands projets d'infrastructures et de trouver des solutions pour décongestionner les routes. Cependant, leur concrétisation se heurte encore à plusieurs obstacles. Cette thèse en identifie trois relativement peu étudiés et apporte quelques solutions.

Nous commençons par examiner si, théoriquement, les décisions d'usagers se déplaçant dans une même direction sont susceptibles d'approcher un équilibre. Nos résultats révèlent l'importance des préférences horaires des usagers et suggèrent que la période de pointe du soir (après le travail) est plus stable que celle du matin. Heureusement, l'hétérogénéité entre les usagers ou la mise en place d'une stratégie de tarification peuvent atténuer les effets de l'instabilité. Les oscillations résiduelles induisent une décomposition du coût de congestion différente de celle de l'équilibre, mais en moyenne, le coût total de la congestion est remarquablement proche de celui à l'équilibre.

Nous étudions ensuite les conditions d'équilibre dans des régions homogènes et isotropes, représentant des réseaux routiers multidirectionnels. Dans ce contexte, les automobilistes peuvent en ralentir d'autres qui ont commencé leur trajet plus tôt, et la longueur de trajet est un déterminant essentiel du choix. Nous montrons qu'avec les préférences horaires les plus communément utilisées dans la littérature, les automobilistes ayant les trajets les plus longs évitent de circuler pendant les périodes les plus congestionnées. Bien que l'ordre des arrivées n'est pas nécessairement le même que celui des départs quand on considère des trajets de différentes longueurs, les trajets sont naturellement organisés de cette manière parmi des groupes d'usagers arrivant en avance ou en retard. Nos simulations suggèrent que le coût de la congestion et sa décomposition diffèrent grandement de ceux observés dans des configurations unidirectionnelles.

Nous proposons finalement une manière de réduire la congestion basée sur la participation facultative à un service de réservation de créneaux horaires, bénéficiant de voies réservées. Du fait de leurs besoins similaires et de certaines synergies, ce service est plus facilement envisageable avec des véhicules autonomes et/ou partagés. Notre étude prend en compte un coût spécifique lié à la réservation et évalue les conséquences d'un tel système pour la société, dépendamment du régime économique et de la proportion de la capacité qui lui

est allouée.
Dans son ensemble, cette thèse améliore notre compréhension des mécanismes à l'origine de la congestion et indique des solutions possibles. L'approche théorique adoptée confère à ce travail une certaine généralité et une valeur explicative. La simulation est utilisée en complément, pour traiter des situations plus réalistes et illustrer nos résultats analytiques.

Mots clés : Heure de départ, Planification, Congestion, Équilibre, Stabilité, Diagramme fondamental de zone, Goulet d'étranglement, Destion de la demande, Flux directionels.

Contents

1 Introduction

1.1 Context and motivation

Road congestion represents a major challenge for modern societies. It arises mainly in cities, where demand for travel is high, and space is limited. It has a wide range of negative consequences on our health, environment and economy. Some of these might be curbed in the future thanks to technological progress (e.g. electric vehicles and better pavement design may reduce noise and air pollution), while others are intrinsic characteristics of congestion. When too many persons want to travel at the same time and do not agree on a joint strategy, the outcome is unavoidably very costly for society. Indeed, the hours stressfully spent behind the wheel could be enjoyed or used productively if it were not for congestion. Besides, congestion also prevents many opportunities from being seized: when some spend hours in traffic jams, others give up traveling, or travel at inconvenient times.

To alleviate the congestion cost, we should understand both the demand (when users depart, how they choose their transportation mode and their route), the supply (how the demand affects travel times), and their interactions. This thesis focuses only on one of these aspects, which is the interaction between the supply and the choice of departure time. From a psychological point of view, one may wonder whether this choice is made consciously, how often it is made, what information is used, etc. This thesis considers such behavioral aspects as inputs. It assumes that commuters are rational and that they try to choose the alternatives that are the best for themselves.

Then, the challenge is to determine how congestion emerges from the interaction between a great number of individuals. To address it, the traffic literature has relied on various concepts from game theory. Nash equilibria for instance are very widely used in the traffic literature, where they are known as user equilibria. With rational users that only switch to alternatives with higher utilities, Nash equilibria represent fixed points. The traffic literature often assumes that the systems considered are at equilibrium and focuses on characterizing equilibria. This convenient “equilibrium assumption” has allowed for many elegant derivations and extremely valuable insight.

The line of research on the departure time equilibrium stems from Vickrey (1969, 1973). In these seminal works, the Nobel Laureate William Vickrey made the case for dynamic congestion pricing by considering a single unidirectional bottleneck of constant capacity. He derived the user equilibrium without toll and showed how the social optimum can be decentralized using a time-varying toll. This idea gained large support among the academic community and Vickrey’s bottleneck model has become “the workhorse for the economic analysis of rush-hour traffic dynamics” (Arnott et al., 2016). Yet, almost 50 years after the first seminal paper, dynamic congestion tolls are still rare. The difficult transfer from theory to practice is generally attributed to the poor popularity of pricing, but other reasons may also explain these difficulties.

First, although the literature has flourished, it has left largely unexamined its most fundamental assumptions. The equilibrium assumption in particular seems to be supported only by the gut feeling that “people do not change their departure time too often, or only because of changes in their constraints (e.g. due to an early meeting)”. While intuition is often helpful, note that people may also adjust their schedule unconsciously. For instance, they may account for the traffic conditions experienced over the last days when scheduling their meetings. Besides, multiple authors have actually reported set-ups involving unidirectional congestion where the system *does not converge* to an equilibrium (de Palma, 2000; Mc Breen et al., 2006; Iryo, 2008; Guo et al., 2018). These examples do not provide a sufficiently strong basis for generalization (they all use the same highly schematic specifications of schedule preferences), but they do show that equilibrium stability cannot be taken for granted.

Second, the congestion phenomena described in the literature are often very different from those occurring in urban areas. Congestion in urban areas typically results from the interactions between a large variety of directional flows, while the literature is typically restricted to unidirectional mechanisms. There is however an emerging branch of the literature that considers congestion in homogeneous and isotropic areas (Small and Chu, 2003; Geroliminis and Levinson, 2009; Fosgerau and Small, 2013; Arnott, 2013; Fosgerau, 2015; Daganzo and Lehe, 2015; Arnott et al., 2016). Isotropic traffic models have advantages in terms of generality and tractability, but they also have very realistic features. If the common unidirectional bottleneck represents the access to a monocentric city, the isotropic area can be considered to represent a large homogeneous urban area, like those produced by conurbation. The isotropic model also provides a natural framework to account for hypercongestion, which is a severe form of congestion where the flow decreases with the number of cars on the road. Importantly, isotropic congestion also allows for studying the influence of trip length on departure time choice. The first works that considered an isotropic model with departure time choice only focused on hypercongestion (Small and Chu, 2003; Geroliminis and Levinson, 2009; Fosgerau and Small, 2013; Arnott, 2013). It is only very recently that Fosgerau (2015) and Daganzo and Lehe (2015) recognized how trip length heterogeneity challenges the fundamental First-In-First-Out (FIFO) assumption and started investigating its impacts on the morning commute. Yet, these two papers reached very different conclusions concerning the role of trip length, especially in terms of sorting patterns. Since sorting patterns are likely to influence stability, their study is a research priority.

Third, the literature does not offer many alternatives to congestion pricing. Some tradable credit schemes have been proposed (Wada and Akamatsu, 2013) but they generally failed to account for the inconvenience related to the trading activity they imply. This is problematic as such an inconvenience might be very large for some persons/trips. Optional participation to the tradable credit scheme may circumvent the issue. Yet, this choice would be somehow artificial if all users then travel together: those choosing the tradable credit scheme would incur an alternative-specific inconvenience, would depart at a time that might not be optimal for themselves and they would incur the same congestion as non-cooperating users. Thus, the cooperative alternative would clearly be dominated for everyone. In order to offer a real choice, it is necessary to separate the flows. Since new roads cannot easily be constructed, this suggests splitting the existing capacity. Such a concept raises questions related to the optimal capacity split and to how such a service should be operated.

1.2 Objectives

This thesis aims at bringing the theory on departure time choice closer to the practice by addressing the three main issues mentioned in Section 1.1. More specifically, our objectives are:

1. Equilibrium assumption with unidirectional congestion
   (a) To determine whether reported instability results can be generalized.
   (b) In case of instability: to estimate to which extent the equilibrium assumption can still be applied.
2. Equilibrium characterization in an isotropic and homogeneous area:
   (a) To identify how trip length and other factors determine the prevailing sorting patterns.
   (b) To compare the unidirectional and the isotropic equilibria.
3. Voluntary cooperation:
   (a) To determine whether dedicated lanes can lead to voluntary cooperation.

(b) To identify an environment in which such a scheme could realistically be implemented and enforced, while remaining attractive.

(c) To provide directions for the best mode of operation (public, private, regulations).

1.3 Contributions

The main scientific contributions of this thesis are summarized hereafter.

1. *Equilibrium assumption with unidirectional congestion*

   Chapter 3 decomposes the game at hand into basic elements and identifies properties of these elements that are related to stability. We demonstrate in particular that with a homogeneous population, the unidirectional bottleneck exhibits features that can be associated to instability when the marginal utility rate at the destination increases (like in the morning commute) and to stability when the marginal utility rate at the destination is non-increasing. We also show that the location of a toll booth (upstream or downstream of the bottleneck) may influence these features.

   We then rely on simulations to study more complex cases, with heterogeneity and pricing in discrete time frameworks. We quantify the error associated with the equilibrium approximation and identify some important biases it entails.

2. *Equilibrium characterization in an isotropic and homogeneous area*

   Chapter 4 only focuses on intra-day dynamics and investigates the equilibrium properties of the morning commute problem at the network level with heterogeneous trip lengths. Congestion is modeled with an isotropic model relating the space-mean speed of a network to the vehicular accumulation. Stability issues are not considered analytically (a good knowledge of static properties is a prerequisite) but simulation-based studies suggest that the equilibrium approximation is reasonable there, at least for light and medium congestion levels. Our main contributions include a proof of continuity of accumulation over time at equilibrium, a proof that a partial First-In, First-Out (FIFO) sorting pattern emerges at equilibrium with widely-accepted schedule preferences and a comparison of the morning commutes with isotropic/unidirectional congestion mechanisms.

3. *Voluntary cooperation*

   Chapter 5 considers the problem of the optimal capacity split at a unidirectional bottleneck in a context where users can choose among unregulated competition and cooperation. We depict the cooperative alternative as an original mobility service, which would be advantageously implemented together with car-sharing and autonomous vehicles, for reasons detailed in the thesis. Non-cooperative users choose their departure time from home and compete for the best departure times from the bottleneck. Cooperative users need to book their trip in advance. As the number of time slots available for booking does not exceed the capacity, cooperative users are guaranteed no delay at the bottleneck. An individual-specific cooperation cost is introduced in the modeling framework to account for the related inconvenience. We then investigate how a central planner should allocate the capacity to these two types of users depending on the regime (laissez-faire, welfare- or profit-maximizing). Two major findings are that the equilibrium demand split Pareto-dominates the case with only competitive users and that the social cost difference between equilibrium and socially optimal demand splits is small compared to their benefits. Profit-maximizing strategies however turn out to be hardly compatible with welfare maximization.

1.4 Thesis structure

Chapter 2 introduces the departure time choice problem and provides the background that is common to all the subsequent chapters. The problem is first presented from a broad perspective including day-to-day dynamics, before focusing on the intra-day dynamics and reviewing the associated literature. Chapter 2 further explains how schedule preferences are usually modeled and introduces the two types of congestion mechanisms that are used in this thesis (uni-directional and isotropic). It then summarizes various equilibrium construction methods proposed in the literature and illustrates the simplest of them with a classic example.

Chapter 3 provides further details on day-to-day dynamics and introduces fundamental concepts from evolutionary game theory that are related to stability. We then build on this framework to derive analytical results valid under ideal conditions (continuous time and homogeneous users) and complement them with simulations of more realistic set-ups. Preliminary ideas are published as:

- Lamotte, R., Geroliminis, N., 2018. (In)stability of departure time choice with the bottleneck model. Presented at the 18th Swiss Transport Research Conference (STRC), Ascona, Switzerland.

Chapter 4 reviews more extensively the literature considering departure time choice with isotropic congestion and analytically derives important features (e.g. continuity, sorting) of the equilibrium. Complementary simulation results illustrate these properties and serve as a basis for comparison with the classic unidirectional problem. The preliminary ideas of this chapter are published as

- Lamotte, R., Geroliminis, N., 2016. The morning commute in urban areas: Insights from theory and simulation. Presented at the 95th Annual Meeting of the Transportation Research Board, Washington, DC.

- Lamotte, R., Geroliminis, N., 2017. The morning commute in homogeneous cities. Presented at the 17th Swiss Transport Research Conference (STRC), Ascona, Switzerland.

The content of this chapter has then been published as

- Lamotte, R. and Geroliminis, N., 2017. The morning commute in urban areas with heterogeneous trip lengths. Transportation Research Part B: Methodological (In Press). *Presented at the 22nd International Symposium on Transportation and Traffic theory (ISTTT), Chicago, Illinois.*

Chapter 5 introduces an original way of fostering cooperation by allocating a proportion of the capacity to cooperative users. The introductory sections discuss the practical advantages of booking, its complementarity with autonomous and/or shared vehicles and the inconvenience it creates. The equilibrium is then characterized mathematically under various regimes (laissez-faire, welfare or profit-maximizing). It is based on the article

- Lamotte, R., de Palma, A. and Geroliminis, N., 2017. On the use of reservation-based autonomous vehicles for demand management. Transportation Research Part B: Methodological, 99, pp.205-227.

Chapter 6 summarizes the contributions of this thesis, highlights its practical implications and identifies future research directions. The emphasis is given mainly to two specific research directions, which are to investigate equilibrium stability with isotropic congestion and to consider time-dependent capacity allocation.

The code I wrote to simulate departure time choice with bottlenecks and within isotropic areas is available online at https://github.com/raplam/departureTimeChoice.

2 Background

2.1 Elements of the problem

This thesis considers that agent behavior is defined by two functions. One function captures the preferences of users and allows comparing the utility (or cost) of all alternatives. The second function defines how users change their decisions from day to day, depending on the current choices and the utilities of all alternatives. Together with the congestion mechanism, these two functions fully define the dynamics of the system. Indeed, given some initial conditions, we can apply iteratively these three functions and imitate the evolution of the traffic conditions from day to day, as illustrated in Fig. 2.1.

Yet, to facilitate the analysis of such systems, most of the departure time choice literature has relied on the equilibrium assumption. Under this assumption, there is no need to specify the protocol that users follow to update their decision. Indeed, the user equilibrium is characterized by the fact that users cannot improve their utility by changing decision unilaterally. If users are rational and non-cooperative, they have no reason to change their decisions once they are at equilibrium. This assumption greatly simplifies the analysis as it suffices to determine the equilibrium to evaluate all sorts of indicators, without accounting for the initial conditions or for convergence issues.

This background section provides an introduction to schedule preferences and to congestion mechanisms, as well as a review of the literature on departure time choice based on the equilibrium assumption. Adjustment mechanisms and systems out of equilibrium are only considered in Chapter 3.

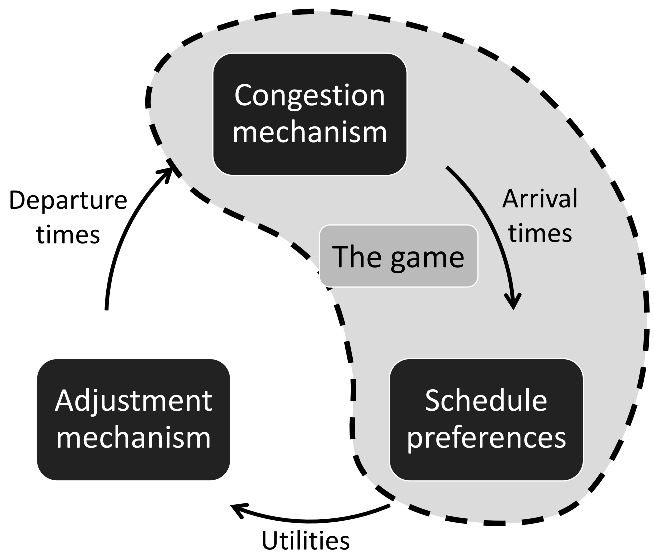

Figure 2.1 – A cyclic description of departure time choice

2.2 Schedule preferences

This thesis follows Vickrey (1973) and assumes that the schedule preferences are expressed by a utility function of the form

$$U(t_o, t_d) = \underbrace{\int_0^{t_o} u_o(t)\,\mathrm{d}t}_{U_o(t_o)} + \underbrace{\int_{t_d}^{0} u_d(t)\,\mathrm{d}t}_{U_d(t_d)}, \tag{2.1}$$

where t_o and t_d are the times of departure (from the origin) and arrival (at destination).[1] In plain words, it means that individuals have values of time that depend on the time of the day and on their location (at home and at work for the commute problem). Note that the above expression implicitly assumes that users do not value the time they spend traveling. While this might sound problematic, it is actually equivalent to defining the marginal utility rates at the origin and at the destination relatively to the marginal utility rate of travel time. The values of the lower limit of the integral U_o and of the upper limit of U_d (both equal to 0 in Eq. (2.1)) do not matter since utility is only defined up to a constant.

This thesis further assumes that the marginal utility rates at the origin (u_o) and at the destination (u_d) are positive everywhere, piece-wise continuous, and that there exists t^* such that for all $t < t^*$, $u_o(t) > u_d(t)$ and for all $t > t^*$, $u_o(t) < u_d(t)$. Positivity ensures

[1]The choice of the specification depends on the problem. For instance when considering multiple consecutive trips, one should account for the fact that the value of time at a location may depend on how much time has already been spent at this location. Such dependencies are traditionally ignored when considering commutes.

that commuters do not simply drive in circles, while the last condition ensures that users have an incentive to move from the origin to the destination. The reader can think of t^* as the "ideal teleportation time" in the sense that if users could simply teleport from their origin to their destination (such that $t_d = t_o$), they would always prefer to do so exactly at t^*.

Alternatively, the literature often specifies the schedule preferences in terms of a generalized cost function, which is the negative of a utility function. More specifically, the generalized cost function typically takes the form $C(t_o, t_d) = \alpha(t_d - t_o) + \mathrm{SP}(t_d)$ for the morning commute and $C(t_o, t_d) = \alpha(t_d - t_o) + \mathrm{SP}(t_o)$ for the evening one. The first term is known as the cost of delay, the second as the schedule penalty (SP). The schedule penalty is typically assumed to be a non-negative unimodal function reaching 0 at t^*. Thus, both terms are always non-negative and since the minimal possible cost is clearly 0, the generalized cost naturally represents the cost of congestion.

The formulation in terms of generalized cost actually corresponds to a special case of Eq. (2.1) where the marginal utility rate at home is equal to α at all times. To see this, replace the limits of integration that are equal to 0 in Eq. (2.1) by t^* (we can do this because utility is only defined up to a constant). Then, if in a morning commute case u_o is constant, Eq. (2.1) boils down to

$$U(t_o, t_d) = -\underbrace{\int_{t_o}^{t_d} u_o(t)\,\mathrm{d}t}_{=\alpha(t_d - t_o)} - \underbrace{\int_{t^*}^{t_d} u_d(t) - u_o(t)\,\mathrm{d}t}_{\triangleq \mathrm{SP}(t_d)}.$$

Similarly, if in an evening commute case u_d is constant,

$$U(t_o, t_d) = -\underbrace{\int_{t_o}^{t_d} u_d(t)\,\mathrm{d}t}_{=\alpha(t_d - t_o)} - \underbrace{\int_{t_o}^{t^*} u_o(t) - u_d(t)\,\mathrm{d}t}_{\triangleq \mathrm{SP}(t_o)}.$$

If the marginal utility of time spent at work is piece-wise constant during the morning commute ($u_d(t) = \alpha - \beta$ if $t < t^*$ and $u_d(t) = \alpha + \gamma$ otherwise), the formulation with marginal utilities of time boils down to the $\alpha - \beta - \gamma$ preferences.[2] The schedule penalty is then equal (up to a constant) to:

$$\mathrm{SP}_{\alpha-\beta-\gamma}(t_d, t^*, \beta, \gamma) = \max(\gamma(t_d - t^*), \beta(t^* - t_d)). \tag{2.2}$$

Although they are widely used in the literature, the use of the $\alpha - \beta - \gamma$ preferences has often been questioned. In fact, the empirical evaluations of marginal utilities systematically relied on more complex models. Small (1982) allowed for a discrete jump

[2] These notations seemingly appeared with Arnott et al. (1990) and became standard.

in utility for late arrivals and for a flexible range of arrival times with lower penalties. Hendrickson and Plank (1984) estimated two quadratic penalty functions for early and late arrivals, which are both equivalent to affine marginal utility functions. Tseng and Verhoef (2008) estimated marginal utilities at different times of the day in a non-parametric way, and Hjorth et al. (2015) used exponential marginal utility functions. However, when comparing different models with stated preference data, Hjorth et al. (2015) found that piecewise constant and affine specifications outperformed the exponential ones. While all agree that constant marginal utilities are not ideal, there is no consensus on the parametric form that should be used. For instance, Hendrickson and Plank (1984) found a marginal utility of time at work that decreases after t^* while the results of the other works mentioned above suggest an increase.

2.3 Congestion models

2.3.1 The fluid approximation

A typical commute involves many thousands of agents. When dealing with such large numbers, it is common to rely on the so-called "fluid approximation". Within this approximation, the number of users is considered as a real-valued variable (we sometimes call it the "mass", or "accumulation" of users). This approximation greatly simplifies the study of the dynamics as accumulation can then potentially evolve continuously, and not necessarily with very small steps of one user. When a strictly positive mass of users departs or arrives simultaneously, the cumulative number of departures or arrivals exhibits a discontinuity. We refer to such phenomena as "mass departures" and "mass arrivals". The two congestion models described hereafter rely on the fluid approximation.

2.3.2 Unidirectional congestion: the constant capacity bottleneck

The constant capacity bottleneck model (a.k.a. Vickrey's bottleneck model) is the archetype of unidirectional congestion mechanisms. It considers a single origin-destination pair, connected by a single route with a bottleneck of constant capacity s, in veh/h. Travel time is the sum of a fixed component T_f to go from the origin to the bottleneck, a waiting time at the bottleneck and a fixed component T'_f to go from the bottleneck to the destination. Both fixed components can be set to 0 without loss of generality (it simply means replacing the functions $t_o \to u_o(t_o)$ and $t_d \to u_d(t_d)$ by $t_o \to u_o(t_o - T_f)$ and $t_d \to u_d(t_d + T'_f)$). Thus, t_o is also the arrival time at the bottleneck and t_d is also the departure time from the bottleneck.

Let $r(t)$ denote the arrival rate at the bottleneck and $q(t)$ the queue length (in vehicles) at time t. The queue length evolution obeys

$$\dot{q}(t) = \begin{cases} r(t) - s, & \text{if } q(t) > 0 \text{ or } r(t) > s \\ 0, & \text{otherwise.} \end{cases} \tag{2.3}$$

Thus, if we know the departure rate $r(t)$, the arrival time is given by $t_d = t_o + q(t_o)/s$.

Note that many alternatives allow modeling unidirectional congestion with a finer spatial resolution (see van Wageningen-Kessels et al. (2015) for a genealogy). For instance, car-following models provide a very intuitive behavioral description of congestion, in which vehicles adjusts their speed based on the vehicles ahead. Such models are useful to describe the spatial distribution of vehicles but they would not allow for an analytical treatment of trip scheduling problems. Fortunately, these models tend to produce very similar arrival times at destination, and these are typically quite close to those predicted with Vickrey's bottleneck model, especially when a strong demand peak exceeds the bottleneck capacity (Verhoef, 2001). The main limitation of Vickrey's bottleneck model is that it cannot model travel time variations when demand is constantly below the bottleneck capacity. The trip scheduling literature leaves this issue aside by focusing on scenarios where demand exceeds capacity. The transitions at the beginning and end of the congested period are admittedly not realistic, but this inaccuracy is negligible if the transition periods are short compared to the peak period.

2.3.3 Isotropic congestion: the MFD (or bathtub)

In the real world, different directional flows interact in many ways: they can compete for the same road, for the same intersection, or interact indirectly through a chain of interactions with other flows. The range of possible configurations is too large to consider an analytical treatment of each case. Instead, we consider another idealization that leaves aside directions altogether and describes the dynamics of vehicles in a homogeneous and isotropic environment: for any time t, the average speed of all the $n(t)$ vehicles traveling is assumed to be given by a function $\mathcal{V}(n(t))$.

The same relation can also be described in terms of total distance traveled per time unit, by all the vehicles currently in the network. This quantity, known as the production, is given by $P(t) = \mathcal{P}(n(t)) = n(t)\mathcal{V}(n(t))$. The production function $\mathcal{P}$ is assumed to be positive and unimodal on an interval $[0, n_{\text{jam}}]$, while the corresponding speed function $\mathcal{V}$ is positive, continuously differentiable, decreasing on $[0, n_{\text{jam}}]$. Both speed and production reach 0 for $n = n_{\text{jam}}$. The accumulation that maximizes the production is assumed to be unique and is referred to as the "critical" accumulation. The free-flow speed $\mathcal{V}(0)$ is denoted v_{f}.

The idea of a such relations with an optimum accumulation was initiated by Godfrey (1969) and similar approaches were introduced later by Herman and Prigogine (1979), Mahmassani et al. (1984) and Daganzo (2007). Geroliminis and Daganzo (2008) was among the first works to empirically observe such relationships at the level of a downtown neighborhood. These relations have been exploited within large-scale traffic control strategies, where they are usually referred to as (Network) Macroscopic Fundamental Diagrams, (N)MFD (see e.g. Kouvelas et al. (2017)). Transportation economists refer to the same concept as "the bathtub model of traffic congestion", a name reportedly coined by W. Vickrey (Arnott, 2013). In this thesis, $\mathcal{P}$ is called the production-MFD and $\mathcal{V}$ the speed-MFD.

In order to have a complete dynamic model, we need to define when users complete their trips. Daganzo (2007) proposed an aggregated approach based on the assumption of slowly-varying conditions. It assumes that at any time t, the trip completion rate (also called outflow) is given by $O(t) = \mathcal{O}(n(t)) = \frac{\mathcal{P}(n(t))}{L}$, where L is the average length of the generated trips. The function $\mathcal{O}$ is referred to as the outflow-MFD. Geroliminis and Daganzo (2008) also provided empirical observations of such a relationship, using data from static loop detectors and taxi traces. While it is possible to use the outflow-MFD as the congestion mechanism of a departure time choice problem (Small and Chu, 2003; Geroliminis and Levinson, 2009), these dynamics cannot easily accommodate for trip length heterogeneity, especially when the distribution varies over time. This is problematic as trip length is intuitively an important determinant of departure time choice. Besides, the conditions are not necessarily slowly-varying during the congestion peaks.

The "trip-based" model offers an attractive alternative to the outflow-MFD. It is based on the principle that a trip of length l that starts at time t_o should finish at the unique[3] time t_d such that

$$\int_{t_o}^{t_d} \mathcal{V}(n(t))\,\mathrm{d}t = l. \tag{2.4}$$

Note that this is automatically the case if all vehicles travel exactly at $\mathcal{V}(n(t))$ (instead of simply traveling *on average* at $\mathcal{V}(n(t))$). This principle can then be used to derive the dynamics of accumulation. Consider the flow entering the zone $I(t)$ as given for all times t and assume that $I(t) = 0$ for all times $t < 0$. If the exit times follow Eq. (2.4), then the accumulation at time t is given by

$$n(t) = \int_0^t I(s) \left(1 - F_{l,s}\left(\int_s^t \mathcal{V}(n(u))\,\mathrm{d}u\right)\right) \mathrm{d}s, \tag{2.5}$$

where $F_{l,s}(\cdot)$ is the cumulative distribution function of the trip length for users entering the network at time s (the associated probability density function is denoted $f_{l,s}(\cdot)$).

[3] Uniqueness is guaranteed if we assume that n always remains bounded away from n_{jam}.

Differentiating this equation leads to:

$$\dot{n}(t) = I(t) - \mathcal{V}(n(t)) \int_0^t I(s) f_{l,s} \left(\int_s^t \mathcal{V}(n(u)) \, \mathrm{d}u \right) \mathrm{d}s. \tag{2.6}$$

Note that the outflow rate depends not only on the current value of accumulation but also on all the previous values. This is in contrast with the outflow-MFD approach. Note also that it also means that vehicles can slow down those that started their trips earlier. In a directional setting with identical origins, this is a violation of the fundamental causality principle. In an isotropic and homogeneous environment however, spatial location and trip advancement are independent.

2.4 Characterizing equilibria

2.4.1 Methods

As explained earlier, a Nash equilibrium is a situation such that no user can become better off by changing decision unilaterally. Although this definition is very simple, finding equilibria can be quite complicated in games with large number of users and alternatives. Departure time equilibria are especially complicated because congestion propagates over time, and therefore affects users choosing different alternatives. This is in contrast with static route choice applications, where the cost of a road often depends only the flow on that same road.[4] Fortunately, the task can be greatly simplified in some specific settings.

The simplest case is perhaps with a bottleneck and homogeneous users. In this case, the equilibrium is such that all used alternatives have the same utility and such that no unused alternative has a strictly greater utility than the used ones. This is the situation considered in Section 2.4.2.

Another setting that is particularly convenient is when the classical first and second order conditions for local optimality are sufficient to identify a unique candidate equilibrium state. Several authors have found situations where this is the case, for instance Vickrey (1969); Fosgerau and de Palma (2012); Fosgerau (2015). In Fosgerau and de Palma (2012) and Fosgerau (2015), the authors consider populations that differs in only one parameter and show that by integrating the derivative of the first-order-condition with respect to that parameter, they can derive the equilibrium. This method is however limited to cases where the decision (i.e. the departure or arrival time) is a continuously differentiable function of the parameter considered at equilibrium.

Then, there also exist two important numerical approaches. Iryo and Yoshii (2007) introduced a method that boils down to solving a simple linear program. This method

[4] Although this is quite reductive, these games are known as “congestion games” in game theory.

only applies when users can only choose among a discrete set of arrival times (each arrival time having a capacity), and all have the same constant marginal utility rate at their origin u_o. This is because in such cases, the socially optimal arrival times coincide with the equilibrium arrival times.

Many heuristic approaches are also possible. For instance, one may try to solve the equilibrium by defining some disequilibrium index and by minimizing it. However the problem at hand is rarely convex, such that generic optimization techniques often fail to find the equilibrium. Last, many authors have proposed iterative methods that imitate to some extent the real world process. The most common approach relies on the Method of Successive Averages (MSA), in which the state at each iteration is given by a weighted average between the previous state and the state that would be obtained if all users updated their decisions using the current conditions. It is used for instance within the METROPOLIS software (de Palma and Marchal, 2002).

2.4.2 The classical example

While departure time equilibria can be quite complex, one example is particularly elegant and insightful: that is the case of a bottleneck with constant capacity and homogeneous users having $\alpha - \beta - \gamma$ preferences.

Let N and S denote the total demand and the bottleneck capacity. In order to serve all users, the bottleneck should be used at full capacity during a total duration N/S. Let c denote the equilibrium cost. Clearly, to satisfy the equilibrium condition, all arrival times corresponding to schedule penalties smaller than c should be used at capacity (otherwise they could be used without queuing and would have a cost smaller than c). Similarly, all arrival times corresponding to schedule penalties strictly larger than c cannot be used at equilibrium. Thus, the set of times $\{t \in \mathbb{R}, \mathrm{SP}(t) < c\}$ should be smaller than N/S and the set of times $\{t \in \mathbb{R}, \mathrm{SP}(t) > c\}$ should be larger than N/S. With the $\alpha - \beta - \gamma$ preferences, these two constraints impose that $c = \delta \frac{N}{S}$, where $\gamma = \frac{\beta\gamma}{\beta+\gamma}$, and the equilibrium is used only during the interval $[t_1, t_2]$, where $t_1 = t^* - \frac{N}{S}\frac{\delta}{\beta}$, $t_2 = t^* + \frac{N}{S}\frac{\delta}{\gamma}$.

The delays are then derived by ensuring that all used alternatives have the same total cost, as illustrated in Fig. 2.2a. The cumulative departure curve (cumulative inflow) is then built by translating the points on the cumulative arrival curve by the delays previously derived (see Fig. 2.2b).

2.4.3 Important results

Traditionally, the literature has often focused on the case of a morning commute with constant u_o. The existence (resp. uniqueness) of such an equilibrium was established by Smith (1984a) (resp. Daganzo (1985)) for the case of a homogeneous population

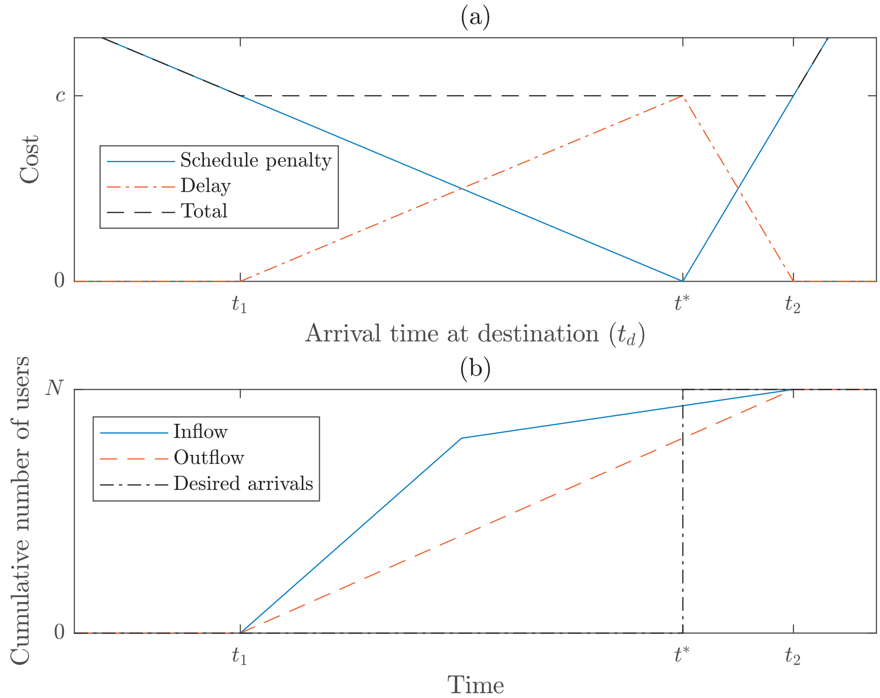

Figure 2.2 – Dynamics with a bottleneck model and homogeneous $\alpha - \beta - \gamma$ preferences

with continuous and (resp. strictly) increasing u_d.[5] Various types of heterogeneity were introduced by Newell (1987) and Arnott et al. (1988). Lindsey (2004) extended these results to allow for a more general dependence of utility on the arrival time.

The evening commute has been relatively less studied. Vickrey (1973) and Fargier (1983) described the analytical solution when users have a piece wise constant decreasing u_o and a constant u_d. de Palma and Lindsey (2002) provided existence and uniqueness conditions similar to those of Lindsey (2004) for more general decreasing u_o and Fosgerau and de Palma (2012) formulated existence and uniqueness results that do not require either u_o or u_d to be constant, and therefore are valid for both commutes. These results are however limited to the case of a continuously distributed free-flow travel time. While there exist some symmetries between the morning and evening commutes (e.g. in terms of total social cost with homogeneous users), they have different queuing dynamics (Vickrey, 1973).

[5]Smith (1984a) (resp. Daganzo (1985)) described their assumptions in terms of schedule penalty (see Section 2.2). Both descriptions are equivalent.

3 Stability of departure time choice with a bottleneck

3.1 Introduction

The problem of departure time choice during the morning and evening commutes involves two levels of dynamics: conditions vary over the day, and from day to day. The abundant research following Vickrey (1969) and Arnott et al. (1990) has focused mostly on the first of these two levels, implicitly assuming that if a unique equilibrium exists, commuters would naturally converge toward it.

Various pieces of evidence have cast doubt on the soundness of this assumption: de Palma (2000) and Mc Breen et al. (2006) reported simulation-based observations of non-converging behavior, Iryo (2008) provided some analytical evidence of instability in continuous time and Guo et al. (2018) recently proved that some dynamic processes cannot converge in discrete time. Yet, all these results rely on specific schedule preferences (the so-called $\alpha - \beta - \gamma$) and adjustment mechanisms, and it is not clear whether they hold under more general assumptions. It is not clear either what this instability means in practice: is it still reasonable to use an unstable equilibrium as an approximation of a dynamic process?

The present work tackles these two issues. On the theoretical side, this is the first work that formulates results valid for wide classes of schedule preferences and rational adjustment mechanisms. It builds on results from route choice and evolutionary game theory, and in particular on the concept of monotonic utility functions. These results, although only valid in continuous time, reveal fundamental differences in the dynamic properties of the morning and evening commutes. It is also shown how the same congestion toll can yield both monotonic and non-monotonic utility functions depending on where the toll booth is installed.

We recognize however that our theoretical derivations still rely on very ideal assumptions and investigate to which extent the results obtained are transferable to the real world.

We conclude with practice-relevant statistics on the error made when approximating trip scheduling problems by their equilibria. Although assumption-specific, these statistics illustrate how the equilibrium approximation can yield relatively good estimates for some metrics (e.g. the social cost), and very poor ones for others (e.g. the proportion of social cost represented by schedule penalties). The analytical and numerical results are presented in Sections 3.3 and 3.4. Section 3.2 provides an introduction to the trip scheduling problem and to selected concepts and results from evolutionary game theory.

3.2 Background

3.2.1 Notations and definitions

The following notations and definitions are largely borrowed from this literature, and in particular from Sandholm (2010b).

Consider a single homogeneous population, i.e. a continuum of users with the same schedule preferences. The demand is assumed inelastic and of size $N \in \mathbb{R}_+^*$, so that N agents commute every day. The choice set of departure times is finite and denoted $\mathrm{DT} = \{t_1, ...t_n\}$. The set of possible populations states is $X = \{(x_1, ...x_n) \in \mathbb{R}_+^n, \sum_{i=1}^n x_i = 1\}$, where x_i denotes the proportion of users choosing departure time t_i. For all $x \in X$, $\mathrm{TX}(x)$ denotes the set of directions that the system can follow without leaving X. Mathematically, $\mathrm{TX}(x) = \{z \in \mathbb{R}^n, \exists y \in X, \alpha \geq 0, z = \alpha(y - x)\}$.

A population game is characterized by a continuous utility function $U : X \to \mathbb{R}^n$, where $U_i(x)$ represents the utility of departing at time i when the population plays according to x.[1] Since agents have no mass, a population state x^* is a Nash equilibrium if $x_i^* > 0$ implies that $U_i(x^*) \geq U_j(x^*)$ for all $j \in \{1, ..., n\}$ or, equivalently, if $\langle y - x^*, U(x^*)\rangle \leq 0$ for all $y \in X$. All population games have at least one Nash equilibrium.[2]

3.2.2 Dynamic system modeling

The evolution of the state x over days can be modeled in many different ways. We chose here the most convenient assumptions: users update their decision according to a deterministic process, based only on the conditions observed on the previous day. To the best of our knowledge, all the previous works considering the trip scheduling problem as a doubly dynamic process relied on the same assumptions. Yet, other alternatives would also be worth considering, as argued by Watling and Hazelton (2003) for the route choice problem.

[1] The notation U_i should not be confused with the notations U_o and U_d.

[2] This result was shown for a broader class of noncooperative games with continuums of agents, cf. Sandholm (2010b)

Besides, the same dynamics can be modeled either in continuous time ($\dot{x}(t) = f(x(t))$) or in discrete time ($x_k = f(x_{k-1})$, where k denotes the day). The discrete alternative provides a more faithful representation of the real world commute, and it also lends itself to numerical simulations. Yet, the continuous time approximation is often more suitable for analytical derivations. The continuous time approximation is standard in evolutionary game theory and also common in route choice applications (Smith, 1984b; Dupuis and Nagurney, 1993). Since this chapter builds on results from these fields, we chose to adopt the continuous time approximation for the theoretical result of Section 3.3 and returned to a more realistic discrete time framework in Section 3.4 for numerical simulations.

3.2.3 Adjustment mechanisms

Population games are based on two main assumptions: inertia and myopia. Inertia means that agents only update their decision sporadically, so that the population state evolves continuously with time. Myopia means that agents only consider the current utilities of each alternative when taking decision, without trying to predict how other users will react. When agents do update their decision, they do so by following a protocol. We focus here on reactive protocols of the form $\rho : X \times \mathbb{R}^n \to \mathbb{R}_+^{n \times n}$, which map population states $x \in X$ and their corresponding utility vectors $u \in \mathbb{R}^n$ to matrices of conditional switch rates ρ_{ij}. Assuming that the revision protocol is Lipschitz continuous, it defines a "deterministic evolutionary dynamic", i.e. a map that assigns to each continuous utility function $U : X \to \mathbb{R}^n$ a system of ordinary differential equations (Sandholm, 2015):

$$\dot{x_i} = \sum_{j=1}^{n} x_j \rho_{ji}(x, U(x)) - x_i \sum_{j=1}^{n} \rho_{ij}(x, U(x)) \; \forall i \in \{1, ..., n\}. \tag{3.1}$$

To best represent reality, the revision protocol should rely on behavioral assumptions and information requirements that are consistent with the selected application. In biology for instance, the revision protocol is often imitative (agents are more likely to choose alternatives that are already widely used). In transportation however, the most commonly accepted principle is utility maximization, which is well illustrated by the best response protocol ($\rho_{ij}(x, u) \in \arg\max_{y \in X}(y'u)$). Yet, it is often more realistic to relax the utility maximization principle, for instance to allow for ρ_{ij} to be continuous, or to account for the fact that users may not know the utilities of all alternatives.

The class of impartial pairwise comparison dynamics addresses this two issues. It is characterized by $\rho_{ij}(x, u) = \phi_j(u_j - u_i)$, where ϕ_j are sign-preserving functions[3]. It can be interpreted as follows: every user regularly revises her decision by comparing her current strategy against a randomly selected alternative. If the alternative provides

[3]Note that $\rho_{ij}(x, u)$ is always non-negative, by definition of conditional switch rates. Thus, the sign-preserving constraint ensures that $u_j \leq u_i \Leftrightarrow \phi_j(u_j - u_i) = 0$.

a larger utility, the user adopts it with a probability that is a function of the utility difference. Otherwise, she retains her current strategy. In full generality, the function ϕ can be specific to the target alternative, but it does not need to be so. For instance, the most common protocol of this class is the proportional swap mechanism of Smith (1984b), defined by $\rho_{ij}(x,u) = [u_j - u_i]_+$ (where $\forall y \in \mathbb{R}, [y]_+ = \max(0,y)$). It leads to the following continuous dynamic:

$$\dot{x_i} = \sum_{j=1}^{n} x_j \left[U_i(x) - U_j(x)\right]_+ - x_i \sum_{j=1}^{n} \left[U_j(x) - U_i(x)\right]_+ . \tag{3.2}$$

A modified version of Smith's protocol is used in the simulations of Section 3.4. Since these simulations are in discrete time, the proportion of the population moving from alternative i to j from one day to the next was computed as

$$\frac{\lambda}{n} x_i \left[U_j(x) - U_i(x)\right]_+ \left(\max\left(1, \sum_{j=1}^{n} \frac{\lambda}{n} \left[U_j(x) - U_i(x)\right]_+\right)\right)^{-1} , \tag{3.3}$$

such that the state remains in the feasible space X. The ratio λ/n captures the sensitivity of users to utility variations and influences the step size. We refer hereafter to λ as the "sensitivity".

Another important category is the class of separable excess payoff dynamics. This class is such that users only switch to strategies associated to a payoff that is larger than the average payoff experienced in the population, and do so independently of their previous strategy. The corresponding protocols are of the form $\rho_{ij}(x,c) = \phi_j\left(\left(\sum_{k=1}^{n} x_k c_k\right) - c_j\right)$, where the functions ϕ_j are sign-preserving. For further details on these various classes of mechanisms, the reader is referred to Sandholm (2015).

3.2.4 Monotonicity and stability

Although the word "stability" is often used in a loose sense, analytical results require precise definitions. The following are three important types of stability, ranked from weakest to strongest.

Definition 1 (stable equilibrium)**.** *A state x_e is Lyapunov stable if for every $\epsilon > 0$, there exists $\delta > 0$ such that, if $|x(0) - x_e| < \delta$ then for every $t > 0$, $|x(t) - x_e| < \epsilon$.*

Definition 2 (asymptotically stable equilibrium)**.** *A state x_e is asymptotically stable if it is Lyapunov stable and there exists $\delta > 0$ such that if $|x(0) - x_e| < \delta$, then* $\lim_{t\to\infty}|x(t) - x_e| = 0$.

Definition 3 (globally asymptotically stable equilibrium)**.** *A state x_e is globally asymptotically stable if it is Lyapunov stable and if $\lim_{t\to\infty}|x(t) - x_e| = 0$ for any $x(0)$.*

The stability of a Nash equilibrium depends both on the game and on the revision protocol. Fortunately, recent progress in evolutionary game theory has identified classes of games and revision protocols that share many important dynamic properties. One such class, the class of stable (or contractive) games, is central to our work. Stable games are population games that are associated to a monotonic utility function. With $\langle \cdot, \cdot \rangle$ denoting the canonical scalar product ($\forall x, y \in \mathbb{R}^n$, $\langle x, y \rangle = \sum_{i=1}^n x_i y_i$), monotonic utility functions are defined as follows.

Definition 4 (Monotonic utility function)**.** *A utility function $U : X \to \mathbb{R}^n$ is monotonic if $\langle y - x, U(y) - U(x) \rangle \leq 0$ for all $x, y \in X$.*

In plain words, monotonicity requires that the average utility improvement of alternatives that are abandoned (weighted by how many users abandon them) is larger than the (also weighted) average utility improvement of alternatives that agents are switching to (improvements can be positive or negative). If U is continuously differentiable, Hofbauer and Sandholm (2009) explains that U is monotonic if and only if its Jacobian matrix $DU(x)$ is negative semidefinite with respect to TX for all $x \in X$. A classical example is the class of negative diagonal dominant games (i.e. games such that for all x and for all $i \in \{1, ...n\}$, $\frac{\partial U_i}{\partial x_i}(x) \leq 0$ and $\frac{\partial U_i}{\partial x_i}(x) \leq \sum_{j \neq i} \left| \frac{\partial U_i}{\partial x_j}(x) \right|$).

From a static point of view, it is easy to show that the set of Nash equilibria of a stable game is convex and that if the utility function is strictly monotonic at some Nash equilibrium (i.e. if the inequality in Definition 4 is strict for all pairs involving this Nash equilibrium), then this Nash equilibrium is unique (Smith, 1979; Hofbauer and Sandholm, 2009). Yet, as the name suggests, stable games also have important dynamic properties. We explained above that dynamic properties normally depend both on the game and on the revision protocol. For stable and continuously differentiable games however, Hofbauer and Sandholm (2009) showed that the set of Nash equilibria is globally asymptotically stable with a wide range of adjustment mechanisms, including all those presented in Section 3.2.3 (best response, impartial pairwise comparison, separable excess payoff) and the class of perturbed best response mechanisms, whose best-known instance is the logit model. Similar local stability results also exist for cases where the utility function is not monotonic everywhere, but only locally (local monotonicity is formalized by the concept of evolutionary stable state) (Sandholm, 2010a).

In order to better understand the role of monotonicity, it is useful to read an example of Lyapunov stability proof (see for instance the proof of Theorem 7.1 in Hofbauer and Sandholm (2009)). These proofs proceed by showing that some global disequilibrium index $V(x)$ (called the Lyapunov function, and specific to the adjustment mechanism chosen) decreases over time when the system follows the adjustment mechanism. One way of proving this is by analyzing $\frac{\mathrm{d}}{\mathrm{d}t}V(x(t)) = \nabla V(x)'\dot{x}$, where $\nabla V(x)$ denote the gradient of V at x. V is typically a function of the form $V(x) = \sum_{i=1}^n x_i \psi_i(x, U(x))$, where $\psi_i(U(x), x)$ is the contribution of users currently playing i to the index $V(x)$. Thus, when users shift to other alternatives, they impact $V(x)$ by (i) changing the weights x_i associated to each contribution $\psi_i(U(x), x)$, and by (ii) changing the contributions $\psi_i(U(x), x)$ themselves. Then, the proof consists in showing that the adjustment mechanism is such that users move on average from alternatives having a large contribution $\psi_i(U(x), x)$ to alternatives having a smaller one (so that the term (i) is strictly negative) and that the term (ii) is always non-positive because of the monotonicity condition.

This summary, although highly schematic, provides some useful intuition. Note in particular that $V(x)$ might also be decreasing if the utility function is not monotonic but the first term dominates the second. This is likely the case far from equilibrium, when users can greatly increase their utility, and thus reduce their contribution to $V(x)$. Yet, if $\langle y - x, U(y) - U(x) \rangle$ is only positive for very specific states x and y, and if the revision protocol does not bring the system in such states, the first term may also dominate the second one close to equilibrium. This highlights the fact that monotonicity is a sufficient condition for stability, but not a necessary one.

3.3 Monotonicity with the bottleneck model

The main result of this Section is Theorem 1, which provides a necessary and sufficient condition for the monotonicity of utility functions with a bottleneck model. It extends the non-monotonicity result established by Iryo (2008) for the well-known $\alpha - \beta - \gamma$ preferences. Before stating it, we propose to analyze a simple but insightful example.

Note also that we consider in this section that departures occur continuously over time, to simplify the analysis. The set of possible departures rates is then the class of non-negative real-valued functions which are measurable and essentially bounded on a compact time interval, denoted $\mathcal{L}^\infty(\mathbb{R})$ (we follow here Mounce (2006)). The concept of a monotonic utility function remains valid with the canonical scalar product of that space $(\forall f, g \in \mathcal{L}^\infty(\mathbb{R}), \langle f, g \rangle = \int_\infty^\infty f(t)g(t)\,\mathrm{d}t)$.

3.3.1 A simple example

Let us first leave aside the schedule preferences and consider that utility is simply the negative of travel time. The congestion mechanism described by the constant capacity bottleneck model implies highly asymmetric externalities: users are only delayed by those traveling before them, and only delay those traveling after them. Also importantly, these externalities do not vanish with time: all users traveling after some perturbation are delayed by the same amount, as long as the queue does not entirely disappear. This is illustrated in Fig. 3.1.

Consider now a constant capacity bottleneck that is consistently congested during some time period, and the departure rate modification illustrated in Fig. 3.2, occurring entirely

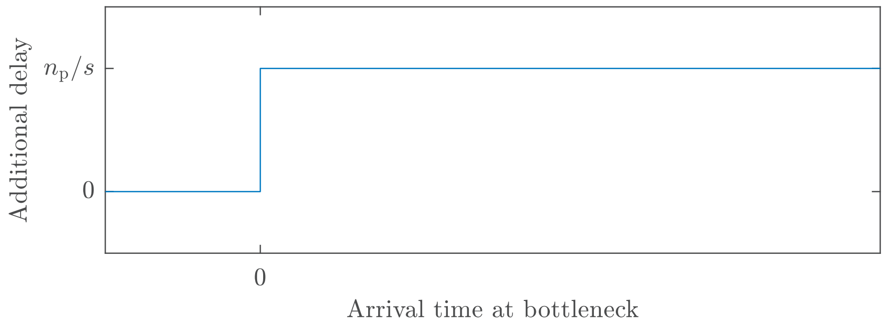

Figure 3.1 – Additional delay imposed on every other user by a user of mass n_p arriving at time $t = 0$ at a bottleneck of capacity s, as a function of others' arrival time at the bottleneck. The bottleneck is assumed to be congested for the entire period of interest.

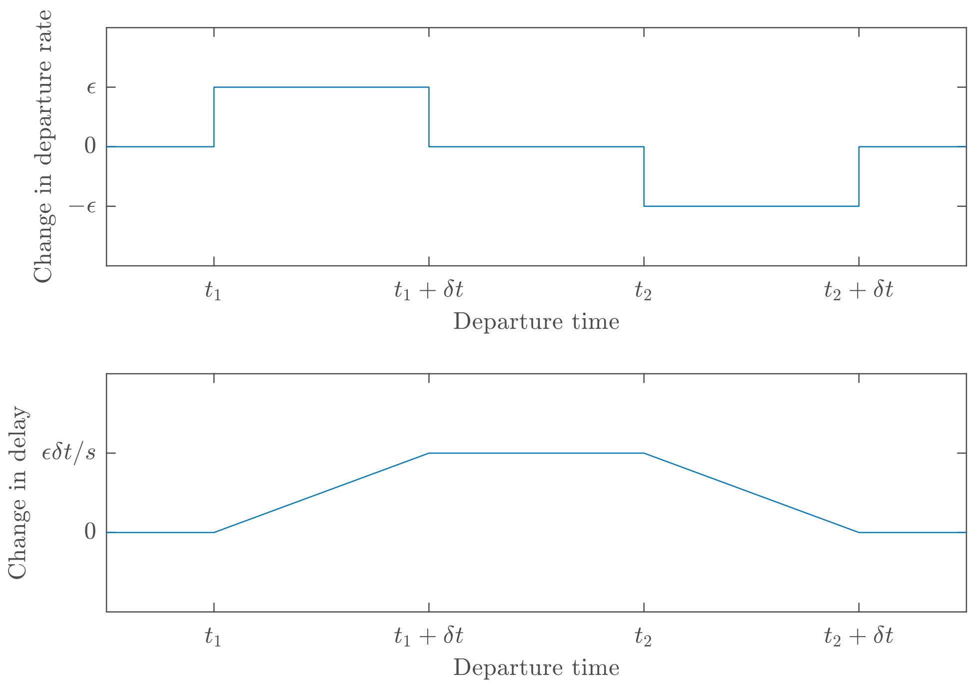

Figure 3.2 – Influence of a small perturbation of the departure rate on experienced delays.

inside the congested period. Let r_1 and r_2 denote the original and modified departure rates. If utility is simply equal to $-(t_\mathrm{d} - t_\mathrm{o})$ (i.e. u_o and u_d are both equal to the same constant), we have that $\langle r_2 - r_1, U_{r_2} - U_{r_1} \rangle = \int_{-\infty}^{\infty} (r_2(t) - r_1(t))(U_{r_2}(t) - U_{r_1}(t))\,\mathrm{d}t = \int_{t_1}^{t_1+\delta t} \frac{\epsilon^2}{s}(t - t_1)\,\mathrm{d}t + \int_{t_2}^{t_2+\delta t} -\frac{\epsilon^2}{s}(t_2 + \delta t - t)\,\mathrm{d}t = 0$. In other words, the change in utility function is perpendicular to the change in departure rate. If users were previously indifferent between the time slots $[t_1, t_1 + \delta t]$ (where users were added) and $[t_2, t_2 + \delta t]$ (where users were removed), they are still indifferent after the departure rate modification. By combining similar changes, we obtain that any departure rate modification that does not affect the congested period and the total number of departures yields changes in utility functions that are perpendicular to the changes in departure rate.

Now imagine that u_d slightly increases between t_1 and $t_2 + \delta t$: the delays would be more costly for users traveling after this increase than for those traveling before it, so that $\langle r_2 - r_1, U_{r_2} - U_{r_1} \rangle$ would be positive, thus violating the monotonicity condition. This small example shows that for the scheduling problem to be monotonic, it is necessary that the marginal utility rate at the destination u_d is non-increasing. This condition is actually also sufficient, as formalized by Theorem 1.

3.3.2 Monotonicity of the laissez-faire policy

Theorem 1. *Assume that $u_d(t) > 0$ for all t. With a bottleneck of constant capacity, the bottleneck utility function is a monotonic function of the flow into the bottleneck if and only if the marginal utility function at the destination u_d is a non-increasing function of time.*

Although this is the first work providing conditions for the monotonicity of the trip scheduling problem with a bottleneck, this result is intricately related to theorems established for route choice applications (without schedule preferences). Indeed, Smith and Ghali (1990) and Mounce (2006) already investigated the monotonicity of the utility function that arises from the bottleneck congestion mechanism when individual costs are reduced to travel times. Smith and Ghali (1990) established the monotonicity of the constant-capacity problem, while Mounce (2006) established a result similar to ours with an exogenous time-dependent bottleneck capacity.

Theorem 2 (Mounce (2006))**.** *Without schedule preferences, the bottleneck delay function is a monotonic function of the flow into the bottleneck if and only if the bottleneck capacity is non-decreasing with respect to time.*

Rather than proving Theorem 1 from scratch, we take advantage of these existing results and simply prove that the case with schedule preferences can actually be seen as another interpretation of Mounce's result.

Proof of Theorem 1. Let $s(t)$ denote the capacity of the bottleneck at time t and let $A(t) = \int_0^t s(\tau)\,\mathrm{d}\tau$ denote the maximum possible cumulative number of arrivals at destination at time t, starting from time 0. If a user arrives at the bottleneck at time t_o, she will leave it at time

$$t_\mathrm{d}(t_\mathrm{o}) = \begin{cases} t_\mathrm{o}, & \text{if } q(t_\mathrm{o}) = 0 \\ t_\mathrm{o} + A^{-1}(A(t_\mathrm{o}) + q(t_\mathrm{o})), & \text{otherwise.} \end{cases}$$

Thus, her utility would be

$$U(t_\mathrm{o}, t_\mathrm{d}(t_\mathrm{o})) = \begin{cases} U_\mathrm{o}(t_\mathrm{o}) + U_\mathrm{d}(t_\mathrm{o}), & \text{if } q(t_\mathrm{o}) = 0 \\ U_\mathrm{o}(t_\mathrm{o}) + U_\mathrm{d}(t_\mathrm{o}) - \int_{t_\mathrm{o}}^{A^{-1}(A(t_\mathrm{o})+q(t_\mathrm{o}))} u_\mathrm{d}(\tau)\,\mathrm{d}\tau, & \text{otherwise.} \end{cases}$$

By making the change of variable $a = A(t)$, her utility becomes

$$U(t_\mathrm{o}, t_\mathrm{d}(t_\mathrm{o})) = \begin{cases} U_\mathrm{o}(t_\mathrm{o}) + U_\mathrm{d}(t_\mathrm{o}), & \text{if } q(t_\mathrm{o}) = 0 \\ U_\mathrm{o}(t_\mathrm{o}) + U_\mathrm{d}(t_\mathrm{o}) - \int_{A(t_\mathrm{o})}^{A(t_\mathrm{o})+q(t_\mathrm{o})} \frac{u_\mathrm{d}(A^{-1}(a))}{s(A^{1-}(a))}\,\mathrm{d}a, & \text{otherwise.} \end{cases}$$

Thus, the utility function does not depend on u_d and s separately, but only on their ratio. In particular, a problem with constant capacity s^0 and time varying marginal utility at work $u_\mathrm{d}^0(t)$ has the same utility function as a problem with time varying capacity $s^0/u_\mathrm{d}^0(t)$ and constant marginal utility at work equal to 1. In such a case, the utility of each alternative decreases linearly with travel time, as in Theorem 2. The terms $U_\mathrm{o}(t_\mathrm{o})+U_\mathrm{d}(t_\mathrm{o})$ play the role of alternative-specific constants, which do not impact monotonicity. Thus, Theorem 1 follows from Theorem 2. $\square$

One of the main insights offered by Theorem 1 is that the morning and evening commutes may have very different dynamic properties. Indeed, u_d is usually an increasing function of time for the morning commute (see for instance the empirical work of Tseng and Verhoef (2008)), while it is typically assumed to be constant in the evening commute. The utility function would thus be non-monotonic during the morning but monotonic during the evening, although this latter case should be considered carefully as a constant u_d represents a limit case. If users were indeed updating their decisions in continuous time, this difference would be extremely significant as it would imply stability for the evening commute (with the caveat that it is a limit case), and possible instability for the morning commute.[4] Thus, our extension brings an entirely new significance to the

[4]Another caveat is that most of the analytical results establishing stability for monotonic utility functions are only established for discrete choice sets. Continuous choice sets involve additional technical difficulties, but Mounce (2006) circumvented them to establish global asymptotic stability with Smith's adjustment mechanism. We are not aware of similar extensions for other mechanisms. Note however that utility functions that are monotonic with Mounce's class of departure rate are also monotonic with the more restrictive class of piece-wise constant departure rates considered in Section 3.2. The converse is not generally true, but if u_d is increasing on a time scale that is larger than the time step between two possible departure times (considered to be 1 min in our numerical applications), then counter-examples

earlier monotonicity results of Smith (1984b) and Mounce (2006).

3.3.3 Pricing and monotonicity

It has long been known that by setting appropriate prices on all alternatives, network authorities can avoid wasteful queuing and ensure that the tolled user equilibrium coincides with the social optimum (Vickrey, 1969). Yet, as for the no-toll equilibrium, it is generally not clear whether the tolled equilibrium is stable. As evidenced hereafter, the answer depends on how the toll is implemented.

Consider a homogeneous population. In order to minimize the total cost, one should find a departure rate $r(t)$ that does not generate any queuing and such that all the used departure times t have a no-queue utility $U(t,t)$ larger than all the unused departure times. Although some additional assumptions are required for uniqueness, such an optimum always exists (because utility is continuous and the choice set compact), and one can associate to it a utility level $\bar{U} = \min_{t \in \mathrm{supp}(r)}\{U(t,t)\}$ (where $\mathrm{supp}(r)$ denotes the support of r). By definition of the social optimum, $\bar{U}$ can also be defined as the maximum of $\{u \in \mathbb{R}, \int [U(t,t) > u] s \,\mathrm{d}t \geq N\}$, where $[X] = 1$ if X is true and 0 otherwise (Iverson brackets).

Then, in order to make the social optimum an equilibrium, one should ensure that all used alternatives have the same utility when there is no queuing. We can ensure this is the case by adding a toll either upstream or downstream of the bottleneck. We propose the same time dependence in both cases:

$$\$(t) = \max(U_\mathrm{o}(t) + U_\mathrm{d}(t) - \bar{U}, 0), \tag{3.4}$$

Note that this formulation depends on the time of the day, but that the toll remains the same from day to day, independently of the traffic conditions.

The choice of the toll booth location has some consequences. If it is downstream of the bottleneck, the toll is applied to t_d, such that whenever $\$(t_\mathrm{d}) > 0$, $U(t_\mathrm{o}, t_\mathrm{d}) = U_\mathrm{o}(t_\mathrm{o}) + U_\mathrm{d}(t_\mathrm{d}) + \bar{U} - U_\mathrm{o}(t_\mathrm{d}) - U_\mathrm{d}(t_\mathrm{d}) = U_\mathrm{o}(t_\mathrm{o}) + \bar{U} - U_\mathrm{o}(t_\mathrm{d}) = U_\mathrm{o}(t_\mathrm{o}) + \bar{U} + \int_{t_\mathrm{d}}^{0} u_\mathrm{o}(t)\,\mathrm{d}t$. This utility function has the same structure as in the no-toll case, except that $u_\mathrm{o}(t)$ in the tolled case plays the role of $u_\mathrm{d}(t)$ in the no-toll case. Thus, if the choice set was reduced to the times such that $\$(t_\mathrm{d}) > 0$, the utility function would be monotonic if and only if u_o is non-increasing (which is likely to be the case during the morning commute). On the other hand, if the toll booth is installed upstream of the bottleneck, the toll is applied to t_o and whenever $\$(t_\mathrm{o}) > 0$, $U(t_\mathrm{o}, t_\mathrm{d}) = U_\mathrm{o}(t_\mathrm{o}) + U_\mathrm{d}(t_\mathrm{d}) + \bar{U} - U_\mathrm{o}(t_\mathrm{o}) - U_\mathrm{d}(t_\mathrm{o}) = U_\mathrm{d}(t_\mathrm{d}) + \bar{U} - U_\mathrm{d}(t_\mathrm{o})$. If the choice set was reduced to the times such that $\$(t_\mathrm{o}) > 0$, this utility function would be monotonic if and only if u_d is non-increasing, as in the no-toll case.

such as the one illustrated in Fig. 3.2 can easily be constructed.

Theoretical results are limited by the fact that users might still choose from time to time to depart at times that are not used in the social optimum, and for which the monotonicity results above do not hold. This problem could be addressed by setting also an appropriate price on suboptimal alternatives. In practice however, such alternatives should not be chosen too often as they are always dominated.

3.4 Discrete time analysis (simulation-based)

3.4.1 Objectives

Combined with the existing results on stable games (see Section 3.2.4), the analytical results covered in Section 3.3 establish some sufficient conditions for the continuous time stability of the trip scheduling problem with homogeneous users. Yet, these are not sufficient for deriving practical recommendations and several issues remain to be investigated.

First, we shall investigate whether systems that are stable in continuous time are also stable in discrete time. The distinction is well explained by Watling (1999). In general, the stability of discrete-time systems depends on how much the state can evolve from iteration to iteration. In previous simulation reports (de Palma, 2000; Mc Breen et al., 2006), the proportion of agents updating their decisions every day was defined exogenously. Here this proportion is endogenous but it can be tuned via the sensitivity λ, in Eq. (3.3).

Second, we need to determine whether non-monotonic situations are unstable. In the simple set-up considered in Section 3.3.1 (homogeneous users with a long queuing interval), monotonicity violations are so common that instability appears very likely. It is not clear however whether this would also be the case in more realistic set-ups. Mc Breen et al. (2006) showed that user heterogeneity can have a significant stabilizing effect. Specifically, the authors considered a population of users having the same function u_d, but translated by various offsets to produce heterogeneity in t^*. It can be interpreted either as heterogeneity in the timing of the utility drop/surge, or as heterogeneity in the constant travel time required between the bottleneck location and the location where marginal utility changes. Intuitively, the stabilizing effect comes from the fact that it reduces both the range of departures that each user considers (and thereby reduces the shifting frequency with Smith's mechanism), and the frequency of unrealistic shifts, where users move from the congestion onset to the offset, and vice versa. Such shifts are problematic as (i) they perturb many users (recall that a user shifting from a departure time t to another t' only affects users that arrive at the bottleneck between t and t') and (ii) they involve great variations in $u_\mathrm{d}(t)$, which were shown in Section 3.3 to violate the monotonicity property for the morning commute. This intuition is supported by another observation of Mc Breen et al. (2006) that heterogeneity in the strength of the preferences alone does not have a similar stabilizing effect (without heterogeneity in t^*, users were indifferent at equilibrium between two departures times that were respectively in the congestion onset and offset). The amount of heterogeneity in t^* is parameterized in the present work by its standard deviation, denoted σ^*.

We shall further precise the influence of monotonicity on stability. To that end, we try to isolate this effect. Without pricing, this is done by considering a family of schedule preferences parameterized by a coefficient δ that influences monotonicity while maintaining some symmetry. With pricing, we simply compare the cases with toll booths installed upstream and downstream of the bottleneck. This comparison differs from the previous one because tolls are designed to avoid queues at equilibrium. Since the externalities behind the monotonicity property are generated by queues, monotonicity may not influence stability as much as in the no-toll case.

Finally, practitioners ultimately want to know whether the equilibrium approximation is reasonable. This issue cannot be reduced to a stability analysis. For instance, a system that would move fast towards its equilibrium and then oscillate close to it could be well approximated by this equilibrium, even though it is unstable. On the contrary, a system that would approach its equilibrium steadily but slowly may always be far from it. Thus, this section also aims at quantifying disequilibrium and its consequences in practice.

3.4.2 Description

Scenarios

In order to mimic the evolution of real world commutes, simulations were run for 200 days, starting from uniformly distributed departures.[5] We could have run the simulations over longer time periods but real world systems do not remain unperturbed for so long. The bottleneck capacity s was chosen such that it takes 2 h to serve all users (with a unit population, $s = 1/2$). The scenarios considered are all symmetric: the average desired arrival time t^* is always 0, and the schedule preferences are such that at equilibrium, the bottleneck is used during the interval $[-1 \text{ h}, 1 \text{ h}]$.

The possible departure times were evenly spaced over the interval $[-1.5 \text{ h}, 1.5 \text{ h}]$, every 1 minute. In practice however, the simulator was designed such that users choosing to depart at time t_i were actually departing uniformly over the interval $[t_i - \frac{\delta t}{2}, t_i + \frac{\delta t}{2}]$, where $\delta t = 1$ min is the time between two consecutive departure alternatives. The resulting departure rate is piece-wise constant, allowing for simple analytical derivations of the queue dynamics. The utility associated to alternative i was computed based on the distribution of arrival times of users departing in $[t_i - \frac{\delta t}{2}, t_i + \frac{\delta t}{2}]$.

Variations in marginal utility rates were produced by the function $f_\delta(t) = \alpha + \frac{\delta}{\pi} \tan^{-1}(w(t-$

[5]The code used for the following simulations is available at https://github.com/raplam/departureTimeChoice.

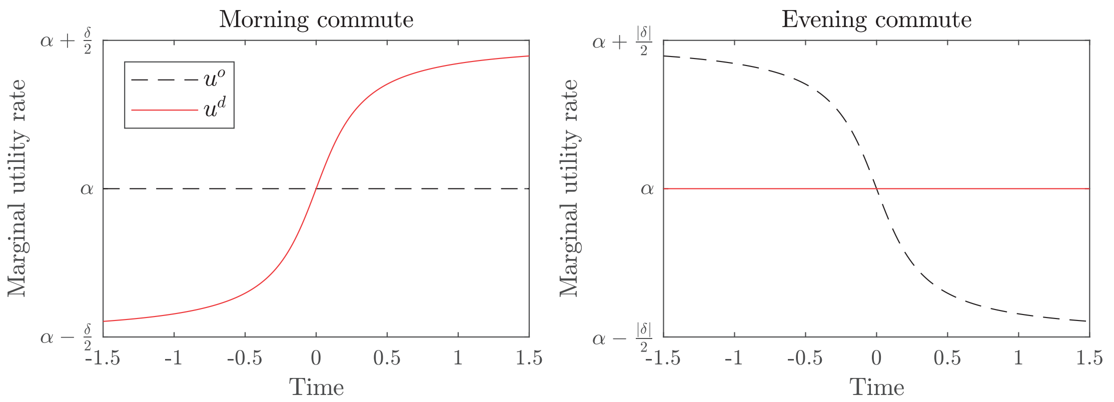

Figure 3.3 – Marginal utility rates for morning and evening commutes

t^*)), with $w = 4$ h^{-1}. We refer to δ as the "amplitude". Depending on the sign of δ, the function f_δ generates either a utility surge or a utility drop. We created "morning commutes" by using $u_{\mathrm{o}}(t) = \alpha = 1$ and $u_{\mathrm{d}}(t) = f_\delta(t)$ with $\delta > 0$ and "evening commutes" by using $u_{\mathrm{d}}(t) = \alpha = 1$ and $u_{\mathrm{o}}(t) = f_\delta(t)$, with $\delta < 0$. These marginal utility rates are illustrated in Fig. 3.3. The magnitude $|\delta|$ determines the strength of the schedule preferences. In a morning commute scenario, users that really need to arrive at work around t^* would have a large and positive δ.

Heterogeneity was introduced by creating $G = 10$ groups and allocating each of them a different t^*.[6] Different standard deviations σ^* were produced by varying the gap between consecutive values. For any $g \in \{1, ...G\}$, the value t^* of that group was set to $t^*_g = \left(g - \frac{G+1}{2}\right)\sqrt{\frac{12}{G^2-1}}\sigma^*$.

Disequilibrium index

Some objective function is needed to quantify convergence. Here, we consider a rather intuitive measure, the "potential gain". It is defined as the average over all users of the maximum utility improvement they could achieve given the current utility functions, divided by their current congestion cost.[7] To simplify its expression, we defined the utility functions U_{o} and U_{d} of each group of user relatively to their own t^*: for any $g \in \{1, ...G\}$, $U^g_{\mathrm{o}}(t_{\mathrm{o}}) = \int_{t^*_g}^{t_{\mathrm{o}}} u_{\mathrm{o}}(t)\,\mathrm{d}t$ and $U^g_{\mathrm{d}}(t_{\mathrm{d}}) = \int_{t_{\mathrm{d}}}^{t^*_g} u_{\mathrm{d}}(t)\,\mathrm{d}t$. With these definitions, the maximum possible utility ($U^g(t^*_g, t^*_g)$) is equal to 0. The congestion cost is thus the negative of utility and if we denote x^g_i the proportion of the population belonging to

[6]We initially considered other values but the aggregate indicators were extremely similar for all $G \geq 10$. $G = 10$ was found to be a good compromise between granularity and computational resources.

[7]The congestion cost of an alternative i for users of group g is defined as the difference between $U^g_i(x)$ and the maximum possible utility without congestion, reached when $t_{\mathrm{o}} = t_{\mathrm{d}} = t^*_g$.

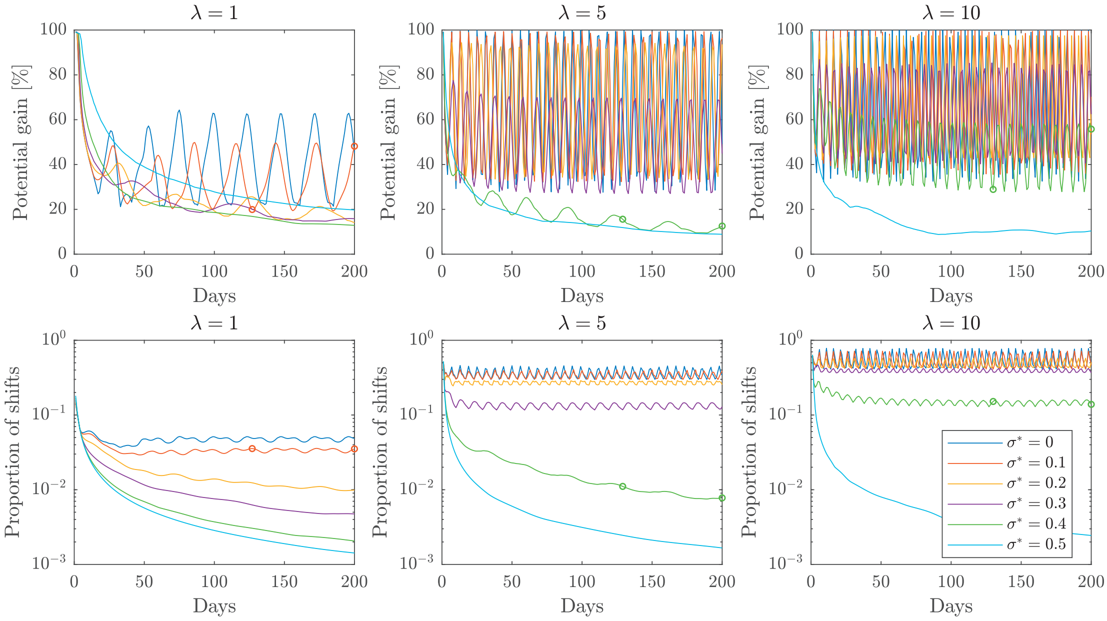

Figure 3.4 – Influence of heterogeneity in t^* on convergence, with different values of the sensitivity λ ($\delta = 1.5$).

group g and choosing alternative i, the potential gain is given by

$$\mathrm{PG}(x) = \sum_{g=1}^{G} \sum_{i=1}^{n} x_i^g \frac{\max_{j\in\{1,\dots,n\}} \left(U_j^g(x) - U_i^g(x) \right)}{|U_i^g(x)|}. \tag{3.5}$$

It belongs to $[0, 1]$ and is equal to zero if x is an equilibrium.

3.4.3 Results

Influence of λ, σ^* and δ on convergence

Fig. 3.4 shows the evolution of the potential gain $\mathrm{PG}(x)$ for multiple heterogeneity levels σ^* and for various sensitivity parameters λ. These runs can be classified in two categories: those where $\mathrm{PG}(x)$ oscillates very early, and those where $\mathrm{PG}(x)$ exhibits a global decreasing trend (although at various rates). When analyzing simulations, we abusively brand the latter behavior as converging. For each value of λ, there seems to be a critical σ^* such that runs with less heterogeneity oscillate, while those with more heterogeneity converge. The critical value of σ^* apparently increases with λ, such that for a given σ^*, the system is more likely to converge with a smaller λ. However, increased stability comes at the cost of slower convergence. With $\sigma^* = 0.5$ for instance, it takes 86 days to reach $\mathrm{PG}(x) = 30\%$ with $\lambda = 1$, but only 12 days with $\lambda = 10$.

These results challenge the idea of a binary relation between the slope of u_d and stability.

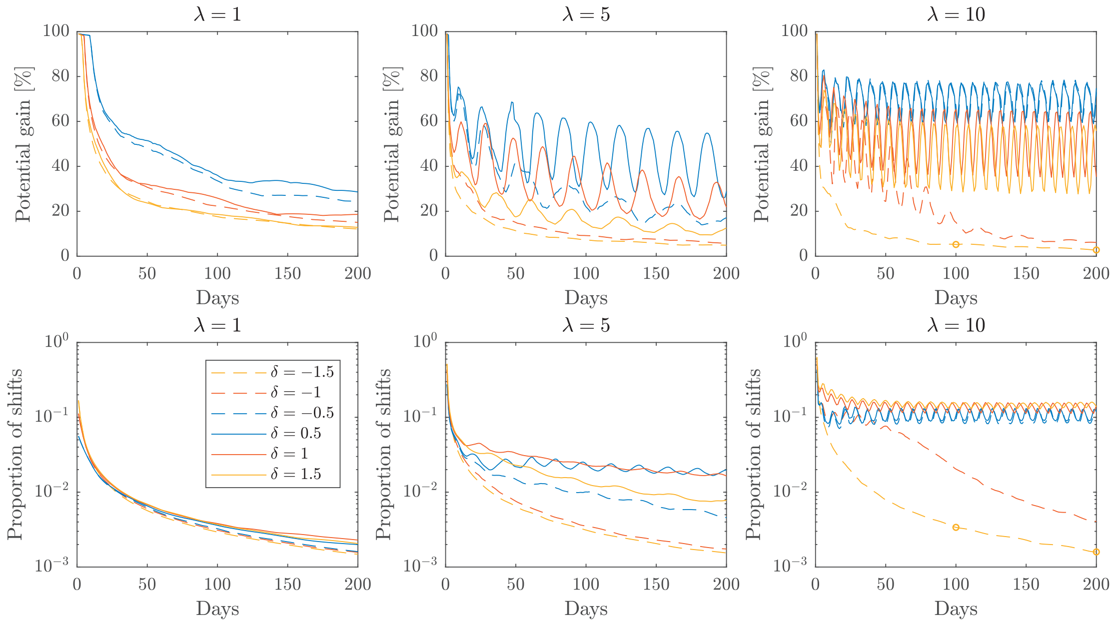

Figure 3.5 – Influence of the amplitude δ on convergence, with different values of the sensitivity λ ($\sigma^* = 0.4$).

Indeed, the above cases all correspond to non-monotonic utility functions[8], and yet some of them converge (at least from a practical point of view). Yet, this does not mean that monotonicity is irrelevant. Fig. 3.5 is the analogue of Fig. 3.4, except that the graphs are replicated for various amplitudes δ, instead of various standard deviations σ^* (for brevity, we only present results for $\sigma^* = 0.4$). The sign of δ is especially important. With our schedule preferences, Theorem 1 implies that in the homogeneous case, the utility function is monotonic if and only if the perturbation has a positive amplitude ($\delta > 0$).

Here, runs with positive and negative amplitudes having the same magnitude $|\delta|$ behave identically during the first iterations (this is because the two commutes are pure mirror images as long as there is no queuing), and then diverge. When they diverge, several trends can be distinguished: (i) both runs exhibit similar oscillations (e.g. when $\lambda = 10$ and $|\delta| = 0.5$), (ii) both runs converge and the run with negative amplitude δ converges at the same rate or faster (e.g. all cases with $\lambda = 1$), (iii) only the run with negative amplitude converges (e.g. when $\lambda = 10$ and $|\delta| = 1$ or 1.5). Thus, even if the sign of δ is not the only determinant of stability, these simulations show that monotonicity does have a stabilizing effect and that evening commutes are generally closer to equilibrium than the symmetric morning ones.

Finally, the evolution of the proportions of users updating their decision displayed in the lower panels of Fig. 3.4 and 3.5 provide some orders of magnitude regarding the decision-

[8]With group-based heterogeneity, clear monotonicity violations can be obtained by considering intra-group movements identical to those considered in Section 3.3.1.

making process. When the system is relatively close to equilibrium, users typically update their departure time only about 0.3% of days, i.e. approximately once a year. Yet users with the same parameters can also react quite quickly to major perturbations, as shown during the first days of the simulation.

Departure rate analysis

The effect of user heterogeneity and the practical consequences of instability are better understood when considering the departure profiles. Each row of Fig. 3.6 shows as functions of time (i) the cumulative input and output of the bottleneck (1^{st} column) and (ii) the departure rates of all groups, for two days of the previous simulations (2^{nd} and 3^{rd} columns). These days, identified by circles in the Figs 3.4 and 3.5, were selected among the last 100 to avoid a strong dependence on the initial conditions. We chose them to be as different as possible in terms of potential gain. Thus, the cumulative plots illustrate the magnitude of variations in queue length (the vertical distance between the input and output curves) and travel time (the horizontal distance) observable over a single run.

The situations illustrated in the first two rows correspond to highly oscillatory behaviors, with large potential gains. Since the output is constrained by the bottleneck capacity, these oscillations mostly affect the cumulative input. These oscillations also influence the distribution of departures within each group and convergence seems highly correlated with a strong segregation between groups. Indeed, the 2^{nd}, 3^{rd} and 4^{th} rows correspond to situations with the same amount of heterogeneity ($\sigma^* = 0.4$) but users are clearly more segregated in situations with smaller oscillations and smaller potential gain. This observation supports the intuition that heterogeneity brings stability by restraining the scope of alternatives considered by users. When heterogeneity is not sufficient to effectively restrain this scope, the system remains unstable.

Finally, note that the amount of delay (i.e. the area between the cumulative input and output curves) is larger in the first row than in the others. Is it due to instability or to different schedule preferences? This is investigated in Section 3.4.3, together with other practical consequences of instability.

Pricing

Pricing is also likely to affect stability, and not only through the concept of monotonicity. To simplify the analysis, we focus here on the morning commute with homogeneous users. Without pricing, this case was the furthest from equilibrium.

Fig. 3.7 represents the potential gain evolution in different situations. All three panels contain 30 curves, corresponding to five initial points, three values of λ and two tolling strategies. The initial conditions considered correspond to the five first iterations of a

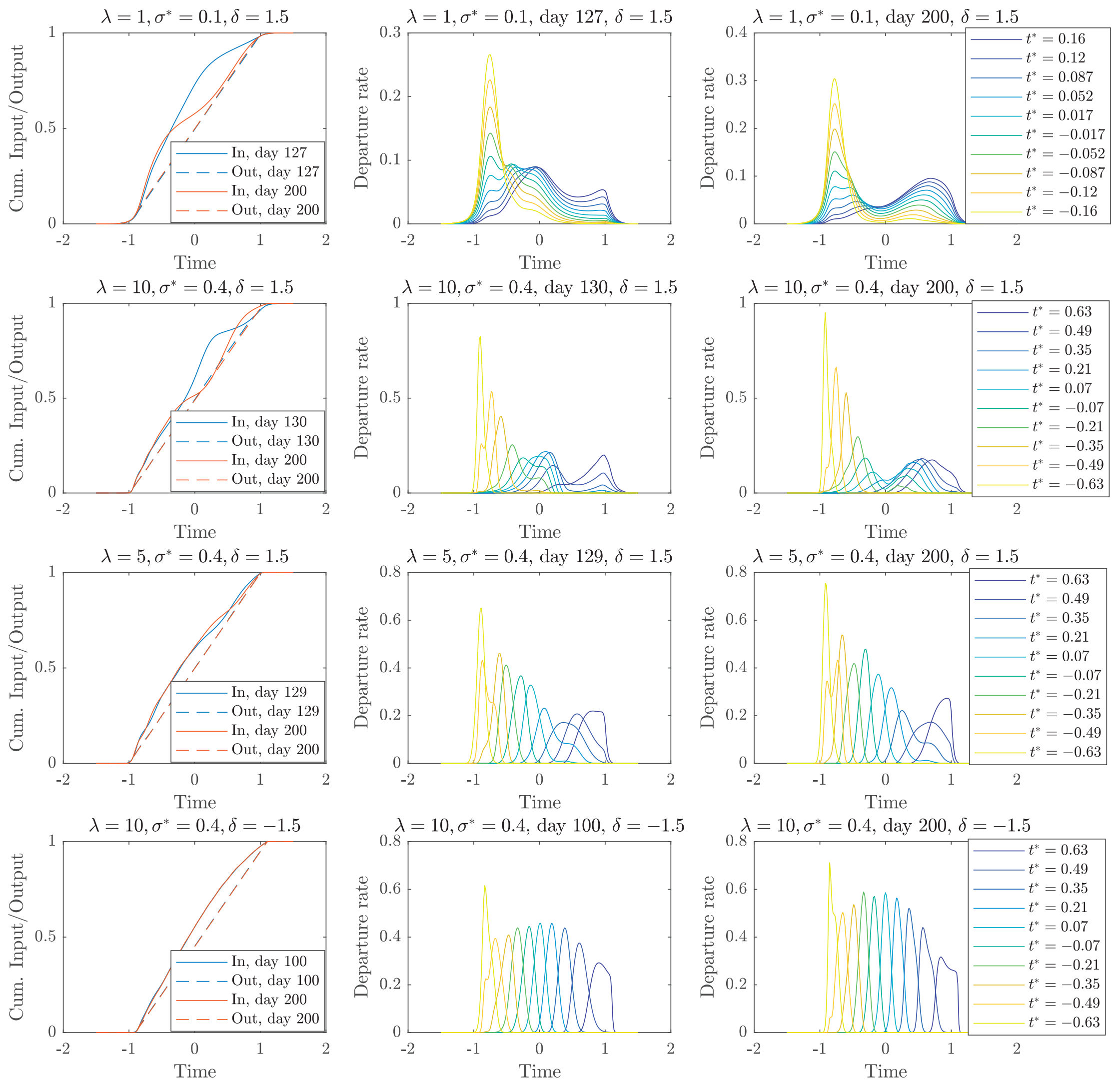

Figure 3.6 – Cumulative input/output curves and disaggregated departure profiles for pairs of days belonging to the same run, for various triplets $(\lambda, \sigma^*, \delta)$.

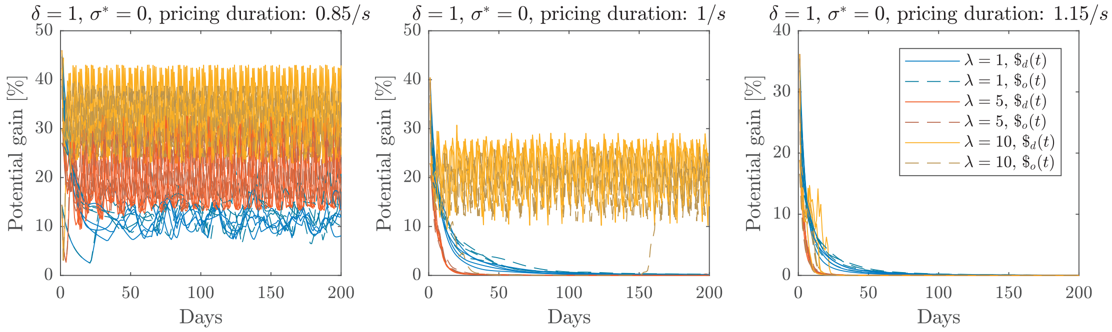

Figure 3.7 – Influence of the pricing duration on convergence, with homogeneous users ($\delta = 1$).

run made with $\lambda = 10$ without tolling. This ensures that the results are robust to queues, which are neither present in the social optimum, nor when departures are uniformly distributed. Each panel then corresponds to a different pricing scheme: the population considered when defining the pricing strategy was either underestimated by 15% (left), perfectly estimated (middle), or overestimated by 15% (right).

The comparison of these three panels shows that the stability greatly depends on the pricing duration, but that there is no noticeable difference in behavior between the scenarios with tolls upstream ($\$_\text{o}$) and downstream ($\$_\text{d}$) of the bottleneck. This can be explained by the shape of the utility preferences. With large pricing time windows, all travelers can use priced alternatives. Without congestion, these alternatives are the most attractive. If some of these alternatives happen to be overused, their utility is lowered. But since priced alternatives cannot be all overused simultaneously, there is always some priced alternative that is not overused and that is more attractive than unpriced ones. Thus, pricing incites users to spread on all the priced time interval, independently of the monotonicity of the utility function. Such a mechanism intuitively has a stabilizing effect. For instance, in continuous time and with a best-response adjustment mechanism, the priced equilibrium would be stable regardless of the monotonicity of the utility function, because users would only switch to alternatives that are priced and not overused. If the pricing duration is too short however, users have to travel also during unpriced intervals. The utility function is not monotonic in that case, irrespective of the toll booth location.

Biases in congestion cost estimation

It should be clear by now that without appropriate tolls, the real world is unlikely to be exactly at equilibrium. Yet, it is also clear that alternatives to the equilibrium approximation are likely to be far more complicated and impractical. We should therefore identify the impacts of this approximation and the biases it entails. While other indicators may be considered, we focus here on the estimation of the congestion cost and on its decomposition between delay and schedule penalties. Since we only consider cases where

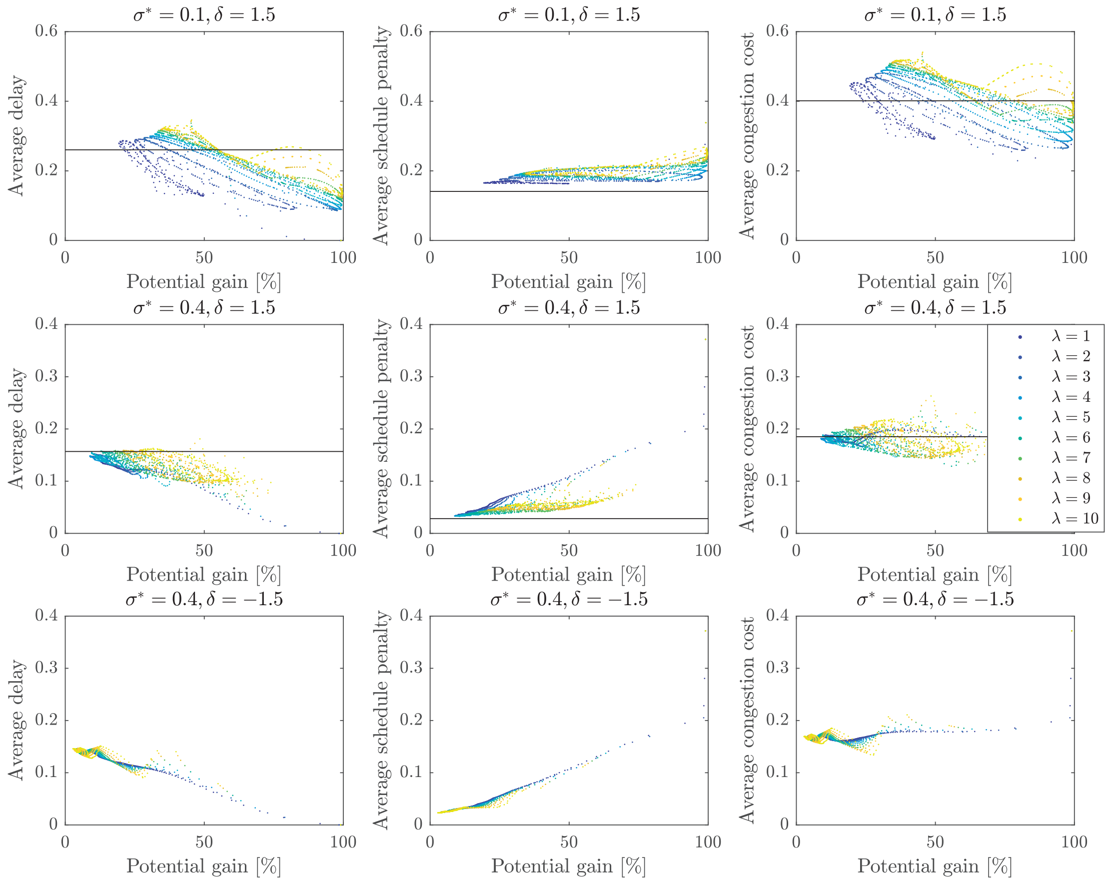

Figure 3.8 – Delay, schedule penalty and congestion cost observed over individual days for various pairs (σ^*, δ) and sensitivity λ.

either u_o or u_d is constant, we define the cost of delay as the travel time $(t_d - t_o)$, multiplied by this constant marginal utility rate. The schedule penalty is then the difference between the congestion cost and the delay cost.

Fig. 3.8 compares these different costs at various stages of the previous simulations for three pairs (σ^*, δ). The equilibrium values are also indicated for the instances with positive amplitude $(\delta > 0)$. They were obtained using an equivalent linear program proposed by Iryo and Yoshii (2007), which only applies when u_o is constant and identical for all users. Since runs with different values of λ share the same equilibrium, they were combined in the same plots, with different colors.

The analysis of the third column suggests that the equilibrium approximation is suitable to estimate the congestion cost. The estimation can admittedly be quite far on specific days (e.g. up to $15 - 20\%$ difference for $(\sigma^*, \delta) = (0.4, 1.5)$ and $\mathrm{PG}(x) = 30\%$). Yet, the cloud of points is approximately symmetric compared to the equilibrium line, such that the average congestion cost over many days is close to its equilibrium value.

In terms of cost decomposition however, the equilibrium approximation entails severe biases. It systematically overestimates the average delay and underestimates the average schedule penalty. The discrepancy can be quite large, especially compared with the average schedule penalty at equilibrium. These results were expected for the first two rows. Indeed, they correspond to a special case where the equilibrium is known to minimize the sum of schedule penalties (see e.g. Iryo and Yoshii (2007)). The amplitude of the bias was however unknown and it is interesting to notice that a similar bias exists in the evening commute scenario considered (last row).

These biases have important practical consequences. When estimating the congestion cost for instance, it is common to measure only the average delay and to deduce the average schedule penalty by applying some proportionality ratio deduced from equilibrium analysis. Our results suggest that such a method could greatly underestimate the congestion cost.

3.5 Discussion

As shown by our literature review, stability analyses raise complex theoretical questions. Yet, stability is an important practical issue, which ultimately requires clear answers and quantitative estimates. Faced with this challenge, we proposed a hybrid approach, based on both analytical derivations and simulations. Analytical derivations give this work a broad scope. By combining results from route choice and evolutionary game theory, it is possible to derive sufficient conditions for continuous time stability that are valid for many rational adjustment mechanisms. Yet, these theories only address part of the problem: they ignore discrete-time issues and do not tell us what happens when utility functions are not monotonic. While simulation results are inherently limited by the range of scenarios considered, they can provide orders of magnitude and illustrate important trends. Here, simulations shed light on the role of some behavioral parameters and on the soundness of the equilibrium assumption in different situations. Given the great heterogeneity existing in the real world, the equilibrium assumption may still provide rather good estimates of the congestion cost. The decomposition of the congestion cost may however be quite far from the one observed in practice. Schedule penalties in particular tend to be severely underestimated by the equilibrium assumption.

This work could be extended in many ways. Empirical work is needed, in particular to quantify the daily variations existing in real world departures. On the theoretical side, an important research priority would be to consider other congestion mechanisms that could be applied to model cities. Previous reports of detailed network simulations (e.g. with the METROPOLIS simulator (de Palma et al., 1997)) and of an ideal isotropic zone-based congestion model (Lamotte and Geroliminis, 2017) suggest that multi-directionality may lead to different dynamic properties.

4 The morning commute in urban areas with heterogeneous trip lengths

This chapter is based on the article:
Lamotte, R. and Geroliminis, N., 2017. The morning commute in urban areas with heterogeneous trip lengths. Transportation Research Part B: Methodological (In Press).

The work has been performed by the candidate under the supervision of Prof. Nikolas Geroliminis.

4.1 Introduction

The morning commute has historically been studied with point bottlenecks, following the seminal paper of Vickrey (1969). Urban networks however, cannot be modeled as collections of independent point bottlenecks. In fact, congestion propagates from one bottleneck to its neighbors, creating connected components of congestion that grow and may extend to the whole network (Ji et al., 2014). In the face of this complexity, an attractive solution consists in considering congestion in homogeneous and isotropic areas (Small and Chu, 2003; Geroliminis and Levinson, 2009). This approach relies on empirically supported relationships, referred to as Macroscopic Fundamental Diagrams (MFD), which describe the dynamics of congestion with just a few variables such as accumulation, space-mean flow, trip completion rate and speed (see Section 2.3.3).

The consequences of this change of scale are not fully understood yet. The first works combining an MFD with departure time choice called attention to the cost of hypercongestion, i.e. the phenomenon by which vehicle flow decreases with accumulation when accumulation exceeds a critical level (Small and Chu, 2003). Geroliminis and Levinson (2009) and Fosgerau and Small (2013) argued that by maintaining the system at the flow-maximizing accumulation, the benefits of congestion pricing with homogeneous users could be even greater than with Vickrey's bottleneck, as the duration of the peak hour could be shortened. Arnott (2013) showed that further gains can be obtained by maintaining the accumulation always *below* its flow-maximizing value, at a level that increases with the peak duration. While they reach similar conclusions, the aforementioned papers utilized different assumptions to deal with complex dynamics involving endogenous delays. Small and Chu (2003) and Geroliminis and Levinson (2009) assumed that travel time is entirely determined by the conditions at a single instant (e.g. at the arrival), Fosgerau and Small (2013) considered a piece-wise constant decreasing branch for the MFD of flow versus accumulation and Arnott (2013) assumed that at any time, all users have the same probability of exiting the network, independently of the time at which they started their trip. Yet, it is only very recently that Fosgerau (2015) and Daganzo and Lehe (2015) recognized how trip length heterogeneity challenges the fundamental FIFO assumption and started investigating its impacts on the morning commute.

Surprisingly however, Fosgerau (2015) and Daganzo and Lehe (2015) reached very different conclusions on the role of trip length. Using mathematically convenient but unconventional exponential-type scheduling preferences, Fosgerau (2015) showed that under some assumptions, the user equilibrium exhibits the "regular sorting" property, i.e. two users differing only by their trip length sort according to a Last-In, First-Out (LIFO) pattern where the user with the longest trip starts earlier and finishes later. On the other hand, Daganzo and Lehe (2015) proved for the more conventional $\alpha - \beta - \gamma$ scheduling preferences but with a less realistic congestion mechanism (actually very similar to a point bottleneck) that the social optimum exhibits the FIFO property. Based on these theoretical considerations, Daganzo and Lehe (2015) also proposed a usage-based toll maintaining the accumulation near its flow-maximizing value and demonstrated numerically its benefits on the same realistic congestion mechanism as in Fosgerau (2015). The question of the prevailing sorting pattern with heterogeneous trip lengths remains not fully solved, despite probably severe consequences of congestion management and stability.

This chapter investigates the morning commute problem with inelastic demand and a MFD relating speed to accumulation. Our focus is primarily on the impact of trip length heterogeneity, but we also study the impact of heterogeneity in the scheduling preferences. In line with most of the literature, we rely on the so-called "fluid approximation" (Newell, 1982), in which the stochastic dynamics of a large number of agents are modeled by a deterministic real-valued process. With Vickrey's constant capacity bottleneck, this assumption, together with some convexity assumptions on the schedule penalties, permitted to prove the existence and the uniqueness of an equilibrium distribution of arrival times (Smith, 1984a; Daganzo, 1985). In this work, we leave aside the existence and uniqueness questions as they would require tedious derivations beyond the scope of the thesis (even for a bottleneck model with much simpler dynamics, the proofs in Smith (1984a) and Daganzo (1985) contain a significant number of modeling assumptions, but also very long derivations). Instead, we analytically characterize an hypothetical equilibrium and confirm using simulations that (i) an attracting steady-state exists for a wide range of demands and that (ii) the patterns theoretically predicted are indeed observed.

The complexity of the problem is mainly due to an endogenous delay in the dynamical model (more information is provided in Section 2.3). This makes a fully analytical solution intractable, as also pointed out by Arnott (2013). Nevertheless, in this work we are able to mathematically prove important properties of the equilibrium solution and develop a numerical simulation of the detailed model that confirms some of the proofs and provides further insight. More specifically, this chapter has three main contributions. First, we show that with a continuum of users having continuously distributed characteristics, accumulation and speed are continuous functions of time. This result is valid for a large class of scheduling preferences and contrasts with the discontinuities resulting from the assumption of homogeneous trip lengths (Arnott et al., 2016). Second, we assume $\alpha - \beta - \gamma$ preferences and, under certain conditions, demonstrate that they result in a partial FIFO pattern among early and late users at equilibrium. This FIFO pattern is strict only within families of users having identical $\alpha - \beta - \gamma$ preferences and heterogeneous trip lengths, or vice versa. Yet, simulation results indicate that it influences the overall properties of the equilibrium, and in particular the cost function. It is also shown that with this more detailed dynamic model, the proportion of on-time users may be larger than previously found under simpler approximations and that, under rapid demand variations, the social optimum may exhibit hypercongestion. This finding goes against the previously proposed pricing strategies and calls for further research on congestion pricing under rapidly varying demand.

Sections 4.2.2 and 4.3 provide analytical treatments of the continuity and sorting properties, while Section 4.4 presents the simulation results.

4.2 General properties

4.2.1 Time-space transformation

As highlighted by Eq. (2.6), the phenomena occurring with a speed-MFD may be difficult to analyze in the time domain. Thus, it is often convenient to change the frame of reference and to study the dynamics of congestion over some distance measure. More specifically, we can introduce the bijection $f : \mathbb{R} \to \mathbb{R}$ given by $f(t) = \int_0^t v(s)\,\mathrm{d}s$, which associates to a time t the distance traveled by a virtual user from the origin of time to t.[1] In the same fashion as we use t to denote a particular time, let $x = f(t)$ denote a particular distance traveled and let V be the speed function in the x space: $V(x) = v(f^{-1}(x))$.

We can now define the function $T(t_\mathrm{w}, l)$ that associates to an arrival time at work $t_\mathrm{w} \in \mathbb{R}$ and a trip length $l \in \mathbb{R}^+$ the corresponding travel time. In the time domain, $T(t_\mathrm{w}, l)$ is only implicitly specified by the constraint $\int_{t_\mathrm{w}-T(t_\mathrm{w},l)}^{t_\mathrm{w}} v(s)\,\mathrm{d}s = l$. In the distance domain

[1] Note that f might not be bijective in some out-of-equilibrium situations reaching (or converging towards) gridlock, but we discard these pathological cases to focus on the equilibrium properties.

however, T admits an explicit expression:

$$T(t_{\mathrm{w}}, l) = \int_{f(t_{\mathrm{w}})-l}^{f(t_{\mathrm{w}})} \frac{1}{V(s)} \, \mathrm{d}s. \tag{4.1}$$

Note that f is continuous and since V is positive almost everywhere at equilibrium, T is continuous as well. In addition, if $v(t)$ is continuous, then $V(x)$ is also continuous and T is continuously differentiable. In particular,

$$\frac{\partial T}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}}, l) = \frac{v(t_{\mathrm{h}}) - v(t_{\mathrm{w}})}{v(t_{\mathrm{h}})}, \tag{4.2}$$

where the departure time from home t_{h} is actually a function of t_{w} and l, given by $t_{\mathrm{h}}(t_{\mathrm{w}}, l) = t_{\mathrm{w}} - T(t_{\mathrm{w}}, l)$.

4.2.2 Continuity of accumulation over time

Besides being extremely numerous, commuters are also very diverse. This heterogeneity is often neglected for the sake of tractability and intuition. Yet, this section shows that trip length heterogeneity radically changes the equilibrium patterns when considering a system governed by a speed-MFD. Indeed, Arnott et al. (2016) showed that with homogeneous trip length, there exists an equilibrium which exhibits piece-wise constant accumulation and non-overlapping trip intervals. We demonstrate that when users have continuously distributed characteristics and especially trip lengths, such an equilibrium cannot occur as accumulation should be continuous over time. More specifically, the continuity result demonstrated in this section requires the following assumption.

Assumption 3 (Distributed characteristics)**.** *Assume that:*

- *All users have marginal utilities of time at home and at work* $h(t, \Theta)$ *and* $w(t, \Theta, \mathcal{T}^*)$*, where* Θ *is a vector of individual-specific parameters and* $\mathcal{T}^*$ *is a continuously distributed scalar parameter.*
- *For all* Θ*,* $t \to h(t, \Theta)$ *is positive and continuous everywhere and for all* Θ *and* $\mathcal{T}^*$*,* $t \to w(t, \Theta, \mathcal{T}^*)$ *is continuous everywhere, except possibly at* $t = \mathcal{T}^*$*.*
- *Trip length and trip length conditioned on* $\mathcal{T}^*$ *are continuously distributed variables.*

Although its statement is technical, Assumption 3 is actually quite reasonable. It simply requires continuous marginal utility functions (but still allows for a discontinuity in the marginal utility at work at a single specific time $\mathcal{T}^*$) and that some user characteristics (trip length, $\mathcal{T}^*$ and trip length conditioned on $\mathcal{T}^*$) are continuously distributed. Note that continuous distributions are overwhelmingly common in nature and there are strong reasons to believe that trip length and $\mathcal{T}^*$ are no exception. Indeed, the trip length distribution arises from a very large number of pairs of origins and destinations (primarily home and work locations). Similarly, the distribution of desired exit times from the road network (when the marginal utility is likely to be discontinuous) depends on both the distribution of official work starting times (which is admittedly not continuously distributed) and on the (continuous) distribution of time spans required to reach the workstations from the time the vehicle exits the road network modeled by the MFD (e.g. when entering a parking lot). Note finally that the continuity result obtained is also corroborated by many empirical studies (see for example Geroliminis and Daganzo (2008); Buisson and Ladier (2009)).

To prove continuity of accumulation, we prove that there cannot be mass departures or mass arrivals at equilibrium, or that they have to exactly cancel out each other. This result only makes sense in the context of the fluid approximation as it requires that a single agent does not have a mass, but only a density.[2]

We first need to describe the framework considered and define some basic notions. A *realization* of the morning commute is an allocation of departure times (or equivalently, arrival times) to all users. An *equilibrium* is a realization such that no user can reduce her cost by unilaterally changing her decision. A *socially optimal realization* minimizes the sum of the costs of all users, excluding potential tolls (which are considered pure financial transfers).

Consider a population of mass $N > 0$. In general, an allocation of departures times or arrival times might be such that there are masses of users that depart or arrive simultaneously. If mass departures and mass arrivals do not exactly cancel out, they create discontinuities in the accumulation. For any realization, we can define the set of times T_{dis} for which such discontinuities occur and m_i the (relative) change in accumulation at time $i \in T_{\text{dis}}$. If, for instance, there is a mass departure at time i but no mass arrival, then m_i is the mass of departures at time i. If there is a mass arrival but no mass departure, then the mass of arrivals is given by $-m_i$. If we further define $d_{\text{c}}(t)$ and $a_{\text{c}}(t)$ as the continuous components of the departures and arrivals, the variation in accumulation during some period $[t_1, t_2)$ is given by $\int_{t_1}^{t_2} d_{\text{c}}(t) - a_{\text{c}}(t)\,\mathrm{d}t + \sum_{i \in T_{\text{dis}} \bigcap [t_1,t_2)} m_i$. We now introduce two Lemmas that are useful to prove the continuity of accumulation in Proposition 1.

Lemma 1. *If trip length is continuously distributed, then for any realization of the morning commute, the set of users that have discontinuities both at their departure and arrival has Lebesgue measure zero.*

Proof. Clearly $\sum_{i \in T_{\text{dis}}} |m_i| \leq 2N$, so the set $\{m_i, i \in T_{\text{dis}}\}$ is summable, and consequently countable. As there is a bijection between each element of $\{m_i, i \in T_{\text{dis}}\}$ and T_{dis}, T_{dis} is countable as well, so its Lebesgue measure (in the time domain) is zero.

[2] A discrete version of this result would be along the lines of "if users all have different trip lengths and desired arrival times, then no user departs or leaves exactly at the same time as another one" but it would be more cumbersome to prove, if it can be proven at all.

Now, one can associate to any pair of times (t_1, t_2) in $(T_{\mathrm{dis}})^2$ the distance separating them: $\int_{t_1}^{t_2} v(u)du$. Since T_{dis} is countable, the set of distances separating two events of T_{dis} is also countable. Since trip length is a continuously distributed variable, the set of users having exactly these trip lengths is of measure zero. $\square$

Lemma 2. *Left and right-hand limits of the the accumulation exist everywhere for any realization of the morning commute.*

Proof. The variation in accumulation during some period $[t_1, t_2)$ is given by $\int_{t_1}^{t_2} d_{\mathrm{c}}(t) - a_{\mathrm{c}}(t)\,\mathrm{d}t + \sum_{i \in T_{\mathrm{dis}} \bigcap [t_1, t_2)} m_i$. The first term converges to zero when t_1 tends towards t_2, whether the endpoints t_1 and t_2 are included or not. For the second term however, the distinction matters as there might be a point mass at t_2. If t_2 is excluded, the function $t_1 \mapsto \sum_{i \in T_{\mathrm{dis}} \bigcap [t_1, t_2)} |m_i|$ is weakly decreasing and remains strictly positive for all $t_1 \in (-\infty, t_2)$. Hence, it converges towards its infimum as $t_1 \to t_2^-$. This infimum is equal to zero as all point masses occurring at times strictly smaller than t_2 are excluded progressively as $t_1 \to t_2^-$. As $|\sum_{i \in T_{\mathrm{dis}} \bigcap [t_1, t_2)} m_i| \leq \sum_{i \in T_{\mathrm{dis}} \bigcap [t_1, t_2)} |m_i|$, $|\sum_{i \in T_{\mathrm{dis}} \bigcap [t_1, t_2)} m_i|$ also converges towards zero as $t_1 \to t_2^-$. If we define a sequence of times $(t_i)_{i \in \mathbb{N}}$ that converges from below towards t_2 but remains strictly smaller, $(n(t_i))_{i \in \mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$, hence it converges. The same reasoning shows that $\lim_{u \to t_2^+} n(u)$ exists. $\square$

Proposition 1 (Continuity)**.** *Consider a system governed by a speed-MFD and a population satisfying Assumption 3. If a realization of the morning commute is an equilibrium, then for this realization the accumulation is a continuous function of time.*

Proof. In order to prove the continuity of the accumulation function, we assume the existence of a discontinuity at some time $t_{\mathrm{his}} \in T_{\mathrm{dis}}$ and show that it implies that some user is not at equilibrium. We first assume that the discontinuity is an accumulation decrease, which implies the existence of a mass arrival. Since $\mathcal{T}^*$ is continuously distributed, only a set of measure 0 can have $\mathcal{T}^* = t_{\mathrm{his}}$ (we will say such users are "on time" by analogy with the $\alpha - \beta - \gamma$ preferences, even though this does not necessarily have the same meaning with other types of schedule penalty functions). Thus, most of the users arriving at t_{his} are not on time. Using Lemma 1, we can select a user from the mass that is not on time such that there was no discontinuity at the time of her departure t_{h}. Let t^* and θ denote the values of $\mathcal{T}^*$ and Θ of this specific user.

Lemma 2 guarantees that left and right-hand limits of accumulation exist everywhere. Consequently, the utility function admits left and right derivatives (with respect to the arrival time) everywhere. A local necessary condition for the selected user to be at equilibrium is that $\dot{U}_-(t_{\mathrm{w}}) \geq 0 \geq \dot{U}_+(t_{\mathrm{w}})$ where $\dot{U}_-$ and $\dot{U}_+$ denote the left and right hand derivatives. Using Eq. (2.1) and Eq. (4.2), this is equivalent to

$$\frac{v(t_{\mathrm{his}}^-)}{v(t_{\mathrm{h}})} h(t_{\mathrm{h}}, \theta) - w(t_{\mathrm{his}}, \theta, t^*) \geq 0 \geq \frac{v(t_{\mathrm{his}}^+)}{v(t_{\mathrm{h}})} h(t_{\mathrm{h}}, \theta) - w(t_{\mathrm{his}}, \theta, t^*), \tag{4.3}$$

where $v(t_{\text{his}}^-) = \lim_{u \to t^-} v(u)$ and $v(t_{\text{his}}^+) = \lim_{u \to t^+} v(u)$ denote the left-handed and right-handed limits of the speed v at the discontinuity. Since $h(t_{\text{h}}, \theta) > 0$, Eq. (4.3) is equivalent to $v(t_{\text{his}}^-) \geq v(t_{\text{his}}^+)$, which is inconsistent with a brutal decrease of accumulation. Thus, a discontinuous decrease of accumulation cannot occur at equilibrium under the assumptions made.

The case of a discontinuous accumulation increase is treated in a similar manner by selecting a user that departs at the discontinuity. Among the users departing with this mass, only a zero-measure set can be on-time (because we assumed that both desired arrival time and trip length conditioned on desired arrival time are continuously distributed and only one trip length corresponds to a pair of departure and arrival time) and only a zero-measure set can arrive at another discontinuity (Lemma 1). Thus, we can select a user who departs at t_{his}, is not on time and arrives at a time t_{w} with no discontinuity. The local equilibrium condition (4.3) becomes

$$\frac{v(t_{\text{w}})}{v(t_{\text{his}}^-)} h(t_{\text{his}}, \theta) - w(t_{\text{w}}, \theta, t^*) \geq 0 \geq \frac{v(t_{\text{w}})}{v(t_{\text{his}}^+)} h(t_{\text{his}}, \theta) - w(t_{\text{w}}, \theta, t^*),$$

i.e. $v(t_{\text{his}}^-) \leq v(t_{\text{his}}^+)$, which is inconsistent with a brutal increase of accumulation. □

In the case where only the trip length is distributed and its distribution does not have any point mass, then one can show with a similar proof that the accumulation $n(t)$ is continuous everywhere except at the common desired arrival time. Note also that while Proposition 1 precludes the existence of an equilibrium consisting only of departure and arrival masses, departure and arrival masses might still coexist if they occur simultaneously and exactly cancel out.

4.3 Properties of the equilibrium with $\alpha - \beta - \gamma$ preferences

Depending on the congestion mechanism and on the scheduling preferences used, the first and second-order equilibrium conditions may lead to different sorting properties. For instance, Fosgerau (2015) showed for users with homogenous exponential-type scheduling preferences and a speed-MFD that, provided some endogenous assumptions on the equilibrium, these conditions translate into a LIFO pattern where all trips overlap and shorter trips are "included" within longer trips, i.e. start later and finish earlier. We demonstrate in this section that with $\alpha - \beta - \gamma$ preferences and a speed-MFD, these conditions constrain the speed evolution (see Proposition 2) in such a way that, provided the accumulation over time follows a unimodal curve, a partial FIFO sorting emerges (see Proposition 3). If parameters of the schedule penalty functions vary between individuals, they might also act locally as a sorting criterion, just as trip length (see Proposition 4). An important consequence of the use of $\alpha - \beta - \gamma$ preferences is that sorting applies even among users having different desired arrival times.

4.3.1 First and second order equilibrium conditions with $\alpha - \beta - \gamma$ preferences

Assume some user (t^*, l) with $\alpha - \beta - \gamma$ preferences arrives at equilibrium at time t_{w}. If her schedule penalty function SP is locally continuously differentiable (i.e. $t_{\mathrm{w}} \neq t^*$), the first order equilibrium condition ($\frac{\partial U}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}}, t^*, l) = 0$) translates into

$$\frac{\partial T}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}}, l) = -\frac{1}{\alpha}\frac{\partial \mathrm{SP}}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}}, t^*). \tag{4.4}$$

Similarly, if $T(t_{\mathrm{w}}, l)$ and $\mathrm{SP}(t_{\mathrm{w}}, t^*)$ are twice differentiable, the second order equilibrium condition becomes

$$\frac{\partial^2 T}{\partial t_{\mathrm{w}}^2}(t_{\mathrm{w}}, l) \geq -\frac{1}{\alpha}\frac{\partial^2 \mathrm{SP}}{\partial t_{\mathrm{w}}^2}(t_{\mathrm{w}}, t^*). \tag{4.5}$$

With $\alpha - \beta - \gamma$ preferences, these conditions can be stated as follows:

Proposition 2. *Consider a system governed by a speed-MFD and a population of users satisfying Assumption 3. For a user with $\alpha - \beta - \gamma$ preferences, trip length l and desired arrival time t^*, the equilibrium departure and arrival times t_h and t_w satisfy:*

- *if $t_w < t^*$, then $v(t_w) = \frac{\alpha-\beta}{\alpha} v(t_h)$;*
- *if $t_w > t^*$, then $v(t_w) = \frac{\alpha+\gamma}{\alpha} v(t_h)$;*
- *if $t_w = t^*$, then $\frac{\alpha-\beta}{\alpha} v(t_h) \leq v(t_w) \leq \frac{\alpha+\gamma}{\alpha} v(t_h)$.*

In addition, if $t_w \neq t^*$ and $v(t)$ is differentiable at t_w and t_h:

$$\frac{\dot{v}(t_h)}{v(t_h)^2} \geq \frac{\dot{v}(t_w)}{v(t_w)^2}. \tag{4.6}$$

Proof. For early and late users, the proportionality result between the speeds at departure and at arrival is obtained by combining the first order equilibrium condition (Eq. (4.4)) with Eqs. (2.2) and (4.2). Similarly, Eq. 4.6 is obtained from the second order condition (Eq. (4.5)) by noticing that $\frac{\partial^2 \mathrm{SP}}{\partial t_{\mathrm{w}}^2}(t_{\mathrm{w}}, t^*) = 0$ for all early and late users, by differentiating Eq. (4.2) and reinserting it into Eq. (4.5). Note that if $v(t)$ is differentiable at t_{w} and t_{h}, then the second order differentiability of the travel time at (t_{w}, l) is a direct consequence of its definition in Eq. (4.1). Regarding on-time users, although their utility function is not differentiable at their arrival time, it admits left and right derivatives. At equilibrium $\frac{\partial_- u}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}}, t^*, l) \geq 0$ and $\frac{\partial_+ u}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}}, t^*, l) \leq 0$, i.e.

$$\begin{cases} \frac{\partial T}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}}, l) \leq -\frac{1}{\alpha}\frac{\partial_- \mathrm{SP}}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}}, t^*), \\ \frac{\partial T}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}}, l) \geq -\frac{1}{\alpha}\frac{\partial_+ \mathrm{SP}}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}}, t^*). \end{cases} \tag{4.7}$$

Combine Eq. (4.7) with Eqs. (2.2) and (4.2) to obtain the result for on-time users. □

Let us provide further intuition for the proposition above. The necessary conditions for some user to be at equilibrium stated in Proposition 2 already impose severe constraints on the shape of the equilibrium. Indeed, if we define the function $Q_l(t) = \frac{v(t)}{v(t_{\text{h}}(t,l))}$, the set of candidate arrival times at which a user may be at equilibrium is $Q_l^{-1}\left(\left\{\frac{\alpha-\beta}{\alpha}\right\}\right) \cup Q_l^{-1}\left(\left\{\frac{\alpha+\gamma}{\alpha}\right\}\right) \cup \{t^*\}$, where $Q_l^{-1}(I)$ denotes the (possibly empty) preimage of $I \subset \mathbb{R}$. Thus, provided desired arrival times are sufficiently similar, the exact value of t^* does not influence the equilibrium arrival times of early and late users, and users sort themselves based on their trip length and $\alpha - \beta - \gamma$ coefficients.

This suggests that the $\alpha - \beta - \gamma$ coefficients and trip length play a similar role and essentially sort users. The desired arrival time however seems to influence the equilibrium arrival time in a very distinct manner. Given some $\alpha - \beta - \gamma$ coefficients and trip length, the function that associates to a desired arrival time t^* the corresponding equilibrium arrival time would be the identity function for some range of times such that arriving on time is optimal, and would be piece-wise constant for the complement of this range. In the remainder of this section, we will focus on the effect of the $\alpha - \beta - \gamma$ coefficients and trip length. The reason for this choice is that the effect of t^* is either trivial (when users arrive exactly on time), or requires comparing the costs of candidate equilibrium arrival times, which means that the variations of accumulation must be known over long time intervals. Thus, understanding the effect of the $\alpha - \beta - \gamma$ coefficients and trip length is a preliminary step to understand the effect of t^*.

Note that this is very similar to the situation with a bottleneck model. When desired arrival time plays little role, the dynamics can be solved easily. For instance, with a S-shape cumulative distribution of desired arrival times and homogeneous $\alpha - \beta - \gamma$ coefficients, the dynamics only depend on two particular points of the distribution of t^* (Smith, 1984a; Daganzo, 1985). Yet, when heterogeneity in desired arrival time is combined with heterogeneity in $\alpha - \beta - \gamma$ coefficients or when the desired arrival time does not have a S-shape cumulative distribution, numerical methods are necessary.

4.3.2 From a demand peak to a congestion peak

Various types of assumptions may be used to limit the influence of heterogeneity in desired arrival times. The S-shape assumption of Smith (1984a) and Daganzo (1985) is an excellent example, being a reasonable idealization and still allowing for strong results in the case of a constant capacity bottleneck. So far, we have been unable to adapt this result to dynamics imposed by a speed-MFD, although early simulation results suggest a similar law might exist. Yet, as with a constant capacity bottleneck, our analytical investigations have shown that the exact distribution of desired arrival times is of little significance. What primarily matters is whether the distribution leads to a unimodal evolution of accumulation or to several peaks.

To obtain the sorting results of Section 4.3.3, we chose to consider the existence of a single peak as granted. More specifically, we made the following endogenous assumption.

Assumption 4 (Single peak)**.** *The distribution of users' characteristics is such that at equilibrium: (i) all commuters travel within an interval* $[t_1, t_2]$ *and (ii) accumulation exhibits a single peak, i.e. there exists* $t_p \in (t_1, t_2)$*, such that for all* $t \in [t_1, t_p)$*, accumulation is strictly increasing and for all* $t \in (t_p, t_2]$*, accumulation is strictly decreasing.*

This assumption has the great advantage that it can be tested empirically. In fact, several empirical observations suggest that unimodal evolutions of accumulation are quite common (Parthasarathi et al., 2011; Yildirimoglu et al., 2015).[3] It also greatly simplifies the analysis, as evidenced by the following corollary of Proposition 2.

Corollary 1. *Consider a system governed by a speed-MFD and assume that users' characteristics are such that the accumulation* $n(t)$ *is continuous and exhibits a single peak, reached at time* t_p*, as specified in the Assumptions 3 and 4. Consider a group of users with* $\alpha - \beta - \gamma$ *preferences. At equilibrium, early users depart and arrive before* t_p *while late users depart and arrive after* t_p*.*

Alternatively, it is shown in Appendix A that sorting results similar to those presented in Section 4.3.3 can be obtained with only *exogenous* assumptions on the distribution of user characteristics (with, in particular, the assumption that desired arrival times are restricted to a compact subset of the time space). It requires however some additional precautions and, in the end, the sorting results apply only out of the range where desired arrival times are distributed. In that sense, it is similar to the numerous works assuming homogenous desired arrival times for Vickrey's bottleneck model, but with a more tedious derivation due to the state-dependent speed. We did not include this in the body of the thesis as we favor realism, simplicity and far-reaching results over strict exogeneity of assumptions.

4.3.3 Sorting properties

This section includes two main sorting results. The first one (Proposition 3) builds on the Propositions 1, 2 and on the single peak assumption (Assumption 4) to provide a FIFO sorting result based on trip length that is valid for any pair of users having similar $\alpha - \beta - \gamma$ preferences and arriving both early (or both late). The second result (Proposition 4) is an extension of the first providing some complementary insight about the sorting phenomena emerging among users having *different* scheduling preferences, at the cost of several additional endogenous assumptions.

[3] Smaller peaks with faster variations (e.g. on a time scale of 15 min) are often observed on top of the main peak but they might be considered as less predictable and therefore less important in the choice of departure time.

Proposition 3. *Consider a system governed by a speed-MFD and a population of users satisfying Assumptions 3 and 4. Consider two users having $\alpha - \beta - \gamma$ preferences and arriving early (resp. late) at equilibrium at the times $t_{w,1}$ and $t_{w,2}$. Denote $(l_1, t_1^*, \alpha_1, \beta_1, \gamma_1)$ and $(l_2, t_2^*, \alpha_2, \beta_2, \gamma_2)$ their respective trip lengths, desired arrival times, α, β and γ coefficients. Assume without loss of generality that for instance $t_1^* \geq t_2^*$. If $\frac{\beta_1}{\alpha_1} = \frac{\beta_2}{\alpha_2}$ (resp. $\frac{\gamma_1}{\alpha_1} = \frac{\gamma_2}{\alpha_2}$) and $t_{w,1} \leq t_2^*$ (resp. $t_{w,2} \geq t_1^*$), then the user that arrives the earliest (resp. latest) has the longest trip length.*

Proof. In the case where both users are early (resp. late), assuming $t_{\mathrm{w},1} \leq t_2^*$ (resp. $t_{\mathrm{w},2} \geq t_1^*$) is equivalent to assuming that both users would remain early (resp. late) after switching their arrival times.

Independently of the order of desired arrival times, let us use the index $i \in \{1, 2\}$ for the user that has the shortest trip length and j for the other. In addition to the notations previously defined, let us also define $t_{\mathrm{h},i}$ and $t_{\mathrm{h},j}$ the departure times of users i and j and $t'_{\mathrm{h},i}$, a departure time such that user i arrives exactly at $t'_{\mathrm{w},i} = t_{\mathrm{w},j}$. Since the original situation was an equilibrium, user i cannot change her decision unilaterally and reduce her cost. Therefore, the cost C'_i of departing at $t'_{\mathrm{h},i}$ must be at least as great as the cost C_i of departing at $t_{\mathrm{h},i}$: $C'_i \geq C_i$.

Let us compare the cost for user j of arriving at $t_{\mathrm{w},i}$ (denoted C'_j) and the cost of arriving at $t_{\mathrm{w},j}$ (denoted C_j). In the case both users are early:

$$\begin{aligned} C'_j &= \alpha_j T(t_{\mathrm{w},i}, l_j) + \beta_j (t_j^* - t_{\mathrm{w},i}), \\ &= \alpha_j \left(T(t_{\mathrm{w},i}, l_i) + T(t_{\mathrm{h},i}, l_j - l_i) \right) + \beta_j (t_i^* - t_{\mathrm{w},i}) + \beta_j (t_j^* - t_i^*), \\ &= \frac{\alpha_j}{\alpha_i} C_i + \alpha_j T(t_{\mathrm{h},i}, l_j - l_i) + \beta_j (t_j^* - t_i^*). \end{aligned}$$

Similarly, $C_j = \frac{\alpha_j}{\alpha_i} C'_i + \alpha_j T(t'_{\mathrm{h},i}, l_j - l_i) + \beta_j (t_j^* - t_i^*)$.

Equilibrium requires that $C'_j \geq C_j$, i.e.

$$\alpha_j \Big(T(t_{\mathrm{h},i}, l_j - l_i) - T(t'_{\mathrm{h},i}, l_j - l_i) \Big) \geq \frac{\alpha_j}{\alpha_i} \left(C'_i - C_i \right),$$

and that $C'_i \geq C_i$. Hence,

$$T(t_{\mathrm{h},i}, l_j - l_i) \geq T(t'_{\mathrm{h},i}, l_j - l_i). \tag{4.8}$$

By applying Corollary 1, both users departed and arrived during the onset of congestion, so speed strictly decreases for the entire duration considered (Assumption 4). Thus, Eq. (4.8) imposes that $t_{\mathrm{h},i} \geq t'_{\mathrm{h},i}$, and therefore $t_{\mathrm{w},i} > t_{\mathrm{w},j}$.

If both users are late at equilibrium, two cases should be distinguished, depending on whether users i and j can exchange their arrival times and still depart both during the offset of congestion. If they cannot (user j would have to depart during the onset), then $t_{\mathrm{w},i} < t_{\mathrm{w},j}$. If they can, then β_i and β_j should simply be replaced by $-\gamma_{\mathrm{i}}$ and $-\gamma_{\mathrm{j}}$ and the equations above remain valid, leading to Eq. (4.8). Since the period considered is after the peak, speed strictly increases during the entire period so Eq. (4.8) imposes that $t_{\mathrm{h},i} < t'_{\mathrm{h},i}$ (i.e. $t_{\mathrm{w},i} < t_{\mathrm{w},j}$). □

The ordering of arrivals demonstrated trivially implies the ordering of departures for early users (if $l_i < l_j$ and $t_{\mathrm{w},i} > t_{\mathrm{w},j}$, then $t_{\mathrm{h},i} > t_{\mathrm{h},j}$). For late arrivals, this is not true in general but it is true with the assumptions made as (i) speed is monotonously increasing during the offset of congestion and (ii) Proposition 2 imposes that if speed if greater at $t_{\mathrm{w},j}$ than at $t_{\mathrm{w},i}$, it is also greater at $t_{\mathrm{h},j}$ than at $t_{\mathrm{h},i}$. This leads to the following corollary.

Corollary 2. *Consider a system governed by a speed-MFD and a population of users satisfying Assumptions 3 and 4. At equilibrium, departures and arrivals among a family of early (resp. late) users having the same ratio β/α (resp. γ/α) follow a First-In-First-Out order.*

This is in sharp contrast with the LIFO pattern found by Fosgerau (2015). The difference stems from the types of scheduling preferences considered. In Fosgerau (2015), the marginal utilities at home and at work decrease and increase exponentially with time, such that users just try to be at the good place at the good time, i.e. travel around the time when both utility curves intersect. Note that this is valid independently of the level of congestion, so the system would unavoidably reach gridlock if there is a number of users having identical scheduling preferences that is greater than the jam accumulation. Here, the marginal utilities of time are respectively constant and piece-wise constant so earliness and lateness is more acceptable. Since travel time is relatively more important for users with long trip lengths, they naturally avoid the peak.

Proposition 3 is however limited to early users having identical ratios $\frac{\beta}{\alpha}$ or late users having identical ratios $\frac{\gamma}{\alpha}$. Since the coefficients α, β and γ are actually continuously distributed in the real world, the scope of this proposition might appear limited. However, one can actually extend Proposition 3 to cover such cases by applying the implicit function theorem (see Proposition 4).

For any early or late user, let r denote the user's active ratio of scheduling preferences, i.e. $r = \frac{\beta}{\alpha}$ for early users and $r = \frac{\gamma}{\alpha}$ for late users. Let also σ denote the sign of $\frac{\partial \mathrm{SP}}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}}, t^*)$ (+1 for late users and -1 for early users).

Proposition 4. *Consider a user with trip length $\tilde{l}$, desired arrival time $\tilde{t^*}$ and $\alpha - \beta - \gamma$ preferences that arrives early or late at equilibrium, at time $\tilde{t_w}$. Denote $\tilde{r}$ her active ratio of scheduling preferences.*

If (i) speed is continuously differentiable at $t_h(\tilde{t_w}, \tilde{l})$ and $\tilde{t_w}$, (ii) $\tilde{t_w}$ is the only arrival time that minimizes the user's cost (i.e. the global optimum is unique) and (iii) $\frac{\partial^2 T}{\partial t_w^2}(\tilde{t_w}, \tilde{l}) > 0$;

then there exists an open set U around $(\tilde{l},\tilde{t^*},\tilde{r})$ such that there exists a unique function $t_w(l,t^*,r)$ on U such that a user with trip length l, desired arrival time t^* and active ratio of scheduling preferences r in U would be at equilibrium by arriving at $t_w(l,t^*,r)$.

The function $t_w(l,t^*,r)$ is continuously differentiable on U and its derivatives are given by:

$$\frac{\partial t_w}{\partial l}(l,t^*,r)=\frac{v(t_w)\dot{v}(t_h)}{v(t_w)^2\dot{v}(t_h)-v(t_h)^2\dot{v}(t_w)}\ ; \tag{4.9a}$$

$$\frac{\partial t_w}{\partial r}(l,t^*,r)=-\sigma\frac{v(t_h)^3}{v(t_w)^2\dot{v}(t_h)-v(t_h)^2\dot{v}(t_w)}\ ; \tag{4.9b}$$

$$\frac{\partial t_w}{\partial t^*}(l,t^*,r)=0\ ; \tag{4.9c}$$

where t_h and t_w are functions of l, t^* and r.

Proof. Since the user considered does not arrive on time at equilibrium and speed is continuously differentiable at $t_{\mathrm{h}}(\tilde{t_{\mathrm{w}}},\tilde{l})$ and $\tilde{t_{\mathrm{w}}}$, the first order equilibrium condition is $\frac{\partial T}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}},l)+\sigma r=0$. Using Eq. (4.2), this is equivalent to $g(t_{\mathrm{w}},t^*,l,r)=0$, where $g(t_{\mathrm{w}},t^*,l,r)=\frac{v(t_{\mathrm{h}}(t_{\mathrm{w}},l))-v(t_{\mathrm{w}})}{v(t_{\mathrm{h}}(t_{\mathrm{w}},l))}+\sigma r$.

Since v is continuously differentiable at $t_{\mathrm{h}}(\tilde{t_{\mathrm{w}}},\tilde{l})$ and $\tilde{t_{\mathrm{w}}}$, $t_{\mathrm{h}}(t_{\mathrm{w}},l)=t_{\mathrm{w}}-T(t_{\mathrm{w}},l)$ is also continuously differentiable (see Eq. (4.2)). Thus g is continuously differentiable. In addition, $\frac{\partial g}{\partial t_{\mathrm{w}}}(t_{\mathrm{w}},l)=\frac{\partial^2 T}{\partial t_{\mathrm{w}}^2}(t_{\mathrm{w}},l)>0$. Hence, we can apply the implicit function theorem to obtain the existence of a unique function $t_{\mathrm{w}}(l,t^*,r)$ on a open neighborhood around $(\tilde{l},\tilde{t^*},\tilde{r})$ such that a user characterized by (l,t^*,r) would be at equilibrium by arriving at $t_{\mathrm{w}}(l,t^*,r)$. Furthermore, $\frac{\partial g}{\partial l}=\frac{v(t_{\mathrm{w}})\dot{v}(t_{\mathrm{h}})\frac{\partial t_{\mathrm{h}}}{\partial l}}{v(t_{\mathrm{h}})^2}=-\frac{v(t_{\mathrm{w}})\dot{v}(t_{\mathrm{h}})}{v(t_{\mathrm{h}})^3}$; $\frac{\partial g}{\partial t_{\mathrm{w}}}=\frac{v(t_{\mathrm{w}})^2\dot{v}(t_{\mathrm{h}})-v(t_{\mathrm{h}})^2\dot{v}(t_{\mathrm{w}})}{v(t_{\mathrm{h}})^3}$; $\frac{\partial g}{\partial t^*}=0$; and $\frac{\partial g}{\partial r}=\sigma$. The implicit function theorem then provides the expressions of the partial derivatives of $t_{\mathrm{w}}(l,t^*,r)$ given in the proposition. $\square$

The expressions of the three partial derivatives in Eq. (4.9) provide three local sorting results. If there is a single peak, $\dot{v}(t_{\mathrm{h}})$ is negative (resp. positive) for early (late) users. In addition, the condition (iii) is equivalent to $v(t_{\mathrm{w}})^2\dot{v}(t_{\mathrm{h}})-v(t_{\mathrm{h}})^2\dot{v}(t_{\mathrm{w}})>0$. Hence the equilibrium arrival time locally decreases (increases) with the trip length, in agreement with Proposition 3. Similarly, $\frac{\partial t_{\mathrm{w}}}{\partial r}(l,t^*,r)>0$ so the equilibrium arrival time increases (resp. decreases) with the ratio $\frac{\beta}{\alpha}$ (resp. $\frac{\gamma}{\alpha}$), i.e. when schedule penalties are relatively more important. Finally, the fact that $\frac{\partial t_{\mathrm{w}}}{\partial t^*}(l,t^*,r)=0$ confirms that among early users, the desired arrival time plays no sorting role. Note that Newell (1987) obtained the same last two sorting properties (without using the gradient of t_{w}) for Vickrey's bottleneck. In fact, these qualitative properties result directly from the first order equilibrium condition, and only the quantitative properties of the gradient depend on the congestion mechanism.

Finally, let us discuss the assumptions used in Proposition 4. The differentiability assumption in Condition (i) is stronger than the continuity result shown in Proposition 1. Yet, we believe that in the same way we proved the continuity of the accumulation requiring only that the distributions of some characteristics of the population do not include any point-mass, the differentiability of the accumulation could be obtained by requiring only the continuity of the probability density functions of the trip length, of the desired arrival time and of the schedule penalty parameters. Although we do not prove this result, the intuition behind it is the following. Continuously distributed desired arrival times ensure the continuity of on-time departure and arrival rates. Continuously distributed trip lengths ensure the continuity of early and late departure and arrival rates. Then, continuously distributed schedule penalty parameters ensure that the transition between e.g. early and on-time departure does not occur simultaneously for many types of users. Condition (ii) simply discards some borderline cases and Condition (iii) is the second order sufficient optimality condition. Note that assuming Condition (i) and that the user is early or late at equilibrium already ensures that $\frac{\partial^2 T}{\partial t_\mathrm{w}^2}(\tilde{t_\mathrm{w}}, \tilde{l}) \geq 0$. However, a strict inequality is necessary to ensure that the conditions are also satisfied in an open neighborhood around the user considered.

4.4 Simulation-based analysis

Simulation-based and analytical approaches are complementary. While analytical considerations provide useful understanding about the key variables involved in congestion mechanisms, simulation can test more complicated configurations and provide clues as to whether systems with higher degrees of freedom and uncertainties can be described with little error by elegant analytical models. Indeed, most of the results proven in the previous sections require some assumptions that are simplifications of the real world (e.g. strict $\alpha - \beta - \gamma$ preferences or any homogeneity assumption). Simulation allows for the relaxation of these assumptions. Furthermore, the stability and the "power of attraction" of equilibria is just as important as their exact characterization. In fact, only a stable and attractive equilibrium is likely to be similar to the ever-changing real world traffic conditions. While these characteristics are usually extremely difficult to identify analytically, simulation inherently addresses these issues as it mimics the evolution of demand from day to day.

4.4.1 Description of the simulation

The simulation-based method proposed mimics the day-to-day adaptation of real drivers to changing conditions in order to identify a potential equilibrium. The basic idea consists in iteratively 1) updating the departure time decisions and 2) running an event-based and agent-based simulation based on Eq. (2.5) to determine the evolution of congestion over time and the cost of each user. In order to provide some stability, the commonly used (Peeta and Mahmassani, 1995; de Palma and Marchal, 2002) method of successive averages (MSA) is applied on the departure time decisions. At each iteration, the departure time decisions of a fraction of the population (e.g. 5% or $\frac{100}{n}$%, where n is the iteration number) were updated based on the results of the previous simulation. Alternatively, MSA could have been applied on the travel times by updating the decisions of all the agents but using average travel times instead of those of the last iteration. The algorithm used is described in more details in Lamotte and Geroliminis (2016).

A quadratic speed-MFD such that $\mathcal{V}(n) = v_{\mathrm{f}}\left(1 - \frac{n}{n_{\mathrm{jam}}}\right)^2$ if $n \in [0, n_{\mathrm{jam}}]$ and $\mathcal{V}(n) = 0$ otherwise is used for all numerical applications (v_{f} denotes the free-flow speed). Note that this MFD has a critical accumulation $n_{\mathrm{cr}} = n_{\mathrm{jam}}/3$, that $\mathcal{V}(n_{\mathrm{cr}}) = 4/9 v_{\mathrm{f}}$ and that the maximum production is $(4/27)n_{\mathrm{jam}}v_{\mathrm{f}} = (4/9)n_{\mathrm{cr}}v_{\mathrm{f}}$. To decrease the degree of freedom, the free-flow speed was assumed to be 1 and the trip length to be uniformly distributed between 0 and 3 in all numerical applications.

Ten types of $\alpha-\beta-\gamma$ preferences were created with $\alpha = 1$, $\beta = 0.4 + \frac{0.2k}{9}$, and $\gamma = 1.5 + \frac{k}{9}$, where $k \in \{0, 1, \ldots, 9\}$ indicates the family. For each family, 400 agents were created with desired arrival time $(i/400 - 1/2)h$, where $i \in \{1, 2, ...400\}$ and h represents the length of the interval with desired arrival times. Hence, the total population is represented by 4000 agents. Various demands are simulated by varying the mass of each agent (or equivalently, the value of the jam accumulation) but keeping the number of agents (i.e. the resolution of the simulation) constant. A value of $h = 5$ was used for most results except when the impact of h is explicitly studied. For each scenario, 2000 iterations were carried out to approximate the equilibrium.[4]

4.4.2 Validation of the theoretical findings and relaxation of some assumptions

The analytical approach presented in the previous sections relied on a number of assumptions to characterize an hypothetical user equilibrium. This section provides some insight on the validity and scope of these results using the agent-based simulation described in Section 4.4.1.

Figures 4.1a and b are the main tools for this investigation, providing respectively aggregated and disaggregated description of the dynamics. Fig. 4.1a represents the natural logarithm of the speed versus $x = f(t)$. Recall that $f(t)$ represents the distance traveled at time t by a virtual user that enters the network with the first user and never exits (see Section 2.3). Since both f and the natural logarithm are bijections, this graph is similar to the evolution of speed over time, but in a different coordinate space. One can readily observe that this is a single peak case. Fig. 4.1b represents each agent by a single point with coordinates the value of $x = f(t)$ at the agent's departure (x_{h}) and

[4] Fewer iterations would have often been sufficient but cases with high congestion levels or little heterogeneity typically require a large number of iterations to converge.

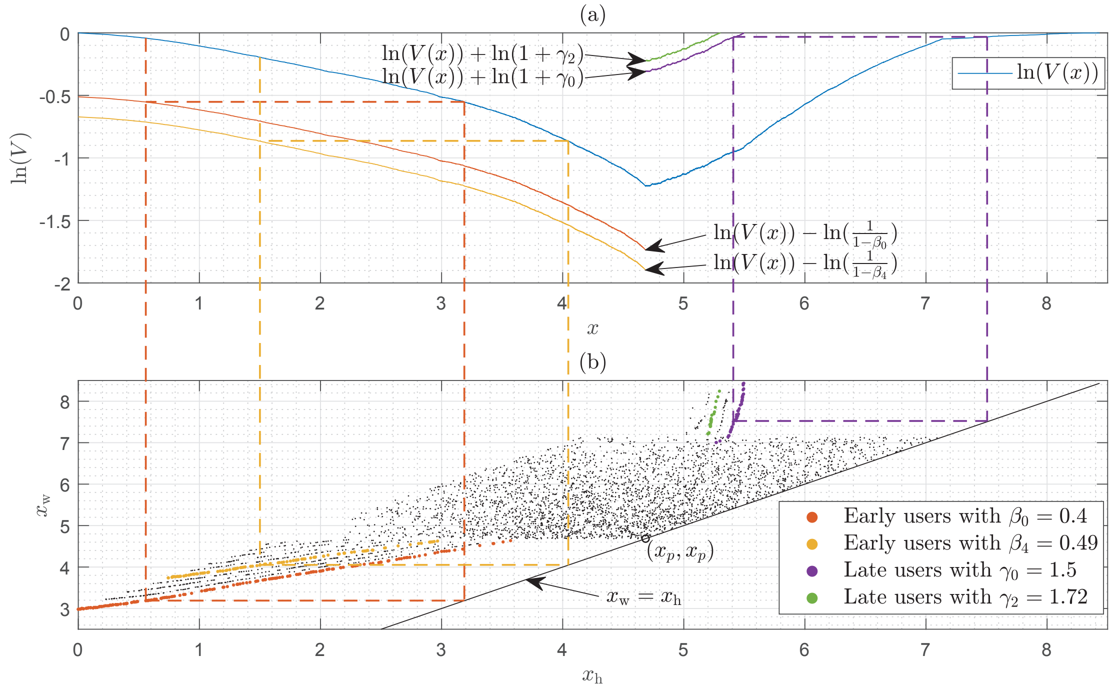

Figure 4.1 – (a) Evolution of the natural logarithm of speed ($\ln(V)$) as a function of x, the distance traveled by a virtual user. (b) Representation of all users in the space (x_h, x_w), illustrating the value of x at the arrival versus the value of x at the departure. All users had $\alpha - \beta - \gamma$ preferences as described in Section 4.4.1, the demand N/n_{cr} was set to 3.

arrival (x_w). Note that since $x_w - x_h$ represents the trip length, all the points are above the line $x_w = x_h$.

Let us define $x_p = f(t_p)$ the value of x that minimizes $\ln(V(x))$. In other words, speed reaches its minimum at t_p. First, note that all the colored points corresponding to early users in Fig. 4.1b are located on the left and below the point (x_p, x_p). This is actually the case for all early users (which form lines parallel to those colored), so all early users depart and arrive during the onset of congestion. Similarly, all colored points corresponding to late users are on the right and above the point (x_p, x_p), i.e. late users depart and arrive during the offset of congestion, as per Corollary 1.

The sorting properties can also be observed in Fig. 4.1b. For each user, the trip length may be read as $x_w - x_h$, i.e. the vertical distance between the point representing the user and the line $x_w = x_h$. Fig. 4.1b shows clearly that within the sets of early users having identical coefficients $\alpha - \beta - \gamma$, those with long trips are on the left and below those with shorter trips, meaning that that they depart and arrive earlier, as per Proposition 3. Similarly, late users with the same coefficient γ sort such that those with short trips depart and arrive earlier than those with longer trips. It is also clear that early users with

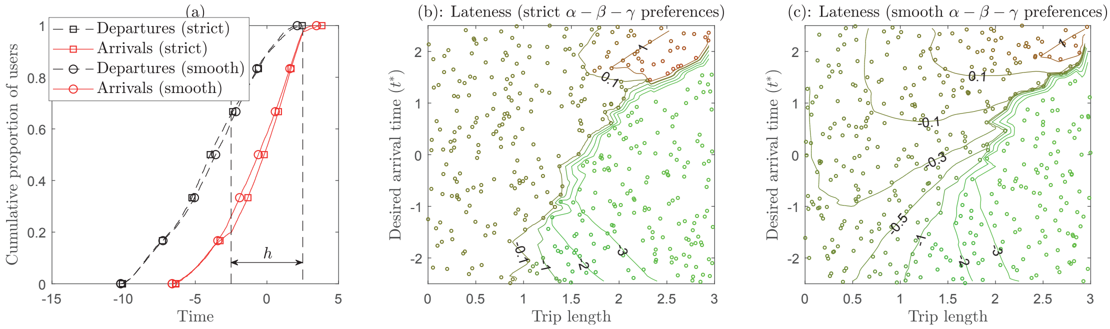

Figure 4.2 – (a) Cumulative departure and arrival curves with strict $\alpha-\beta-\gamma$ preferences and the smooth approximation $h(t)=\alpha$ and $w(t)=\frac{2+\gamma-\beta}{2}+\arctan(4(t-t^*))\frac{\gamma+\beta}{\pi}$. (b) and (c): Contour plots representing the amount of lateness for all 400 users having $\beta=0.4$ and $\gamma=1.5$ with strict $\alpha-\beta-\gamma$ scheduling preferences (b) and their smooth approximation (c). The results were obtained with a demand $N/n_{\mathrm{cr}}=3$.

$\beta=0.49$ depart and arrive later than those having approximately the same trip length but with $\beta=0.4$. In between, the points are located along three lines, corresponding respectively to the families $\beta=0.4+\frac{0.2k}{9}$ with $k=1$, 2 and 3. The same applies to late users, as per Proposition 4.

Finally, the joint analysis of Figs. 4.1a and b illustrates the first and second order equilibrium conditions studied in Section 4.3.1. Indeed, in the coordinate system $(x,\ln(V))$, the first order condition (4.4) simply becomes $\ln(V(x_{\mathrm{w}}))=\ln(V(x_{\mathrm{h}}))-\ln(\frac{\alpha}{\alpha-\beta})$ for early users and $\ln(V(x_{\mathrm{w}}))=\ln(V(x_{\mathrm{h}}))+\ln(\frac{\alpha+\gamma}{\alpha})$ for late users. These conditions are illustrated for two early users and one late user by the dashed lines of different colors joining the Figs. 4.1a and b. For any user $(x^0_{\mathrm{h}},x^0_{\mathrm{w}})$ materialized in Fig. 4.1b, $\ln(V(x^0_{\mathrm{h}}))$ can be read at the intersection of the vertical line $x_{\mathrm{h}}=x^0_{\mathrm{h}}$ with the curve $\ln(V)$ (in Fig. 4.1a). Similarly, $\ln(V(x^0_{\mathrm{w}}))$ may be read by starting from the point materializing the user, drawing an horizontal line $x_{\mathrm{w}}=x^0_{\mathrm{w}}$ until the line $x_{\mathrm{w}}=x_{\mathrm{h}}$ in Fig. 4.1b and then a vertical line $x_{\mathrm{h}}=x^0_{\mathrm{w}}$ until the curve $\ln(V(x))$ in Fig. 4.1a. One can then check that $\ln(V(x^0_{\mathrm{w}}))$ is at the same vertical level as the curve $\ln(V(x))-\ln(\frac{1}{1-\beta})$ at x^0_{h} for early users, or at the same level as $\ln(V(x))+\ln(1+\gamma)$ for late users. Besides, the second order condition (4.5) for early or late users becomes $\frac{\mathrm{d}}{\mathrm{d}x}\ln(V(x))\Big|_{x=x_{\mathrm{h}}}\geq\frac{\mathrm{d}}{\mathrm{d}x}\ln(V(x))\Big|_{x=x_{\mathrm{w}}}$. One can easily verify that the slope of $\ln(V(x))$ systematically decreases between the departure and the arrival of early and late users.

Last, we investigate whether the phenomena obtained for the $\alpha-\beta-\gamma$ can be expected to hold with more realistic scheduling preferences that do not exhibit a discontinuity in t^*. The simulation above was replicated with a smooth approximation of the $\alpha-\beta-\gamma$ preferences, modeled by $h(t)=\alpha$ and $w(t)=\frac{2+\gamma-\beta}{2}+\arctan(4(t-t^*))\frac{\gamma+\beta}{\pi}$. The resulting cumulative departure and arrival curves are displayed in Fig. 4.2a, together with those obtained with strict $\alpha-\beta-\gamma$ preferences. While departures and arrivals occur almost exactly over the same periods, the evolution of the cumulative arrival curve is smoother with the smooth preferences than with the $\alpha - \beta - \gamma$ preferences. The peak is also less pronounced and the maximum accumulation (i.e. the vertical distance between the two curves) smaller.

Despite these minor differences in the dynamics, the analysis of the individual decisions suggests the existence of similar patterns. Fig. 4.2b and c illustrate the amount of lateness (a negative amount corresponds to earliness) of all the 400 users characterized by $\beta = 0.4$ and $\gamma = 1.5$ with both types of preferences. With strict $\alpha - \beta - \gamma$ preferences, only users with long trips choose not to arrive on time. They reschedule earlier or later depending on their desired arrival time t^*. With smooth preferences, the same observation is valid for users with long trips. In this case however, short trips are also rescheduled in such a way that arrival times remain close to desired arrival times. This is caused by the slow variations in utility around t^*: with smooth scheduling preferences, there exists a range of arrival times which are almost equivalently desirable. Thus, in both cases, users with short trips are approximately on time, while those with long trips are early or late, depending on their desired arrival time. Overall, $\alpha - \beta - \gamma$ preferences produce results that are remarkably similar to those associated with a smooth approximation, thus justifying the use of the $\alpha - \beta - \gamma$ preferences for analytical derivations.

4.4.3 Impact of demand on the user equilibrium

Demand severity is described throughout this chapter in terms of total demand (N/n_{cr}) and duration over which desired arrival times are distributed (h). Fig. 4.3a and 4.3b illustrate the time series of accumulation and speed for different values of N/n_{cr}. As expected, the peak duration and the maximum congestion level both increase with the total demand (keeping h constant). Note that for the four scenarios considered, the maximum congestion level is reached for $t = -\frac{h}{2}$, i.e. at the earliest desired arrival time. With Vickrey's bottleneck model and homogeneous $\alpha - \beta - \gamma$ preferences, the maximum would be reached later, for $t = \frac{\gamma-\beta}{\beta+\gamma}\frac{h}{2}$. Intuitively, as users have to travel some distance in the congested area, they are necessarily in the network before their desired arrival time, so the accumulation peak occurs earlier than with a constant capacity bottleneck.

As shown by Fig. 4.3c, the detailed dynamic model used allows for a very large share of users arriving during the interval $[-h/2, h/2]$, even though large accumulations may be observed earlier. To explain this phenomenon, we analyze the trajectory of the outflow in the accumulation-outflow space (see Fig. 4.3f). The common steady-state (and static average trip length L) approximation assumes that this trajectory lays along a well-defined curve $\mathcal{O}(n) = \mathcal{P}(n)/L$ (represented by a black dashed line in Fig. 4.3f). Here, the observed outflow is very far from this idealized model and during the interval $[-h/2, h/2]$, it is about twice as large as the outflow-MFD capacity. Two reasons can explain this discrepancy. First, the average trip length of users in the network is time-dependent, as

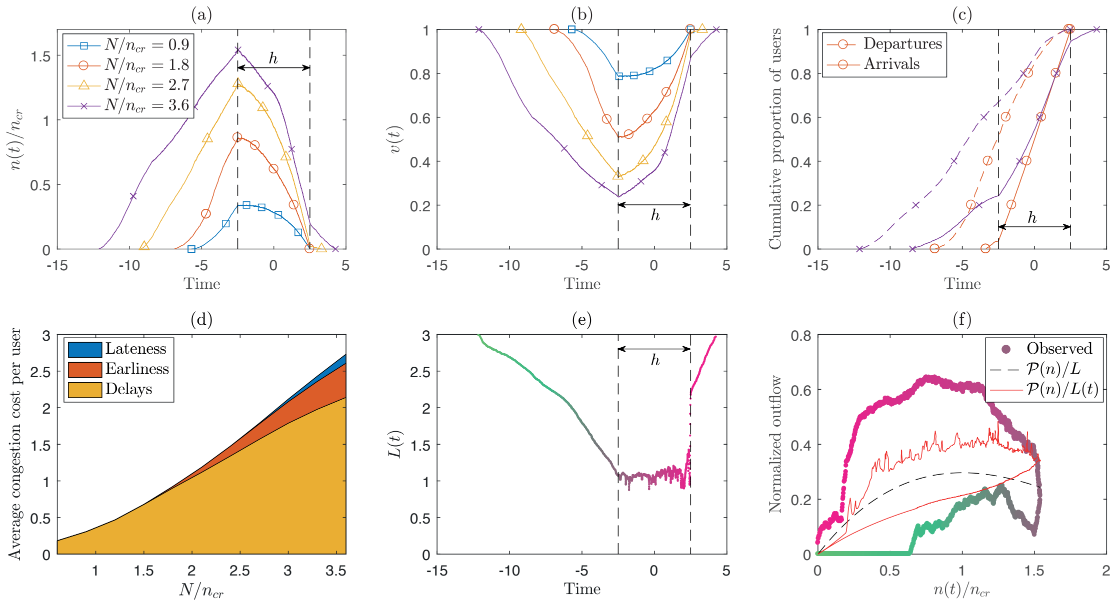

Figure 4.3 – User equilibrium comparison for different demands of the (a) accumulation over time ; (b) speed over time ; (c) cumulative departure and arrival curves ; (d) decomposition of the average congestion cost per user into its components associated to delays, earliness and lateness. The sub-figures (e) and (f) represent the harmonic average of trip length of all users in the network as a function of time (smoothed with a moving average) and the observed outflow (also smoothed with a moving average) in the accumulation-outflow space, for the scenario $N/n_{\text{cr}} = 3.6$. The observed outflow is compared in (f) with $P(t)/L$ (where L=3/2) and with $P(t)/L(t)$, where $L(t)$ is the harmonic average trip length represented in (e). Desired arrival times are distributed on an interval of length $h = 5$.

shown in Fig. 4.3e.[5] By replacing the constant trip length of the outflow model by the time-dependent average trip length $L(t)$, an anti-clockwise hysteresis loop is obtained, represented by the red curve in Fig. 4.3f.[6] Yet, this hysteresis loop remains smaller than the one observed. The rest of the discrepancy might be explained by the complex endogenous delay in Eq. (2.6). Indeed, users that are about to finish their trip are under-represented in the network during the onset of congestion but over-represented during the offset. Thus, even if all the generated trips had the same length, an hysteresis loop would still be obtained. Note that the hysteresis phenomenon obtained here is extremely severe but that it would be reduced with slower time variations.

This high outflow during the most desired period allows for very little schedule penalties overall, as shown in Fig. 3d. In contrast, schedule penalties account for half the congestion cost with a bottleneck of constant capacity and homogeneous users. Note also that the individual congestion cost increases approximately linearly with the demand, at least for the range $N/n_{\mathrm{cr}} \in [1.8, 3.6]$. This is similar to Vickrey's bottleneck problem (Arnott et al., 1990) and to the results of Arnott (2013) but in sharp contrast with those of Fosgerau (2015), in which a LIFO sorting means that a population larger than the jam accumulation of users having homogeneous scheduling preferences cannot be served at a finite cost. This result is related to the sorting pattern and is expected to hold whenever the system converges towards an equilibrium.

4.4.4 Impact of staggered work hours

Staggered work hours may be used instead or in addition to congestion pricing to flatten the peak and reduce schedule penalties (Henderson, 1981). With homogeneous $\alpha - \beta - \gamma$ preferences and a S-shape distribution of desired arrival times, Vickrey's bottleneck predicts a schedule penalty reduction but no impact on the dynamics as long as all desired arrival times remain within the congested period (Vickrey, 1969).

With a speed-MFD however, Fig. 4.4 shows that staggered work hours do impact the dynamics, in a way that depends on the total demand N/n_{cr}. Fig. 4.4a compares the average congestion cost for different demands and ranges h of desired arrival times. Fig. 4.4b and Fig. 4.4c provide similar comparisons for the time series of accumulation (only for $N/n_{\mathrm{cr}} = 1.8$ and 3) and the proportions of early and late users (only for $h = 2$ and 5). $h = 2$ may correspond to an initial state while $h = 5$ may correspond to a state after staggering work hours. Fig. 4.4b suggests that the impact of h on the dynamics of accumulation is larger when the demand is lower, i.e. when the range of desired arrival times is of the same order of magnitude as the peak duration. When the peak period is significantly longer than the range of desired arrival times, staggering work hours

[5] The harmonic average considered in Fig. 4.3e would be equal to the arithmetic average trip length of the trip-generating process if the system was in a steady state.

[6] The noise in the upper part of the loop is caused by users having trip lengths almost equal to 0 and traveling during the period $[-h/2, h/2]$.

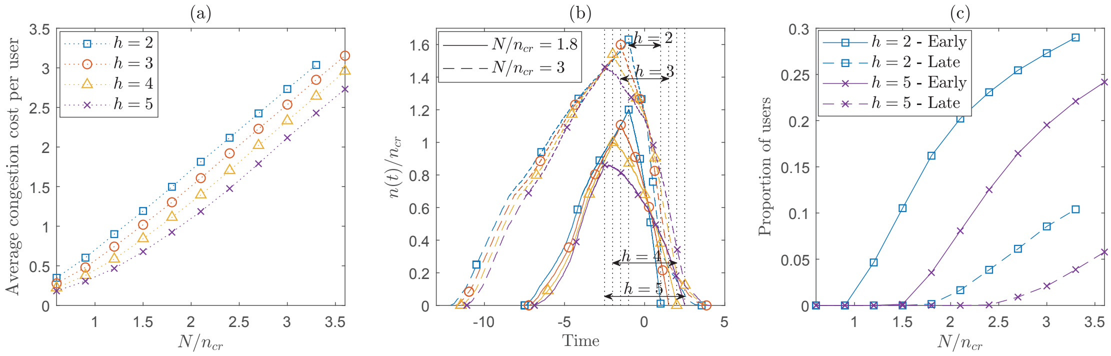

Figure 4.4 – (a) Comparison for $h = 2$, 3, 4 and 5 of the average individual congestion cost as a function of the demand N/n_{cr} ; (b) time series of accumulation for a demand of $N/n_{\mathrm{cr}} = 1.8$ and $N/n_{\mathrm{cr}} = 3$; (c) proportions of early and late users as functions of the demand for desired arrival times distributed on intervals of length $h = 2$ and $h = 5$.

only slightly reduces the maximum congestion level. Yet, it still allows for significant cost reductions, as evidenced by Fig. 4.4a. These cost reductions are explained by the reduced congestion level but also by a large reduction of the proportions of early and late users (see Fig. 4.4c). Note that we only focused here on the impact of staggering work hours on the congestion cost. In reality, staggering work hours also entails other costs. For instance, for many types of professions, the productivity of a worker at a given time depends on whether others also work at that time (Henderson, 1981).

4.5 Discussion

This chapter has investigated the equilibrium properties and appropriate congestion pricing strategies for the morning commute problem with a regional bottleneck modeled by a speed-MFD. Our investigations have shed light on the fundamental role of trip length heterogeneity. We first showed that with the fluid approximation, a continuously distributed trip length ensures that the main variables of the morning commute (accumulation, speed) are continuous functions of time for a large class of scheduling preferences. This essential property served then as the basis for several analytical derivations. We showed in particular that if the users' characteristics are such that the morning commute consists of a single peak, a FIFO sorting pattern naturally emerges within families of early and late users having identical $\alpha - \beta - \gamma$ preferences and heterogeneous trip lengths, or vice versa. Simulations suggest that these sorting properties influence the global characteristics of the equilibrium and produce patterns that are also observed with similar but smooth utility functions.

Yet, many questions remain open. The comparison of the present work with Fosgerau (2015) underlines the importance of empirical measurements. Different estimations of scheduling preferences reaching different conclusions have highlighted the complexity of an apparently simple choice. While this would not fully characterize the scheduling preferences, the analysis of a large number of observations of real world trips may answer the question of the prevailing pattern (FIFO or LIFO). On a more theoretical note, an important issue is the determination of necessary and sufficient conditions for the emergence of a single peak. Early results towards this direction are detailed in Appendix A and deserve to be extended. This might also be completed by empirical works investigating whether commuters schedule their trip mostly based on a smoothed profile of the dynamics (that would exhibit a single peak in most cases) or whether they also account for small oscillations.

The consequences of rapidly varying demand on congestion pricing should also be investigated. Different tolls have been proposed that maintain the accumulation below its critical value, either at a constant (Daganzo and Lehe, 2015) or at a time-varying level (Arnott, 2013). These tolls were shown to be optimal under some simplifying assumptions and are expected to provide significant benefits as long as demand does not vary too rapidly. Under more rapid variations however, such pricing strategies might prove inefficient and a complete change of paradigm may be needed. Consider for instance a system governed by a triangular production-MFD, such that the production-maximizing accumulation is N. Assume that there are exactly N long-distance travelers that travel (at free-flow speed) from 7 am to 8 am. Now assume that there is an additional short-distance traveler, who would like to arrive at 7:30 am. If the short trip length is short enough, it is socially optimal that all users arrive on-time, even if it creates transient hypercongestion.[7] Note that the fact that demand results from scheduling preferences is of little importance. Indeed, even if we considered departures times as given, it would not be optimal to prevent the $N+1^{th}$ user from entering the network until 8 am. Such a measure would impose 30 min delay to this user while her impact on the others can be arbitrarily small (with arbitrarily small trip lengths). Thus, when the dynamics are governed by a speed-MFD, the total distance traveled in the next time unit by all vehicles is not the only performance criterion: its distribution among users matters as well. In some situations, finishing as many trips as possible in the near future might be a more appropriate myopic objective than keeping the system in a state that maximizes the number of vehicle kilometers traveled (i.e. the production). This is also supported by recent control applications in networks governed by multi-region MFD dynamics (Ramezani et al., 2015). Estimation of the trip completion rate is difficult with traditional measurement methods (e.g. loop detectors) but will become more practical with the increasing use of trajectory data.

[7]To prove this, first note that the situation proposed is a local optimum. Indeed, small changes of departure times only add schedule penalties and do not reduce travel time. Second, adding one on-time user with a very short trip length is equivalent to keeping the number of users constant but continuously switching the trip length of the $N+1^{th}$ user from zero to some non-zero value. By continuity of the social cost function with respect to the $N+1$ pairs of arrival times and trip length (gridlock is not possible here), as there is only one global optimum when the trip length of the $N+1^{th}$ user is zero, the global optimum changes continuously in a small neighborhood around this situation. Thus, if the short trip is short enough, the situation proposed is the global optimum.

5 Splitting the road to foster cooperation

This chapter is based on the article:
Lamotte, R., de Palma, A. and Geroliminis, N., 2017. On the use of reservation-based autonomous vehicles for demand management. Transportation Research Part B: Methodological, 99, pp.205-227.

The work has been performed by the candidate under the supervision of Prof. André de Palma and Prof. Nikolas Geroliminis.

5.1 Introduction

Shared autonomous vehicles, also known as autonomous taxis or automated mobility on demand, are widely seen as a very likely future for urban passenger mobility. According to the International Transport Forum (2015a), it represents one of the two main pathways envisioned towards the large-scale development of autonomous vehicles, together with the gradual automatization of privately owned vehicles. For several decades already, studies have been conducted to evaluate the potential impacts of automation in fields such as road safety (Jamson et al., 2013), traffic flow on highways (Varaiya, 1993), or on urban roads (Fajardo et al., 2012; Qian et al., 2014). Yet, the potential of this mobility revolution for demand management remains largely unexplored. We investigate in this chapter how a central planner should administer systems involving both conventional and bookable autonomous vehicles on separated roads.

We consider a single bottleneck dynamic setting, in which the capacity of a freeway is divided between conventional and bookable autonomous vehicles (referred to hereafter as independent and cooperative, respectively). Users of conventional vehicles freely choose their departure time from home, while users of bookable vehicles choose in advance their time slot among those proposed by a central operator. Booking users are guaranteed no delay at the bottleneck. The main challenges addressed in this chapter are then to determine how a central planner should allocate the capacity to these two vehicle types depending on the regime (laissez-faire, welfare- or profit-maximizing) and in the cases of the welfare or profit-maximizing regimes, which constant tolls/subsidies should be applied on each route. The bottleneck capacity is assumed to be perfectly divisible.

Booking is a particularly well-suited congestion management tool under unpredictable conditions. With congestion pricing, a time-dependent toll is set based on predictions of the characteristics of the population. Then, given this toll, users compete to minimize their own cost. If both the road operator and the users were able to predict perfectly the traffic conditions, queues could be avoided with appropriate time-dependent tolls (Vickrey, 1969). In practice however, the demand on a specific day is only revealed once the vehicles are on the road and have selected a departure time. Thus, the efficiency of the toll is determined by the quality of the predictions, which are notoriously difficult to make. On the contrary, booking does not rely on predictions. Time-slot booking has regularly been proposed over the last two decades as an alternative to congestion pricing for demand management (Wong, 1997; De Feijter et al., 2004; Edara and Teodorović, 2008; Liu et al., 2015). Probably because it would impose some important habit change and appears difficult to implement with conventional vehicles, it has never gained the same support as congestion pricing. Yet, the emergence of large-scale car sharing permitted by automation may challenge this status quo as booking is already well accepted for fleet management and autonomous vehicles remove the need for enforcement.

In theory, the cooperative system described could also be implemented with conventional private cars. Yet, the fact that booking has remained rather anecdotic in the literature on road congestion suggests that some important drawback has not been accounted for and prevents this development. In this work, an individual-specific cost of cooperation is introduced.[1] This cost accounts for all inconveniences related to the cooperative service, except the price, the travel time and the schedule penalties, which are modeled as in Vickrey (1969). The most important inconvenience modeled is certainly the need for users to schedule their tasks in advance. A similar cost was already proposed by Tisato (1992) and Fosgerau (2009) in the context of public transit to distinguish planning from non-planning users.

Of course, the cooperation cost is likely to depend on the flexibility that the system offers to its customers. While quantifying this effect is beyond the scope of this thesis, the rapid emergence of diverse forms of mobility services involving a similar scheduling inconvenience (car-sharing, ride-sharing and, to a lesser extent, e-hailing and on-demand public transit) suggests that a significant proportion of the population is willing to bear this cost in exchange for some benefits associated with mobility as a service. Since a cooperative service would bring the additional benefit of avoiding congestion for essentially the same cost, its materialization in the context of mobility as a service seems plausible.

[1]If registration to the service is considered on single-trip basis, this cost would become de facto trip-specific.

The proposed interpretation of the reservation service as a car-sharing system based on autonomous vehicles has important consequences from a practical point of view. First, self-driving vehicles and car-sharing systems are likely to expand together. Indeed, self-driving technologies and car-sharing are complementary: their combination allows for a door-to-door service and addresses issues related to relocation, liability and maintenance (International Transport Forum, 2015a,b). Second, isolating autonomous vehicles from conventional vehicles would remove several obstacles to their development and would allow for higher capacity utilization (Varaiya, 1993), hence bringing another justification for the allocation of a reserved infrastructure. Third, self-driving technologies would naturally resolve the difficulties associated to the enforcement of the slot allocation mechanism as the vehicles would pick up their passengers only at the scheduled time. Fourth, booking also facilitates fleet management. In fact, as many car-sharing services are already reservation-based, the implementation of demand management strategies would simply bring more benefits to the current users and potentially attract new ones. In terms of our analytical results however, the only impact of automation stems from the gain in capacity associated to introducing autonomous vehicles on a separate roadway. This gain is accounted for by considering a different capacity for autonomous vehicles.

The remainder of this chapter is structured as follows: Section 2 provides further details on the cooperation cost, suggests some possible interpretations for the booking service and discusses the impacts of the service on the effective capacity of the roadway. The differences with existing congestion management techniques are also highlighted. Section 3 lays the background by introducing the general assumptions and the expressions of the different costs. Then, the user equilibrium and social optimum are treated in Section 4 for both fixed capacity splits and optimal capacity splits. It is shown in particular that the user equilibrium can be made socially optimal with a simple constant toll (decentralization of the social optimum) but that unlike the social optimum, the user equilibrium with no toll and with a socially optimal capacity split Pareto-dominates the case with no cooperation. It is also shown numerically that the equilibrium and the social optimum allow very similar social cost reductions. The management of a route by a private operator is considered in Section 5 to assess the impact of profit-maximizing strategies on the social cost and finally, conclusions are drawn and suggestions for future research are stated in Section 6.

5.2 Motivation and main concepts

5.2.1 The cost of cooperation

The cost of cooperation can be interpreted as the alternative-specific constant generally introduced in mode choice models (Small and Verhoef, 2007). Here, it is individual specific, i.e. it is distributed in the population. Although this cost accounts mostly for the need to schedule a trip in advance, user experiences might also differ in many other dimensions, depending on the interpretation chosen for the booking service.

For example, the proposed interpretation of the service as a car-sharing system utilizing autonomous vehicles entails several such differences. Indeed, not only the car running costs might differ but also the safety, the value of the in-vehicle time, the free-flow travel time itself, the travel time variability, etc. In this chapter, we assume that all these differences between the two routes or types of vehicles can be approximated by constant terms in the generalized cost function (i.e. that the associated costs do not depend on the demand and capacity splits) such that these inconveniences can be considered as parts of the cooperation cost. This assumption seems reasonable for the car running cost (if we exclude the fuel consumed while waiting), for the safety consequences, for the free-flow travel time and for the in-vehicle value of time.[2] The impact of the cooperative scheme proposed on the travel time variability would deserve to be further studied. While the travel time with autonomous vehicles might be considered as fixed, it is not clear yet how the demand and capacity split would influence travel time variability for the conventional vehicles and how this variability should be valued. This is still an active research topic (see e.g. Fosgerau (2010), Li et al. (2010), Fosgerau and Karlström (2010), Carrion and Levinson (2012), Xiao et al. (2017)).

5.2.2 Implementation of advance booking

"Advance booking" is understood in this thesis as a means to force the early allocation of time slots, thus avoiding queuing but introducing a scheduling inconvenience. In order to distribute the demand over time as efficiently as possible, the choice of the slot allocation mechanism is paramount. At the frontier between game theory and economics, such mechanism-design problems have been extensively studied (Hurwicz and Reiter, 2006) and a variety of tools have been proposed. For the problem at hand, Liu et al. (2015) have already proposed to distribute time slots of the bottleneck capacity via auctions. Alternatively, the time slots could be issued by the road manager before being exchanged on a trading market (Wada and Akamatsu, 2013). More ideas have been suggested in the context of other well-known transportation-related problems such as airport slot allocation (Rassenti et al., 1982) or coordination of autonomous vehicles at intersections (Schepperle et al., 2008).

The only restriction imposed by our methodology is that the system should be *fair*, in the sense described hereafter. In the general case, cooperative users experience a cost $C^c(j,t) = C_{coop}(j) + \mathrm{SP}(t) + C^c_{run} + C_{comp}(t)$, where $C_{coop}(j)$ is the cooperation cost of individual j, $\mathrm{SP}(t)$ is the schedule penalty for arriving at time t (as in Vickrey (1969)), C^c_{run} is the running cost (vehicle purchase, maintenance, fuel...) and $C_{comp}(t)$ is an

[2]This would most likely not be true if autonomous and conventional vehicles had to share the same road as in van den Berg and Verhoef (2016). Nevertheless, since autonomous vehicles are assumed to be separated spatially from other vehicles and since users have the same scheduling preferences, the individual costs do not depend on the value of the in-vehicle time (Arnott et al., 1990).

additional cost, that might be positive or negative and acts as a *compensation mechanism*. Similarly, independent users experience a cost $C^i(j,t) = W(t) + \text{SP}(t) + C^i_{run}$, where $W(t)$ is the waiting cost if arrival occurs at t and C^i_{run} is the running cost of driving a private car. Since the demand is inelastic and the running costs are constant, only their difference matters. Yet, by defining the cooperation cost as the sum of the inconvenience related to cooperating, plus the reduction/increase in running cost that results from sharing vehicles ($C^c_{run} - C^i_{run}$), running costs can be ignored altogether. Then $C_{comp}(t)$ is a term that compensates for individual variations in schedule penalties incurred (this will play a role in the acceptability of such a system). It has a zero-average, such that the cooperative service as a whole is neither taxed nor subsidized. Formally, if $\bar{\text{SP}}$ denotes the average schedule penalty experienced by all users of the cooperative service, setting $C_{comp}(t) = \bar{SP} - \text{SP}(t)$ ensures that the average compensation cost is equal to zero and leads to $C^c(j,t) = C_{coop}(j) + \bar{\text{SP}}$.[3] Note that even though $C_{comp}(t)$ may be negative, there is no need to actually pay users as long as $C_{comp}(t) < |C^c_{run}|$. In this case, $C_{comp}(t)$ is merely an off-peak discount (or peak charge). This is only one example of compensation mechanism. Further research can identify additional mechanisms, which will be fair and acceptable to potential users. We refer to this situation as the "no-toll equilibrium". The impact of a tax or subsidy is studied in Section 5.4. The notations introduced in this Section were only aimed at making explicit the different cost components but are replaced by more convenient notations in the remainder of the chapter.

5.2.3 Improvement of the effective capacity

Besides avoiding waiting time, the reservation service proposed is expected to increase the effective capacity of the roadway, i.e. the rate at which vehicles can pass the bottleneck. If a part of the bottleneck normally serves up to z conventional vehicles per time unit, it would serve up to gz vehicles per time unit when allocated to the cooperative service, where g is called the capacity improvement factor. The amplitude of g highly depends on the interpretation adopted.

In case the service is assumed to be provided by autonomous vehicles only, the difference in effective capacity could be dramatic. Although there is no consensus on the scale of this change, the impact is globally expected to be positive on highways, with authors estimating capacity improvements ranging from 20% to more than 100% (Varaiya, 1993). For urban networks, the impact is even more uncertain. Different research groups working on intersection control showed that communication technologies could allow for reductions in delays at intersections (Fajardo et al., 2012; Qian et al., 2014). On the contrary, Le Vine et al. (2015) argued that users of autonomous vehicles would have a lower

[3] In practice, users have different values of time and different schedule preferences. Since users would be able to acquire their preferred combination of schedule penalty/price paid, the average cost would represent an upper bound, given the social efficiency of second-price sealed-bid auction (which would be implemented via a mobile app for instance) (Vickrey, 1961).

acceptance to accelerations and decelerations, which would actually slow down traffic in urban areas. Progress in terms of safety could also significantly improve the capacity overall. Indeed, since traffic incidents are responsible for approximately 10-30% of the total congestion delay (see e.g. Skabardonis et al. (2003)) and since human error is the critical factor in the vast majority of crashes, reducing crashes would also increase the average capacity of roadways.

Based on the literature above, various values of g between 1 and 2 are used for numerical applications, corresponding to improvements between 0% and 100%. Fortunately, the theoretical results obtained remain valid for any value of the capacity improvement factor $g > 0.5$, which seems to be an extremely reasonable assumption.

5.2.4 Comparison with other demand management strategies

As highlighted in the introduction, congestion pricing and advance booking differ mostly in terms of acceptability and timing. In more practical terms, the simultaneous booking of the vehicle and of the road utilization confers to the proposed system several advantages over congestion pricing. Since a potential time-dependent price can be included within the vehicle fare, this price can (i) have a zero-average (a negative price is simply a discount on the vehicle fare) and hence be more acceptable, (ii) be collected without additional administrative overhead. Besides, although this is difficult to evaluate, peak pricing might be politically more acceptable and easier to implement if it is applied by a fleet operator to its mobility service than by the government to the road network. In fact, peak pricing is already implemented by many airlines, railways and more recently by mobility companies like Uber (Surowiecki, 2014).

Besides these differences, congestion pricing and advance booking share many common characteristics and the analysis of the literature on partial road pricing can provide some precious insight. Systems in which a fine toll is applied only on one of two parallel routes have already been considered in the past by many authors Braid (1996); Liu and McDonald (1999); de Palma and Lindsey (2000). Although the social optimum in the cases studied appeared to be obtained by pricing the entire capacity, partial tolling was justified by legal, political and technical constraints requiring the existence of a free-access alternative. By introducing wealth heterogeneity, Verhoef and Small (2004) and van den Berg and Verhoef (2011) shed some light on these political constraints, using respectively static and dynamic models of congestion. Both studies found first-best pricing to provide larger benefits on average but in a much less egalitarian way than second-best partial pricing. By modeling in addition the reduction of capacity that is commonly caused by congestion, Hall (2018) demonstrated that partial tolling can be used to produce a Pareto improvement before redistributing the toll revenues. Thus, while in these cases the social optimum imposes a unique system, a more diverse offering is relevant because it is more egalitarian. The idea of reserving access to some part of the road to a group of users (e.g. carpools) during some time slot is also proposed by Fosgerau (2011). This work differs from ours however as the "privileged" users still compete with each others, thus generating queuing.

5.3 Optimum and Equilibrium

This section focuses on the socially optimal and equilibrium demand splits, for fixed or optimal capacity splits. The socially optimal capacity split of course depends on the regime considered (equilibrium or optimum). Note also that the expression "social optimum" refers in this chapter to the socially optimal demand split, not to the absence of congestion on both routes. The notations used in this chapter are listed in Table 5.1.

5.3.1 Background

Setting

Let us consider a single origin/single destination situation with only one route and a bottleneck of capacity S and a total (inelastic) demand N. This route can be divided into two parallel routes, which have bottlenecks at the same location as the original route and which are reserved for independent and cooperative users, respectively. The proportion of the demand that is cooperative and the proportion of the bottleneck capacity that is allocated to them are denoted by x and $y \in [0,1]$. However, for the reasons given above, the effective capacity of the cooperative route is likely to differ as it is used by autonomous vehicles. Thus, a cooperative infrastructure with a "traditional-vehicle-capacity" of yS would have an effective capacity of gyS, where the capacity improvement factor g would characterize how much more/less efficiently the facility is used (g would be equal to 1 if the cooperative service was implemented with conventional cars).

Individual costs

We consider that all users have $\alpha - \beta - \gamma$ preferences (described in Section 2.2) with the same coefficients α, β, γ and t^*. On the part of the road allocated to independent users, $(1-x)N$ users want to pass a bottleneck with capacity $(1-y)S$, hence the total duration of congestion is $\frac{(1-x)N}{(1-y)S}$. Note that it corresponds to the situation studied in Section 2.4.2. The individual equilibrium congestion cost, denoted c_{i}, is given by:

$$c_{\mathrm{i}} = \begin{cases} \frac{\delta N}{S}\frac{1-x}{1-y} & \text{if} \quad y \in [0,1) \\ \infty & \text{if} \quad y = 1. \end{cases} \tag{5.1}$$

Since their departure times are properly scheduled, cooperative users do not have any queuing time. However, they still incur a schedule penalty cost, which is a uniformly

Variables	Unit	Description
α	\$/h	unit cost of travel time
β	\$/h	unit cost of arriving early
γ	\$/h	unit cost of arriving late
δ	\$/h	$\triangleq \frac{\beta\gamma}{\beta+\gamma}$
t^*	h	common desired departure time from the bottleneck
S	veh/h	capacity of the bottleneck
N	veh	demand
x	-	proportion of the demand that is cooperative
y	-	proportion of the capacity that is reserved for cooperative users
g	-	capacity improvement factor
θ	-	type of user or (normalized) cooperation cost
$\hat{\theta}$	-	critical value of θ separating cooperative from independent users
κ	\$	unit cooperation cost
f, F	-	probability (pdf) and cumulative (cdf) density function of θ
Θ	-	support of f
$\underline{\theta}, \bar{\theta}$	-	infimum and supremum of Θ
τ	\$	a toll imposed for each trip on the independent route
$\tau_{\mathrm{pc}},\ \tau_{\mathrm{pi}},\ \tau_{\mathrm{gc}},\ \tau_{\mathrm{gi}}$	\$	toll imposed for each trip by the government (g) or the private operator (p) on the cooperative (c) or on the independent (i) route
$\hat{\theta}^{\mathrm{pc}}, \hat{\theta}^{\mathrm{pi}}$	-	critical value of θ at equilibrium with the tolls τ_{pc} and τ_{pi}
$c_{\mathrm{c}}, c_{\mathrm{i}}$	\$	individual congestion cost for cooperative and independent users
r	-	cost ratio equal to $\frac{\delta N}{\kappa S}$
$y^{\mathrm{o}}(\hat{\theta}),\ \hat{\theta}^{\mathrm{o}}(y),\ x^{\mathrm{o}}(y)$	-	socially optimal values, given the variables in parentheses
$\hat{\theta}^{\mathrm{e}}(y), x^{\mathrm{e}}(y)$	-	user equilibrium values (no toll) for a given capacity split y
$\theta^{\mathrm{sol}}(a, b)$	-	solution of $\hat{\theta} + aF(\hat{\theta}) - b = 0$ (see Lemma 3)
$\hat{y}$	-	$\triangleq F\left(\left(1 - \frac{1}{2g}\right) r\right)$, capacity split such that at equilibrium $x = y$
$\mathrm{SC}(\hat{\theta}, y)$	\$	social cost as a function of $\hat{\theta}$ and y
$\mathrm{SC}(x, y)$	\$	social cost as a function of x and y
$\mathrm{SC}(\hat{\theta})$	\$	$\triangleq \mathrm{SC}\left(\hat{\theta}, y^{\mathrm{o}}(\hat{\theta})\right)$, social cost function with a socially optimal capacity split
$\mathrm{SC}(y)$	\$	$\triangleq \mathrm{SC}\left(\hat{\theta}^{\mathrm{o}}(y), y\right)$, social cost function with a socially optimal demand split
$\mathrm{SC}^{\mathrm{ref}}, c^{\mathrm{ref}}$	\$	social and individual cost in the reference scenario (no cooperation)
$\widetilde{\mathrm{SC}}$	-	$\triangleq \frac{\mathrm{SC}}{\mathrm{SC}^{\mathrm{ref}}}$; the same tilde notation is used for other variables and it always indicates the ratio of the variable divided by its value in the reference scenario (no cooperation).

Table 5.1 – Notations

distributed random variable taking values between 0 and $\frac{\delta N}{S}\frac{x}{gy}$. By assuming the existence of a compensating mechanism (see Section 5.2.2), we can consider that each user experiences the average schedule penalty, that is $\frac{\delta N}{S}\frac{x}{2gy}$. In addition, cooperative users also incur a cooperation cost, which is a characteristic of each individual. The probability density function (pdf) of this cost in the entire population is assumed to be known and to satisfy the following condition:

Condition 1. *The support of the probability density function of the cooperation cost is an interval including* 0*, positive for at least some users, bounded below but not necessarily above.*

Note that this condition allows the cooperation cost to be negative for some users. This might happen, for instance, if travel time reliability is highly valued or if the technology used for cooperative vehicles has significant advantages. Since the cooperation cost is the only source of heterogeneity, users can be characterized by their individual cooperation cost, relatively to the entire population: an individual of type θ has the cooperation cost $\kappa\theta$, where κ is referred to as the unit cooperation cost. The pdf of the type θ is denoted by f, its support by Θ, and its infimum and supremum by $\underline{\theta}$ and $\bar{\theta}$. Condition 1 imposes that $\underline{\theta} \in \mathbb{R}^-$ and $\bar{\theta} \in \mathbb{R}^{+*} \cup \{+\infty\}$. With this notation, the average cost for a cooperative individual of type θ is:

$$c_{\mathrm{c}} + \kappa\theta = \begin{cases} \frac{\delta N}{S}\frac{x}{2gy} + \kappa\theta & \text{if} \quad y \in (0,1] \\ \infty & \text{if} \quad y = 0, \end{cases} \tag{5.2}$$

where c_{c} is referred to as the congestion cost for cooperative users.

Social cost

Since individuals differ only by their cooperation cost, there exists a critical type denoted by $\hat{\theta}$ such that all individuals of type $\theta < \hat{\theta}$ are cooperative, while all individuals of type $\theta > \hat{\theta}$ are independent. In other words, the cooperative population consists only of the individuals with the smallest values of θ, both under user equilibrium and under social optimum. Note that $\widehat{\theta}$ might potentially be equal to $\underline{\theta}$ or $\bar{\theta}$, in which case all users belong to the same category. The proportion of the demand that is cooperative is simply given by the cumulative distribution function (cdf) F of the cooperation cost θ evaluated at $\hat{\theta}$ (see Fig. 5.1):

$$x(\hat{\theta}) = F(\hat{\theta}). \tag{5.3}$$

By assumption (cf Condition 1), $f(u) > 0$ for all $u \in (\underline{\theta}, \bar{\theta})$ so x is a strictly increasing function of $\hat{\theta}$. Consequently, $\hat{\theta} \to x(\hat{\theta})$ is a bijection from $[\underline{\theta}, \bar{\theta}]$ to $[0, 1]$. Then, the social

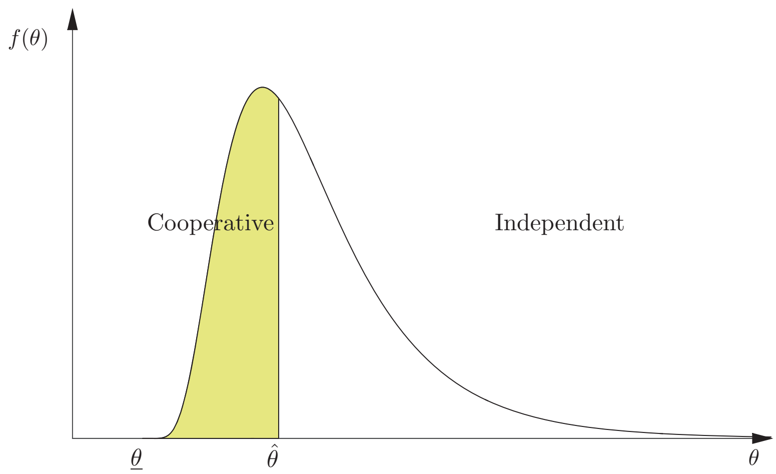

Figure 5.1 – Example of a distribution of the cooperation cost among the population and separation between cooperative and independent users

cost is expressed as a function of the demand and capacity splits by:

$$\mathrm{SC}(\hat{\theta}, y) = \begin{cases} N\kappa \displaystyle\int_{\underline{\theta}}^{\bar{\theta}} u f(u)\,\mathrm{d}u + \frac{\delta N^2}{S}\frac{1}{2g} & \text{if } (\hat{\theta}, y) = (\bar{\theta}, 1) \\ N\kappa \displaystyle\int_{\underline{\theta}}^{\hat{\theta}} u f(u)\,\mathrm{d}u + \frac{\delta N^2}{S}\frac{x^2}{2gy} + \frac{\delta N^2}{S}\frac{(1-x)^2}{1-y} & \text{if } (\hat{\theta}, y) \in [\underline{\theta}, \bar{\theta}] \times (0,1) \\ \frac{\delta N^2}{S} & \text{if } (\hat{\theta}, y) = (\underline{\theta}, 0), \end{cases} \tag{5.4}$$

where $x = x(\hat{\theta})$ according to Eq. (5.3). The social cost is infinite for $\hat{\theta} \neq \underline{\theta}$ and $y = 0$ and for $\hat{\theta} \neq \bar{\theta}$ and $y = 1$.

Besides the demand and capacity splits x and y, Eq. (5.4) also involves the exogenous variables N, S, κ and δ. To further simplify this expression and those derived thereafter, we now introduce the relative social cost, that is the social cost divided by the social cost in a reference scenario. The reference scenario chosen here is the situation with no cooperation at all ($x = 0$, $y = 0$). This case is referred to hereafter as the "user equilibrium with no cooperation". In this case, the social cost is simply $\mathrm{SC}^{\mathrm{ref}} = \frac{\delta N^2}{S}$. Similarly, we will also consider individual costs relative to the individual cost in the reference scenario $c^{\mathrm{ref}} = \frac{\delta N}{S}$ and the duration the route is used, relative to $\frac{N}{S}$. With this transformation, the relative social cost is given by:

$$\widetilde{\mathrm{SC}}(\hat{\theta}, y) \triangleq \frac{\mathrm{SC}(\hat{\theta}, y)}{\mathrm{SC}^{\mathrm{ref}}} = \begin{cases} \frac{1}{r} \displaystyle\int_{\underline{\theta}}^{\hat{\theta}} u f(u)\,\mathrm{d}u + \frac{1}{2g} & \text{if } (\hat{\theta}, y) = (\bar{\theta}, 1) \\ \frac{1}{r} \displaystyle\int_{\underline{\theta}}^{\hat{\theta}} u f(u)\,\mathrm{d}u + \frac{x^2}{2gy} + \frac{(1-x)^2}{1-y} & \text{if } \left(\hat{\theta}, y\right) \in [\underline{\theta}, \bar{\theta}] \times (0,1) \\ 1 & \text{if } (\hat{\theta}, y) = (\underline{\theta}, 0), \end{cases} \tag{5.5}$$

where $r \triangleq \frac{\delta N}{\kappa S}$. Note that r has a physical interpretation: it represents the ratio of the congestion costs (if all users are independent, $\frac{\delta N}{S}$ is the individual cost, i.e. the sum of the schedule delay penalty and of the travel time cost) and the unit cooperation cost. Note that the relative social cost is only a function of the capacity y, the capacity improvement factor g, the cost ratio r and the critical cooperation cost $\hat{\theta}$. If both the demand split and the capacity split are optimally chosen, the relative social cost depends only on g and r. Similarly, it is shown in Section 5.3.2 that the demand split under user equilibrium is only a function of g, r and y and that hence, the relative social cost under user equilibrium with a socially optimal capacity split is also a function only of g and r.

We will also denote by $\widetilde{\mathrm{SC}}$ the relative social cost functions that use other input arguments. However, to avoid any ambiguity, the relevant input arguments will always be mentioned. Later in this work, the functions $\widetilde{\mathrm{SC}}(\hat{\theta}) = \widetilde{\mathrm{SC}}(\hat{\theta}, y^{\mathrm{o}}(\hat{\theta}))$ and $\widetilde{\mathrm{SC}}(y) = \widetilde{\mathrm{SC}}(\hat{\theta}^{\mathrm{o}}(y), y)$ will be used to refer to relative social cost functions of one variable only, assuming that the other split is fixed (or optimal).

Specific distributions and numerical values

An effort was made throughout this work to keep a general scope and assumptions about specific distributions or numerical values were avoided when appropriate. However, some analytical expressions and the graphical illustrations require assumptions. When necessary (and it is always mentioned), a uniform distribution is assumed for the cooperation cost. Some numerical applications are also replicated with a log-normal or an exponential distribution with the same expected value.

Based on Eq. 5.5, the results depend only on dimensionless relative quantities that are used in all graphical illustrations. However, some intuition about reasonable values of $r = \frac{\delta N}{\kappa S}$ is critical to assess the scale of the benefits that should be expected. The numerical evaluation of the reference individual cost $\frac{\delta N}{S}$ is relatively common. Typical values of earliness ($\beta = 0.5\alpha$) and lateness ($\gamma = 2\alpha$) (Small, 1982) lead to $\delta = \frac{\beta\gamma}{\beta+\gamma} = 0.4\alpha$, where α is the value of time at home while $\frac{N}{S}$ is simply the length of the peak period ($\sim$ 2 h for instance).[4]

Concerning the scale κ of the cooperation cost, mode choice models can suggest an educated guess since most of them also require that users plan their trips in advance (e.g. carpooling, buses with long headways, trains, and planes). Bhat (1995) for instance proposed different models for mode choice between cities, including a multinomial logit.

[4] Since user homogeneity artificially increases the average cost, we chose a peak period that is shorter than what is commonly observed to produce realistic average costs. For a peak period of 2 h, homogeneous users and the scheduling preferences mentioned above, the earliest user would be 96 min early, the latest 24min late and the on-time user would have a trip duration that is 48 min longer than the free-flow travel time. These costs are rather large, but would be doubled for a 4h peak period (which is closer to the real duration of the peak period in large cities).

The ratio of the mode-specific parameter and of the parameter associated to in-vehicle travel time leads to the following estimates: the train has a mode-specific cost that is approximately equal to the cost of 51 min of travel time and the mode-specific cost of taking the plane is about 62 min (compared to the car). However, this intrinsic utility does not only account for the cooperation cost but also for other characteristics of the mode which penalize public transit (e.g. comfort). Thus, the cooperation cost would most likely have a smaller value if personal vehicles were used, say around 30 min. If the time unit is 1h, then the average value of $\kappa\theta$ should be around 0.5α so with a uniform distribution on $[0,1]$, this leads to $\kappa = \alpha$. Altogether, these estimates lead to the best guess $r = \frac{\delta N}{\kappa S} \sim 0.8$.

5.3.2 Fixed capacities

In this Section, the capacity split is considered as a parameter, such that the only variable is the demand split. The results of this restricted case are not only useful to derive results in the relaxed case (in which the capacity split is also a variable, see Section 5.3.3) but they also have direct implications for the real world, as the capacity allocated to the car-sharing service might be constrained for various reasons (e.g. road geometry).

Optimal demand split with fixed capacities

Since the cooperation cost plays a crucial role in this section, we will consider the relative social cost $\widetilde{\mathrm{SC}}$ as a function of the type $\hat{\theta}$ rather than the demand split x. Before studying potential optimal demand splits, we state the following technical lemma, which will be useful for different propositions.

Lemma 3. *Let $G(\hat{\theta}) = \hat{\theta} + aF(\hat{\theta}) - b$, where $a, b > 0$. Then $G(\hat{\theta}) = 0$ has a unique interior solution denoted by $\hat{\theta}^{sol}$ if and only if $\bar{\theta} + a - b > 0$. If the probability density function f is continuous, then $\hat{\theta}^{sol}$ is a locally continuous and differentiable function of a and b, decreasing with a and increasing with b.*

Proof. Since $\underline{\theta} \leq 0$, $G(\underline{\theta}) = \underline{\theta} - b < 0$. Besides, $\lim_{\hat{\theta}\to\bar{\theta}} G(\hat{\theta}) = \bar{\theta} + a - b$. Thus, if $\bar{\theta} + a - b > 0$, the intermediate value theorem ensures the existence of a solution and since G is strictly increasing, this solution is unique. Else, $G(\hat{\theta}) < 0$ for all $\hat{\theta} < \bar{\theta}$ so there can be no interior solution.

Let us now see G as a function of $\hat{\theta}$, a and b. G is simply affine with a and b so if f is continuous, G is continuously differentiable with: $\frac{\partial G}{\partial \hat{\theta}} = 1 + af(\hat{\theta})$, $\frac{\partial G}{\partial a} = F(\hat{\theta})$ and $\frac{\partial G}{\partial b} = -1$. As $\hat{\theta}$ is interior, all the previously mentioned derivatives are invertible so the implicit function theorem implies that $\hat{\theta}^{\mathrm{sol}}$ is locally continuous and differentiable with:

$$\frac{\partial \hat{\theta}^{\mathrm{sol}}}{\partial a}(a,b) = -\frac{F(\hat{\theta}^{\mathrm{sol}})}{1+af(\hat{\theta}^{\mathrm{sol}})} < 0 \text{ and } \frac{\partial \hat{\theta}^{\mathrm{sol}}}{\partial b}(a,b) = \frac{1}{1+af(\hat{\theta}^{\mathrm{sol}})} > 0.$$

□

Proposition 5. *For a given capacity split and for a cooperation cost verifying Condition 1, there exists a unique demand split that minimizes the social cost. If the capacity split is interior ($y \in (0,1)$), then the optimal demand split is interior and is the unique solution of*

$$\frac{1}{r}\hat{\theta}^o + \frac{1+(2g-1)y}{gy(1-y)}x(\hat{\theta}^o) - \frac{2}{1-y} = 0. \tag{5.6}$$

In addition, $\hat{\theta}^o$ is a continuous function of the capacity improvement factor (for $g \in [0.5,+\infty)$), the capacity split (for $y \in [0,1]$) and the cost ratio (for $r \in (0,+\infty)$), increasing with g and y. It increases (resp. decreases) with r if $\hat{\theta}^o > 0$ (resp. $\hat{\theta}^o < 0$).

Proof. If $y \in \{0,1\}$, there is only one demand split that yields a finite social cost, so the result is trivial. Let us now consider $y \in (0,1)$. By differentiating Eq. (5.5): $\forall \hat{\theta} \in \Theta$,

$$\begin{aligned}\frac{\mathrm{d}\widetilde{\mathrm{SC}}}{\mathrm{d}\hat{\theta}}(\hat{\theta}) &= \frac{1}{r}\hat{\theta}f(\hat{\theta}) + \frac{x(\hat{\theta})}{gy}f(\hat{\theta}) - \frac{2\left(1-x(\hat{\theta})\right)}{1-y}f(\hat{\theta}) \\ &= \frac{1}{r}f(\hat{\theta})G(\hat{\theta}),\end{aligned}$$

where we define: $G(\hat{\theta}) \triangleq \hat{\theta} + r\left(\frac{1}{gy}+\frac{2}{1-y}\right)x(\hat{\theta}) - \frac{2r}{1-y}$. G has the form of the function required by Lemma 3 with:

$$\bar{\theta} + a - b = \bar{\theta} + r\left(\frac{1}{gy}+\frac{2}{1-y}\right) - \frac{2r}{1-y} = \bar{\theta} + \frac{r}{gy} > 0.$$

Thus, there exists a unique demand split $\hat{\theta}^{\mathrm{o}} \in (\underline{\theta},\bar{\theta})$ satisfying $G(\hat{\theta}^{\mathrm{o}}) = 0$, i.e. satisfying Eq. (5.6).

Finally, since $\frac{f(\hat{\theta})}{r} > 0$, $\forall \hat{\theta} \in S_\theta$, $\frac{\mathrm{d}\widetilde{\mathrm{SC}}}{\mathrm{d}\hat{\theta}}(\hat{\theta})$ has the same sign as $G(\hat{\theta})$, i.e. $\frac{\mathrm{d}\widetilde{\mathrm{SC}}}{\mathrm{d}\hat{\theta}}(\hat{\theta}) \gtreqless 0$ for $\theta \gtreqless \hat{\theta}^{\mathrm{o}}$ so $\widetilde{\mathrm{SC}}$ reaches its global minimum for $\hat{\theta} = \hat{\theta}^{\mathrm{o}}$.

The second part of Lemma 3 implies that $\hat{\theta}^{\mathrm{o}}$ is locally continuous and increases with g. Since this is valid for every $g \in [0.5,+\infty)$, $\hat{\theta}^{\mathrm{o}}$ is continuous and increasing everywhere. Similarly, by the implicit function theorem, $\hat{\theta}^{\mathrm{o}}$ is continuous and increases (resp. decreases) with r on $(0,+\infty)$ if $\hat{\theta}^{\mathrm{o}} > 0$ (resp. $\hat{\theta}^{\mathrm{o}} < 0$) and is continuous and increases with y on $(0,1)$. To generalize this last result to the closed interval $y \in [0,1]$, note first that $\hat{\theta}^{\mathrm{o}} = \underline{\theta}$ for $y = 0$ and $\hat{\theta}^{\mathrm{o}} = \bar{\theta}$ for $y = 1$. Then Eq. (5.6) can be rewritten as: $\frac{x(\hat{\theta}^{\mathrm{o}})}{gy(1-y)} = -\frac{\hat{\theta}^{\mathrm{o}}}{r} - \frac{2g-1}{g(1-y)}x(\hat{\theta}^{\mathrm{o}}) + \frac{2}{1-y}$. As the left-hand term is clearly positive, so must be the right-hand term. In addition, the right-hand term is bounded above by $-\frac{\underline{\theta}}{r} + \frac{2}{1-y}$. Consequently, $x(\hat{\theta}^{\mathrm{o}})$ is non-negative and bounded above by $gy(1-y)\left[-\frac{\underline{\theta}}{r} + \frac{2}{1-y}\right]$, and thus converges toward 0 when y tends towards 0, which ensures continuity in 0. Similarly, Eq. (5.6) can also be rewritten as $1 - y = \frac{r}{\hat{\theta}^{\mathrm{o}}}\left[2 - 2x(\hat{\theta}^{\mathrm{o}}) - \frac{1-y}{gy}x(\hat{\theta}^{\mathrm{o}})\right]$. The left-hand term clearly converges towards 0 when y tends towards 1, so the right-hand term must do so as well. The right-hand term however is the product of two terms that are related: if the first is small (i.e. $\hat{\theta}^{\mathrm{o}}$ is big), then y must be close to 1 so the second term must be small as well, and vice versa. Thus it is trivial to show that both terms tend towards 0, or, equivalently, that y converges towards 1. □

The existence and uniqueness results stated in Proposition 5 confirm an intuitive result: if the road capacity is split into two, the optimal demand split is such that both roads are used and its exact value depends on how much users dislike cooperating. As we have not specified the distribution of the cooperation cost yet, the exact value of the optimal demand split is given by an implicit equation. This equation could be made explicit by assuming some particular distribution, as done in Section 5.3.2.

Note also that if we set $\kappa = 0$ (i.e. no cooperation cost), the problem studied is equivalent to solving for the socially optimal fine toll when the capacity of a road is divided into two and only one road can be tolled. Eq. (5.6) then reduces to $x^{\mathrm{o}} = \frac{2gy}{1+(2g-1)y}$, or equivalently $\frac{x^{\mathrm{o}}}{gy} = \frac{2(1-x^{\mathrm{o}})}{1-y}$. This imposes that the ratio of demand to capacity (or equivalently, the duration the route is used) is twice as big for the cooperative route as for the independent one, as found by Braid (1996).

User equilibrium with fixed capacities

At equilibrium, no user can reduce her cost by changing her decision. Hence, either only one route is used and its cost is smaller than the cost on the other road for all users, or both roads are used and their costs are equal for some critical user (because of the continuity assumption included in Condition 1). If tolls are small enough, both roads are necessarily used at equilibrium as Condition 1 imposes that some user has zero cost of cooperation. The equilibrium condition is then:

$$\kappa\hat{\theta}^{\mathrm{e}}(y,\tau) + c_{\mathrm{c}} = c_{\mathrm{i}} + \tau, \tag{5.7}$$

where τ is a constant toll on the independent route (if the toll is on the cooperative route, then τ is simply negative) and $\hat{\theta}^{\mathrm{e}}(y,\tau)$ characterizes the critical user that is indifferent between the two routes. Hereafter, in cases with no toll, $\hat{\theta}^{\mathrm{e}}(y,0))$ is simply denoted $\hat{\theta}^{\mathrm{e}}(y)$.

Proposition 6. *For a given capacity split and with no toll, a demand split satisfies* *the user equilibrium equation* (5.7) *with a unit cooperation cost* $\kappa = K$ *if and only if the demand split satisfies the social optimum equation* (5.6) *with a unit cooperation cost* $\kappa = 2K$*.*

Proof. Eq. (5.7) is equivalent to

$$\kappa\hat{\theta}^{\mathrm{e}}(y) + \frac{\delta N}{S}\frac{x(\hat{\theta}^{\mathrm{e}}(y))}{2gy} = \frac{\delta N}{S}\frac{(1 - x(\hat{\theta}^{\mathrm{e}}(y)))}{1-y},$$

or

$$r\left(\frac{1+(2g-1)y}{2gy(1-y)}\right)x(\hat{\theta}^{\mathrm{e}}(y)) = r\frac{1}{1-y} - \hat{\theta}^{\mathrm{e}}(y). \tag{5.8}$$

It suffices now to note that Eq. (5.8) with $\kappa = K$ is identical to the social optimum equation (5.6) for $\kappa = 2K$. □

A consequence of Proposition 6 is that the results obtained for the social optimum remain valid after multiplying κ by two. In particular:

Corollary 3. *For a given capacity split and with no toll, there exists a unique demand split satisfying the user equilibrium equation. It is a continuous function of the capacity improvement factor g (on $[0.5, +\infty)$), of the cost ratio r (on $(0, +\infty)$), and of the capacity split y (on $[0, 1]$).*

In addition, Proposition 6 also highlights the important role played by the cooperation cost in the problem considered. Indeed, by taking the extreme case $\kappa = 0$, one obtains the following corollary:

Corollary 4. *If cooperation is not costly, the equilibrium demand split with no toll is socially optimal.*

By adding a cooperation cost, we impose that on one hand, the critical user at equilibrium equalizes her own cooperation cost with the difference in *her* congestion cost on the two routes $c_i - c_c$. On the other hand, the social optimum is obtained by equalizing this same critical cost of cooperation *for one user* with the change in the total congestion cost *for all users.* Hence, the choice of the route entails externalities which are not accounted for at the equilibrium. Without a cost of cooperation however, one can observe that imposing $\frac{\partial \widetilde{\mathrm{SC}}}{\partial x} = 0$ is mathematically equivalent to imposing $c_c = c_i$, i.e. the externality imposed by the route choice of the critical user at the user equilibrium is equal to 0. This is in agreement with the result of Braid (1996), who found that, if a welfare-maximizing operator can only toll one of two parallel routes, it should add to Vickrey's fine toll a negative flat component, that attracts more demand to the route with no queuing. The negative flat component is such that the toll is equal to zero on average, i.e., it corresponds to our compensating mechanism (see Section 5.2.2). This leads us to state this additional corollary:

Corollary 5. *For any interior capacity split, there exists a toll of the same sign as the critical cooperation cost that increases the social welfare when applied to the independent route.*

If a toll is imposed, it can be proven that there still exists a unique demand split satisfying the user equilibrium condition but in this case, the solution is not necessarily interior. The proof is left to the reader as it is similar to the proof of Proposition 5.

Decentralization of the social optimum

The objective of this section is to determine, given some capacity split, how the routes should be tolled to shift the user equilibrium to socially optimal conditions. Corollary 5 implies that the independent route should be tolled and Proposition 6 can be used to determine the exact amount without additional calculations.

Indeed, if one solves the social optimum problem to find $\hat{\theta}^{\mathrm{o}}$ and then sets the toll to $\tau = \frac{1}{2}\kappa\hat{\theta}^{\mathrm{o}}$, Eq. (5.7) becomes:

$$\kappa\left(\hat{\theta}^{\mathrm{e}}\left(y, \frac{1}{2}\kappa\hat{\theta}^{\mathrm{o}}\right) - \frac{1}{2}\hat{\theta}^{\mathrm{o}}\right) + c_{\mathrm{c}} = c_{\mathrm{i}}.$$

Proposition 6 implies that $\hat{\theta}^{\mathrm{o}}$ is a solution and since the solution is unique, $\hat{\theta}^{\mathrm{e}}\left(y, \frac{1}{2}\kappa\hat{\theta}^{\mathrm{o}}\right) = \hat{\theta}^{\mathrm{o}}$, i.e. the equilibrium is the social optimum.

Intuitively, this toll forces some naturally independent users that are close to being cooperative to become cooperative. Hence, it is natural that it should increase with the cooperation cost of these users ($\kappa\hat{\theta}^{\mathrm{o}}$).

It is of practical interest to notice that this toll is only paid by independent users, is time-independent and is relatively small (numerical applications show that for the range of κ considered, it is about half the average toll required by Vickrey's time-dependent tolling strategy). Note however that the social optimum reached has a higher social cost of congestion than the one produced by Vickrey's fine toll since only cooperative users avoid queuing. Thus, while the user acceptability of such a pricing strategy in a possible implementation is expected to be higher, it does not convey the same extraordinary benefits as the ideal fine toll.

Special case: uniform distribution

In order to find closed-form expressions of the demand split for the social optimum and user equilibrium, we assume a uniform distribution of the type θ: $f(\theta) = 1$ if $\theta \in [0,1]$ and $f(\theta) = 0$ elsewhere. Hence, $x = \hat{\theta}$ for all $\hat{\theta}$ in $[0,1]$, i.e. the demand split is equal to the critical type of user. Under these conditions, the social optimum equation Eq. (5.6) has an explicit solution given by

$$\hat{\theta}^{\mathrm{o}} = \frac{2rgy}{gy(1-y) + r(1 + (2g-1)y)}. \tag{5.9}$$

It can be easily verified that this solution is interior, i.e. that $\forall y \in (0,1), \hat{\theta}^{\mathrm{o}} \in (0,1)$. The expression of the equilibrium demand split can be derived immediately from Eq. (5.9) using Proposition 6:

$$\hat{\theta}^{\mathrm{e}} = \frac{2rgy}{2gy(1-y) + r(1 + (2g-1)y)}. \tag{5.10}$$

The demand split and the times both routes are used are displayed in Fig. 5.2a and Fig. 5.2b for $r = 0.2$, $r = 0.4$, $r = 1$ and $r = 2$ for the social optimum. Note that by exploiting the correspondence between social optimum and user equilibrium, the same curves also describe the user equilibrium for $r = 0.4$, $r = 0.8$, $r = 2$ and $r = 4$. Some simple considerations and empirical observations in other contexts requiring scheduling (public transit, airlines) suggest a realistic value of r would be in the order of 0.8 (see 5.3.1).

It is clear in Fig. 5.2a that the optimal demand split increases with both r and y (see Proposition 5). Note that besides the very specific cases where the capacity is entirely allocated to one alternative ($x = 0$ and $x = 1$), the curves for different cost ratios r are quite different: those associated to small values of r have a S shape while those associated to large values of r appear to be concave. Nevertheless, all curves are similar for the range of small capacity splits y, which is a natural consequence of the limited number of cooperative users. Indeed, as only the users with almost no cooperation cost are cooperative when the capacity split y is small, such situations can all be approximated by the case with no cooperation derived at the end of Section 5.3.2. Hence, as the independent users have a demand-to-capacity ratio close to 1, the demand-to-capacity ratio of the cooperative user (i.e. the slopes of the curves in Fig. 5.2a or value on the y-axis in Fig. 5.2b) must tend towards 2 as the capacity split y tends towards 0, regardless of the cost ratio r.

Fig. 5.2b illustrates the balance of the flows across the two routes. For small capacity splits y, the demand-to-capacity ratio (i.e. the time the road is used) is twice as large on the cooperative road as on the independent one. For $y = 1$ however, the route that is used during the longest period depends on the cost ratio r. Smaller cost ratios r lead

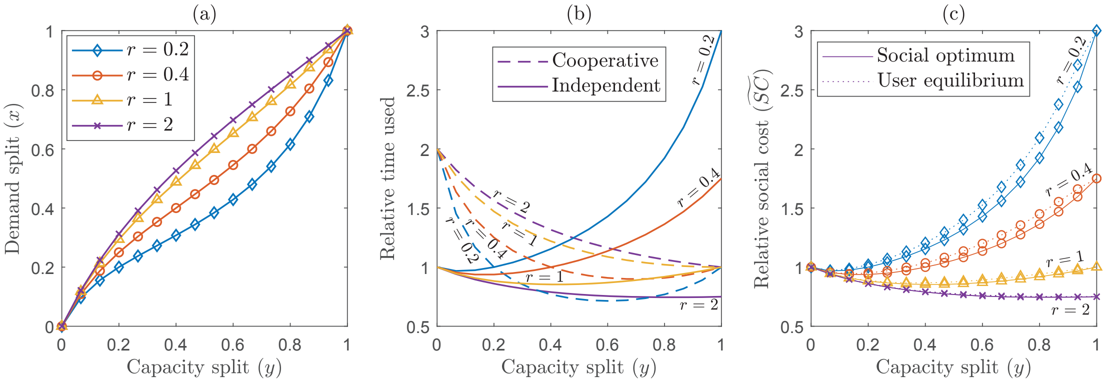

Figure 5.2 – Graphical representations as functions of the capacity split y of (a) the socially optimal demand split (x), (b) the relative time each route is used in the social optimum (1 corresponds to the reference scenario), and (c) the relative social cost for both the social optimum and the user equilibrium. All results were obtained with a uniform distribution of the cooperation cost and are provided for different values of the cost ratio r and for $g = 1$.

to a greater usage of the independent road, which can be used for a period up to three times longer than the cooperative route. Such an imbalance is intuitively not desirable as it implies that the capacity is most of the time only partially used.

The performance of the different regimes can be assessed by the relative social cost, which is represented in Fig. 5.2c for both the social optimum and the user equilibrium. Although the social optimum by definition always yields a smaller social cost than the user equilibrium, the difference between these two regimes appears to be relatively small, especially for large values of the cost ratio r. Intuitively, large values of r correspond to cases in which cooperation has a small cost compared to congestion, and hence to cases where the user equilibrium is close to being socially optimal (see Corollary 4 and the discussion following it).

Thus, cooperation can reduce the social cost in some cases, an ill-adapted capacity split leads to an imbalance between the two routes, which translates either into a very little gain (large r and small y) or into a severe increase in social cost (small r and large y). This highlights the importance of the role of the central planner when allocating road capacity to the two vehicle types.

5.3.3 Optimal capacities

Optimal demand and optimal capacity splits

We now assume that both the demand and the capacity splits can be governed by a central planner and we determine the pair of splits that minimizes the social cost. This problem is most easily solved by first identifying the explicit expression that associates a demand split x to the capacity split y^{o} that minimizes the social cost given the demand split x, and then minimizing the function $\widetilde{\mathrm{SC}}(\hat{\theta}) \triangleq \widetilde{\mathrm{SC}}(\hat{\theta}, y^{\mathrm{o}}(x(\hat{\theta})))$, where we use the same notation with a slight abuse of notation.

Proposition 7. *For any given demand split $x \in [0,1]$, there exists a unique capacity split that minimizes the social cost. It is independent of the cooperation cost and congestion parameter δ, and it is given by:*

$$y^o(x) = \frac{x}{\sqrt{2g}(1-x)+x}. \tag{5.11}$$

This capacity split is continuous and strictly increasing on $[0,1]$. The cooperative route is used $\sqrt{2g}$ times longer than the independent route.

Proof. If $x = 0$ (or $x = 1$), the socially optimal capacity split is trivially given by $y^{\mathrm{o}} = 0$ (resp. $y^{\mathrm{o}} = 1$).

Now, if $x \in (0,1)$, the choices $y = 0$ and $y = 1$ lead to infinite values of the social cost. Thus, we can restrict the search to $y \in (0,1)$. By differentiating Eq. (5.5), we get for $y \in (0,1)$:

$$\frac{\mathrm{d}\widetilde{\mathrm{SC}}}{\mathrm{d}y}(y) = -\frac{x^2}{2gy^2} + \frac{(1-x)^2}{(1-y)^2}.$$

Thus, $\frac{\mathrm{d}\widetilde{\mathrm{SC}}}{\mathrm{d}y}(y^{\mathrm{o}}) = 0$ is equivalent to:

$$\sqrt{2g}\frac{1-x}{1-y^{\mathrm{o}}} = \frac{x}{y^{\mathrm{o}}}, \tag{5.12}$$

so that we obtain Eq. (5.11) for $x \in (0,1)$.

It is trivial to show that this solution is interior for $x \in (0,1)$ and that $\frac{\mathrm{d}^2\widetilde{\mathrm{SC}}}{\mathrm{d}y^2}(y)$ is strictly positive for $y \in (0,1)$. Therefore, the social cost reaches its global minimum for $y = y^{\mathrm{o}}(x)$.

Finally, the function y^{o} is clearly continuous on $(0,1)$ while:

$$\lim_{x \to 0^+} (y^{\mathrm{o}}(x)) = 0 = y^{\mathrm{o}}(0) \ ; \ \lim_{x \to 1^-} (y^{\mathrm{o}}(x)) = 1 = y^{\mathrm{o}}(1).$$

Hence, y^{o} is continuous on $[0,1]$. □

As highlighted in Proposition 7, Eq. (5.12) has the intuitive interpretation that the cooperative route is used $\sqrt{2g}$ times as long as the independent route. Since the total cooperation cost is a constant, the social optimum simply minimizes the total congestion cost and as routes are always either used at capacity or not used at all, the duration a route is used is simply the demand-to-capacity ratio. It is natural that the social optimal requires a higher ratio on the cooperative route as there is no queueing, only schedule penalties.

We then derive an explicit formulation of the relative social cost as a function of only one variable: the demand split x (or alternatively, the critical user $\hat{\theta}$).

Lemma 4. *The relative social cost with a socially optimal capacity split $\widetilde{SC}(\hat{\theta})$ is a continuous function on $[\underline{\theta}, \bar{\theta}]$ and, with $x = x(\hat{\theta})$,*

$$\widetilde{SC}(\hat{\theta}) = \frac{1}{r}\int_{\underline{\theta}}^{\bar{\theta}} uf(u)\,\mathrm{d}u + \frac{\left(\sqrt{2g}(1-x)+x\right)^2}{2g}. \tag{5.13}$$

Proof. $\widetilde{\mathrm{SC}}$ is trivially continuous on $(\underline{\theta}, \bar{\theta})$. Let us now show that $\widetilde{\mathrm{SC}}$ is continuous on the closed interval $[\underline{\theta}, \bar{\theta}]$. By combining Eq. (5.5) and Eq. (5.11) and after some manipulations, we obtain: $\forall \hat{\theta} \in (\underline{\theta}, \bar{\theta})$,

$$\begin{aligned}
\widetilde{\mathrm{SC}}(\hat{\theta}) &= \frac{1}{r}\int_{\underline{\theta}}^{\bar{\theta}} uf(u)\,\mathrm{d}u + \frac{x^2}{2g\frac{x}{\sqrt{2g}(1-x)+x}} + \frac{(1-x)^2}{1-\frac{x}{\sqrt{2g}(1-x)+x}} \\
&= \frac{1}{r}\int_{\underline{\theta}}^{\bar{\theta}} uf(u)\,\mathrm{d}u + \frac{1}{2}\left(\sqrt{2g}(1-x)+x\right)\left[\frac{x}{g} + \frac{2(1-x)^2}{\sqrt{2g}(1-x)}\right] \\
&= \frac{1}{r}\int_{\underline{\theta}}^{\bar{\theta}} uf(u)\,\mathrm{d}u + \frac{1}{2g}\left(\sqrt{2g}(1-x)+x\right)^2.
\end{aligned}$$

We can now evaluate this expression at the boundaries of its domain:

$$\begin{aligned}
\lim_{\hat{\theta}\to\underline{\theta}} \left(\widetilde{\mathrm{SC}}(\hat{\theta})\right) &= \frac{1}{2g}\left(\sqrt{2g}\right)^2 = 1 = \widetilde{\mathrm{SC}}\left(\underline{\theta}\right)\ ; \\
\lim_{\hat{\theta}\to\bar{\theta}} \left(\widetilde{\mathrm{SC}}(\hat{\theta})\right) &= N\kappa\int_{\underline{\theta}}^{\bar{\theta}} uf(u)\,\mathrm{d}u + \frac{1}{2g}\left(\sqrt{2g}(1-1)+1\right)^2 \\
&= \frac{1}{r}\int_{\underline{\theta}}^{\bar{\theta}} uf(u)\,\mathrm{d}u + \frac{1}{2g} = \widetilde{\mathrm{SC}}(\bar{\theta}).
\end{aligned}$$

Therefore, $\widetilde{\mathrm{SC}}$ is continuous on $[\underline{\theta}, \bar{\theta}]$. □

Finally, the results of Lemmas 3 and 4 can be combined to obtain the following proposition.

Proposition 8. *If the cooperation cost satisfies Condition 1, there exists a unique pair of demand and capacity splits that minimizes the social cost. This solution is interior if and only if $r < \bar{\theta}\frac{g}{\sqrt{2g}-1}$. In this case, the demand split $x(\hat{\theta})$ is given implicitly by:*

$$\frac{1}{r}\hat{\theta} + \frac{\left(\sqrt{2g}-1\right)^2}{g}x(\hat{\theta}) - \left(\sqrt{2g}-1\right)\sqrt{\frac{2}{g}} = 0. \tag{5.14}$$

Both the demand split x and capacity split y increase with the cost ratio r and with the capacity improvement factor g. If $r \geq \bar{\theta}\frac{g}{\sqrt{2g}-1}$, at the social optimum all vehicles should cooperate.

Proof. Since the expression of $\widetilde{\mathrm{SC}}$ obtained in Eq. (5.13) is valid on $[\underline{\theta}, \bar{\theta}]$, we will now use it to avoid handling different cases. This function is continuous, differentiable and we have:

$$\begin{aligned}\frac{d\widetilde{\mathrm{SC}}}{d\hat{\theta}}(\hat{\theta}) &= \frac{1}{r}\hat{\theta}f(\hat{\theta}) - \frac{1}{2g}\left[2\left(\sqrt{2g} - \left(\sqrt{2g}-1\right)x(\hat{\theta})\right)(\sqrt{2g}-1)f(\hat{\theta})\right] \\ &= \frac{1}{r}f(\hat{\theta})H(\hat{\theta}),\end{aligned}$$

where $H(\hat{\theta}) \triangleq \hat{\theta} + \frac{r}{g}\left((\sqrt{2g}-1)^2 x(\hat{\theta}) - (2g - \sqrt{2g})\right)$.

Lemma 3 can be applied on function H with $\bar{\theta}+a-b = \bar{\theta}+\frac{r}{g}\left(1+2g-2\sqrt{2g}-2g+\sqrt{2g}\right) = \bar{\theta} - r\frac{\sqrt{2g}-1}{g}$. Thus, given g and $\bar{\theta}$, the existence of an interior solution depends on the cost ratio r.

If $r < \bar{\theta}\frac{g}{\sqrt{2g}-1}$, then $\bar{\theta} > r\frac{\sqrt{2g}-1}{g}$, so according to Lemma 3 there exists a unique $\hat{\theta}^{\mathrm{o}}$ such that $H(\hat{\theta}^{\mathrm{o}}) = 0$. Since $\frac{\mathrm{d}\widetilde{\mathrm{SC}}}{\mathrm{d}\hat{\theta}}$ is negative on $[\underline{\theta}, \hat{\theta}^{\mathrm{o}})$, positive on $(\hat{\theta}^{\mathrm{o}}, \bar{\theta}]$ and equal to zero at $\hat{\theta}^{\mathrm{o}}$, $\hat{\theta}^{\mathrm{o}}$ is the unique global minimum of $\widetilde{\mathrm{SC}}$. Although the comparative statics of lemma 3 do not apply directly here, the implicit function theorem can be used in a very similar fashion to demonstrate that $\hat{\theta}^{\mathrm{o}}$ increases with r and g. Since the demand split is an increasing function of $\hat{\theta}^{\mathrm{o}}$ and the capacity split an increasing function of the demand split, both splits increase with r and g.

If $r \geq \bar{\theta}\frac{g}{\sqrt{2g}-1}$, then $\forall\hat{\theta} \in [\underline{\theta}, \bar{\theta})$, $\frac{\mathrm{d}\widetilde{\mathrm{SC}}}{\mathrm{d}\hat{\theta}} < 0$ and $\frac{\mathrm{d}\widetilde{\mathrm{SC}}}{\mathrm{d}\hat{\theta}}(\bar{\theta}) \leq 0$. Thus, there is a unique global minimum and it is reached for $\hat{\theta} = \bar{\theta}$ (and $x = 1$). From a practical point of view, this means that if the cooperation cost is small enough for all the population, the social optimum is an entirely controlled infrastructure. □

Thus, there is an interior solution if and only if the maximum cooperation cost within the population ($\kappa\bar{\theta}$) is more than $\frac{\sqrt{2g}-1}{g}$ times the individual cost when all users are independent. Since $\frac{\sqrt{2g}-1}{g} \in (0, 0.5]$ for $g > 0.5$, the following sufficient condition holds:

Corollary 6. *If the maximum cooperation cost within the population is greater than half the individual cost with no cooperation ($\kappa\hat{\theta} > \frac{\delta N}{2S}$), the optimal demand and capacity splits are interior.*

The assumption included in Condition 1 mentioning that the closure of the cooperation cost support includes 0 is critical here. Thanks to this assumption, allocating at least some part of the capacity to cooperative users is always socially beneficial. Else, allocating capacity to cooperative users would be detrimental unless there is at least some minimum level of congestion, which would slightly complicate the derivations and properties. However, as waiting time represents on average half the individual congestion cost for homogeneous independent users, it suffices that some users have a cooperation cost smaller than half the individual congestion cost without cooperation to ensure that the optimum features a cooperative route.

Optimal capacity split with equilibrium demand

Even though Proposition 6 shows the existence of a strong relationship between the user equilibrium and the social optimum, the properties of these two regimes are fundamentally different. In particular, while the user equilibrium assumes purely selfish users, the numerical applications provided in Section 5.3.3 highlight how the social optimum requires the "sacrifice" of some cooperative users, for which the cost is worse than in the reference scenario. We demonstrate in this section that unlike the social optimum, the user equilibrium is Pareto-improving for the socially optimal capacity split. This result is obtained without actually determining the socially optimal capacity split under user equilibrium as the calculations involved are particularly tedious even with simplistic assumptions[5].

Before actually demonstrating this result with Propositions 9 and 10, let us define

$$\hat{y} = F\left(\left(1 - \frac{1}{2g}\right) r\right).$$

As shown in the following lemma, the capacity $\hat{y}$ plays a very particular role as it separates situations where the demand-to-capacity ratio is greater for cooperative users from situations where it is greater for independent users.

Lemma 5. *Let y denote a capacity split and $x^{\mathrm{e}}(y)$ denote the associated demand split at user equilibrium with no toll. If the distribution of the cooperation cost satisfies Condition 1, then $x^{\mathrm{e}}(y) > y$ for all $y \in (0, \hat{y})$, $x^{\mathrm{e}}(y) < y$ for all $y \in (\hat{y}, 1)$ (if $\hat{y} < 1$), and $x^{\mathrm{e}}(y) = y$ for $y \in \{0, \hat{y}, 1\}$.*

Proof. First, it is trivial that $x^{\mathrm{e}}(0) = 0$ and $x^{\mathrm{e}}(1) = 1$ as users have no choice in these conditions. Second, one can verify that that if $y = \hat{y}$, then $x = \hat{y}$ is a user equilibrium. Indeed, if the critical user is of type $\hat{\theta} = (1 - \frac{1}{2g})r$, the total individual cost for the critical user is (following (5.2)) $\kappa\hat{\theta} + \frac{\delta N x}{2gSy} = \frac{\delta N}{S}(1 - \frac{1}{2g}) + \frac{\delta N}{2gS} = \frac{\delta N}{S}$. As the demand split x is equal to the capacity split y, this is also the individual cost for independent users,

[5] Assuming a uniformly distributed cooperation cost, the differentiation of the function associating to a capacity split its social cost under user equilibrium leads to a rational function whose numerator is a 4^{th} order polynomial.

so this is a user equilibrium. Corollary 3 ensures uniqueness so $x = \hat{y}$ is the only user equilibrium for the capacity split $y = \hat{y}$.

Now, let $y \in (0, \hat{y})$. The reasoning above shows that $x = y$ is not a user equilibrium because the critical cooperation cost would be smaller than $(1 - \frac{1}{2g})r$, i.e. it would be in the interest of some independent users to become cooperative. Hence, as the individual costs for a given capacity split are monotonous functions of the demand split, the user equilibrium necessarily verifies $x^{\mathrm{e}}(y) > y$.

Conversely, if $\hat{y} < 1$ and $y \in (\hat{y}, 1]$, $x = y$ is not a user equilibrium either because it would be in the interest of some cooperative users to become independent. Thus, the equilibrium verifies $x^{\mathrm{e}}(y) < y$. □

While these comparisons of demand-to-capacity ratios might seem abstract, we argue below that a capacity split smaller than $\hat{y}$ is also necessary and sufficient to have a Pareto improvement.

Proposition 9. *For any distribution of the cooperation cost verifying Condition 1, the user equilibrium with no toll Pareto-dominates the user equilibrium with no cooperation if and only if the capacity split y is in the interval $(0, \hat{y}]$.*

Proof. By applying Lemma 5:

$$y \in (0, \hat{y}] \Leftrightarrow \begin{cases} x^{\mathrm{e}}(y) \geq y \\ y > 0 \\ \text{if } y = 1, \text{ then } \hat{y} = 1. \end{cases} \tag{5.15}$$

Then, in order to obtain a Pareto improvement, three conditions should be verified: (i) the independent users are not worse-off, (ii) the user with the critical cooperation cost is not worse-off on the cooperative route, and (iii) at least one user is better off.

Let us first show that $y \in (0, \hat{y}]$ implies that there is a Pareto improvement. Based on the equation of the individual cost for independent users (5.1), Condition (i) is equivalent to $x \geq y$, i.e. the proportion of the demand that is cooperative should not be smaller than the proportion of the capacity they are allocated. Note then that if Condition (i) is verified and $y < 1$, Condition (ii) must be verified as well since the critical user is indifferent at equilibrium. If $y = 1$, users have no choice, so we cannot use the indifference argument. However, as highlighted by Eq. (5.15), then we must have $\hat{y} = 1$, so the cooperative user with the biggest cost is still not worse-off. In addition, if Condition (ii) is verified and $x > 0$ (which is guaranteed by Condition 1 as long as $y > 0$), then there are some users with a cooperation cost strictly smaller than the critical one who are better off. Thus, $y \in (0, \hat{y}]$ implies that there is a Pareto improvement.

Conversely, $y = 0$ corresponds to the reference scenario, which violates Condition (iii). If $y > \hat{y}$, then Lemma 5 implies that either $x^{\mathrm{e}}(y) < y$ or $y = 1$ and $\hat{y} < 1$. The first case clearly violates Condition (i) while the second violates Condition (ii). Indeed, in this last situation $x^{\mathrm{e}}(y) = y$ so the congestion cost for cooperative users is equal to $\frac{\delta N}{2gS}$. The cooperation cost that makes a user indifferent between the reference scenario and cooperation in this scenario is reached for $x(\hat{y}) = \hat{y}$. As $y > \hat{y}$, there are more cooperative users for $y = 1$ and these additional cooperative users are all worse-off. □

Intuitively, even though there may be a higher demand-to-capacity ratio on the cooperative route at equilibrium, cooperative users must always have a cost that is smaller than independent users (otherwise they would be independent). Thus, we should simply ensure that independent users are better-off. This requires that they have a demand-to-capacity ratio that is smaller than in the reference scenario, which, according to Lemma 5, happens if and only if $y \leq \hat{y}$ (and $y > 0$). For $y > \hat{y}$, the capacity split allocated to cooperative users exceeds the proportion of the demand that is disposed to cooperate. Hence, independent users are not better off unless some users with large cooperation cost cooperate, which does not occur at equilibrium.

Note that this result might not strictly hold when diverse but highly correlated sources of heterogeneity co-exist. If, for instance, the cooperation cost is strongly negatively correlated with the value of travel time α, then providing cooperation might separate users into two groups with little intra-group heterogeneity. Thus, the travel time of the independent users (who have high cooperation cost and hence small values of α) arriving near t^* may increase even if independent users have a better demand-to-capacity ratio. Such effects should be further studied as individual characteristics are often highly correlated in the real world.

Pareto improvements are extremely important from a political point of view. While socially optimal policies should theoretically be sought, it happens that sub-optimal measures are implemented instead, simply because they are Pareto-improving, which is not necessarily the case of socially optimal measures. In the situation at hand however, we argue below that the socially optimal capacity split exists and Pareto-dominates the user equilibrium with no cooperation.

Proposition 10. *For any distribution of the cooperation cost verifying Condition 1, there exists a capacity split that minimizes the social cost of the user equilibrium with no toll. The associated user equilibrium Pareto-dominates the user equilibrium with no cooperation.*

Proof. By applying Corollary 3, the demand split is a continuous function of the capacity split for $y \in [0, 1]$. Thus, the social cost under user equilibrium is also a continuous function of the capacity split on $[0, 1]$ and the extreme value theorem guarantees the existence of a capacity split minimizing the social cost.

If $\hat{y} = 1$, all capacity splits $y > 0$ are Pareto-improving so the result is trivial. Let us now assume that $\hat{y} < 1$. Let $Y > \hat{y}$ and X be a capacity split and its associated demand split at user equilibrium. We demonstrate hereafter that the social cost at user equilibrium for $y = Y$ is bigger than for $y = \hat{y}$.

First, although the pair $(x = X, y = X)$ is not an equilibrium, we can show that $\widetilde{\mathrm{SC}}(X, Y) > \widetilde{\mathrm{SC}}(X, X)$ (we use in this proof the relative social cost $\widetilde{\mathrm{SC}}$ as a function of the capacity split y and of the demand split x, rather than of the type $\hat{\theta}$). Indeed, note that the number of cooperative users is identical so the cooperation cost is exactly the same and we just have to compare the costs of congestion. We consider here the demand split as given so the total congestion cost is simply a function of the capacity split: $g(y) = \frac{X^2}{2gy} + \frac{(1-X)^2}{1-y}$. Differentiating this expression leads to $g'(y) = -\frac{X^2}{2gy^2} + \frac{(1-X)^2}{(1-y)^2}$, which is positive for all $y > X$. Thus, $\widetilde{\mathrm{SC}}(X, Y) > \widetilde{\mathrm{SC}}(X, X)$.

Second, since the demand split is a strictly increasing function of the capacity split, $X > x(\hat{y}) = \hat{y}$. We can then show that $\widetilde{\mathrm{SC}}(X, X) > \widetilde{\mathrm{SC}}(\hat{y}, \hat{y})$. In fact, the individual costs are the same in these two scenarios for all users that do not change their decision (i.e. those that are independent or cooperative in both scenarios). In addition, $x = y = \hat{y}$ is an equilibrium so all users that are independent in these conditions would be worse-off if they were forced to be cooperative with the same congestion conditions, which is exactly what happens in the situation $(x = X, y = X)$.

Thus, $\widetilde{\mathrm{SC}}(X, Y) > \widetilde{\mathrm{SC}}(X, X) > \widetilde{\mathrm{SC}}(\hat{y}, \hat{y})$ so the socially optimal capacity split verifies $y \leq \hat{y}$. To conclude, note that $y = 0$ is clearly not socially optimal as it is Pareto-dominated by any $y \in (0, \hat{y}]$. □

In other words, Proposition 10 states that the socially optimal capacity split exists and cannot be greater than $\hat{y}$, i.e. that "sacrificing" the users with a high cooperation cost to reduce the travel time of those with a lower cooperation cost is overall detrimental. Note however that if we introduce a new source of heterogeneity by considering that users have different values of time (but the same relative value of earliness $\frac{\beta}{\alpha}$ and lateness $\frac{\gamma}{\alpha}$) and that the time of users with low cooperation cost is more valuable (shift-workers usually have lower wages), Proposition 10 might not stand anymore. This is considered as a future research direction.

Special case: uniform distribution

As in the case with a fixed capacity split, we assume a uniform distribution to provide explicit solutions for the social optimum and numerically compare it with the user equilibrium.

Proposition 11. *If the cooperation cost is uniformly distributed on* $[0, 1]$*, then the socially optimal demand and capacity splits and the corresponding social costs are given* *by:*

$$\begin{array}{llll} \textit{For all } r < \bar{\theta}\frac{g}{\sqrt{2g}-1}: & \hat{\theta}^o = \frac{r\sqrt{2g}\left(\sqrt{2g}-1\right)}{g+r\left(\sqrt{2g}-1\right)^2}, & y^o = r\frac{\sqrt{2g}-1}{g}, & \widetilde{SC}^o(\hat{\theta}^o) = \frac{g}{g+r\left(\sqrt{2g}-1\right)^2}. \\ \textit{For all } r \geq \bar{\theta}\frac{g}{\sqrt{2g}-1}: & \hat{\theta}^o = 1, & y^o = 1, & \widetilde{SC}^o(\hat{\theta}^o) = \frac{1}{2r} + \frac{1}{2g}. \end{array}$$

Proof. In the case of a uniform distribution and assuming that the optimum is interior (i.e. $r < \bar{\theta}\frac{g}{\sqrt{2g}-1}$), the optimality condition given by Eq. (5.14) becomes

$$\hat{\theta}^{\text{o}} + \frac{r}{g}\left(\left(\sqrt{2g}-1\right)^2\hat{\theta}^{\text{o}} - \left(2g-\sqrt{2g}\right)\right) = 0,$$

which is equivalent to the expression of $\hat{\theta}^{\text{o}}$ provided in the proposition. Then, with a uniform distribution Eq. (5.11) can be rewritten:

$$y^{\text{o}} = \frac{\hat{\theta}^{\text{o}}}{\sqrt{2g}-(\sqrt{2g}-1)\hat{\theta}^{\text{o}}},$$

and by using the expression of $\hat{\theta}^{\text{o}}$ found above, we obtain the required expression for y^{o}. For $r \geq \bar{\theta}\frac{g}{\sqrt{2g}-1}$, Proposition 8 imposes that $\hat{\theta}^{\text{o}} = y^{\text{o}} = 1$, regardless of the distribution chosen.

The expression of the relative social cost is also greatly simplified for a uniform distribution of the cooperation cost on $[0, 1]$. Indeed, Eq. (5.13) becomes:

$$\widetilde{\text{SC}}(\hat{\theta}) = \frac{\hat{\theta}^2}{2r} + \frac{\left(\sqrt{2g}\left(1-\hat{\theta}\right)+\hat{\theta}\right)^2}{2g}.$$

The final expressions provided in the proposition are then simply obtained by replacing $\hat{\theta}$ by the values of $\hat{\theta}^{\text{o}}$ found above. □

As explained in Section 5.3.3, it is much more complicated to obtain similar analytical expressions for the user equilibrium. Hence, the optimization was carried out numerically. The demand and capacity splits obtained and the resulting social costs are displayed together in Fig. 5.3a. Fig. 5.3b provides a comparison of the costs borne by different individuals and Fig. 5.3c provides a decomposition of the social cost into its different components (waiting time, schedule penalties and cooperation cost).

The most striking result of this numerical application is certainly the very small social cost difference between the user equilibrium and social optimum in Fig. 5.3a. Even when the difference is the largest (around $r = 1$), the user equilibrium achieves more than 90% of the social cost reduction associated to the social optimum. This suggests that decentralizing the social optimum might not be necessary for practical applications. Note also that the social costs at the optimum and at the equilibrium are both decreasing

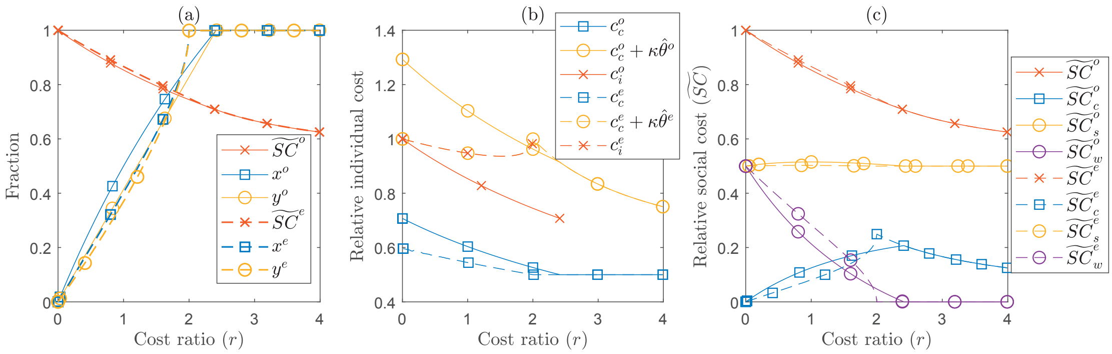

Figure 5.3 – (a) Demand and capacity splits and the resulting relative social costs for the social optimum and user equilibrium as functions of the cost ratio r. (b) Comparison of the relative individual costs of the independent users (c_i) on one hand and of the two extreme cooperative users (the one with no cooperation cost (c_c) and the one with the critical cooperation cost ($c_c + \kappa\hat{\theta}$)) on the other hand. (c) Decomposition of the relative social cost obtained into waiting time $\widetilde{\mathrm{SC}}_w$, schedule penalties $\widetilde{\mathrm{SC}}_s$ and cooperation costs $\widetilde{\mathrm{SC}}_c$ for the social optimum and user equilibrium. All results were obtained with a cooperation cost uniformly distributed between 0 and 1 and with the capacity improvement factor $g = 1$. The superscripts "o" and "e" denote respectively the social optimum and user equilibrium.

functions of the cost ratio r. Although it is not visible here, both actually converge towards $1/2g$ as r tends towards infinity.[6]

Besides this small difference in social cost, the optimum and equilibrium differ in terms of demand-to-capacity ratios. One can see in Fig. 5.3a that for a large range of values of the cost ratio r, the demand split x is greater than the capacity split y under socially optimal conditions, while the difference is much smaller under equilibrium conditions. Hence, even though the proposed system is aimed at removing queues for cooperative users, under socially optimal conditions it would be biased in favor of the independent users, in the sense that they benefit from a proportionally larger share of the infrastructure than the cooperative users. This inequity is particularly obvious in Fig.5.3b, which shows the individual costs of some particular users. Indeed, for values of r between 0 and about 1.8, the cooperative user with the largest cooperation cost is worse-off compared to the reference scenario (i.e. her relative individual cost is greater than 1), while the individual cost of the independent users is significantly reduced. Note that such inequalities cannot occur at the user equilibrium as by construction, the cooperative user who has the largest individual cost has the same cost as the independent users. Hence, the savings from the cooperation benefit mostly the cooperative users at the equilibrium, while the social optimum forces more users to cooperate, thereby generating a redistributive effect.

Finally, while the social cost is a convenient measure of performance, it aggregates several

[6]This is the classic result for Vickrey's bottleneck model that can be obtained without cooperation cost, with $g = 1$ and with a fine toll on the entire route - cf Arnott et al. (1990).

inconveniences that are very different: schedule penalties, waiting time and cooperation. The decomposition of the social cost into these different components is displayed in Fig. 5.3c. While the sum of schedule penalties hardly changes, cooperation greatly reduces the total waiting time. In fact, while the social cost is only reduced by approximately 14% for $r = 1$, the waiting time is reduced by about 56% at the social optimum and 42% at the user equilibrium. Since the external costs ignored in this chapter are related primarily to the presence of vehicles on the road (e.g. pollution and travel time variability), the estimation of the benefits of cooperation provided in this paper is considered conservative.

Numerical applications with other distributions

In order to obtain simple closed-form expressions for the social optimum, it was assumed in the previous part that the cooperation cost was uniformly distributed. For some values of r, the results suggest that the entire infrastructure should be allocated to cooperative users. However, common sense suggests that such a uniform distribution is not realistic. In fact, even if most trips could be scheduled in advance at a rather low cost, there will still remain some emergency trips that cannot be scheduled, such that a more realistic distribution should allow for large positive values of the cooperation cost. Hence, two additional distributions with supports $[0, +\infty)$ are considered: the exponential and log-normal distributions. To allow for a fair comparison, the parameters of the log-normal distribution were set to $\mu = \ln(\frac{1}{2}) - \frac{1}{8}$ and $\sigma = 0.5$ (μ and σ are the expected value and standard deviation of $\ln(\theta)$) while the rate parameter λ of the exponential distribution was set to two. Thus, the expected value of the type θ is equal to 0.5 with both distributions, as for the uniform distribution. These distributions are represented in Fig. 5.4a. The demand split and the relative social cost were numerically evaluated with these distributions and the results are plotted in Fig. 5.4b and Fig. 5.4c, together with the uniform distribution case.

Impact of an improvement of the effective capacity

As discussed in Section 5.2.3, the scheme proposed is likely to improve the effective capacity of the roadway due to the introduction of autonomous vehicles, their isolation on allocated roadways, and hypercongestion avoidance. We investigate how this potential capacity improvement would impact the way the cooperative system should be implemented.

As expected, a more efficient use of the roadway by cooperative vehicles allows for further reductions in the social cost (see Fig. 5.5) and this gain naturally increases with r, as the proportion of cooperative users increases. For instance for $r = 1$ and $g = 1$, the optimum reduces the social cost by about 14.5% while for $r = 1$ and $g = 2$, the reduction exceeds 33%. Note also that the maximum social cost difference between equilibrium and optimum increases with the capacity improvement factor g. Intuitively, the capacity improvement factor g only amplifies the effect of cooperation. The congestion cost is

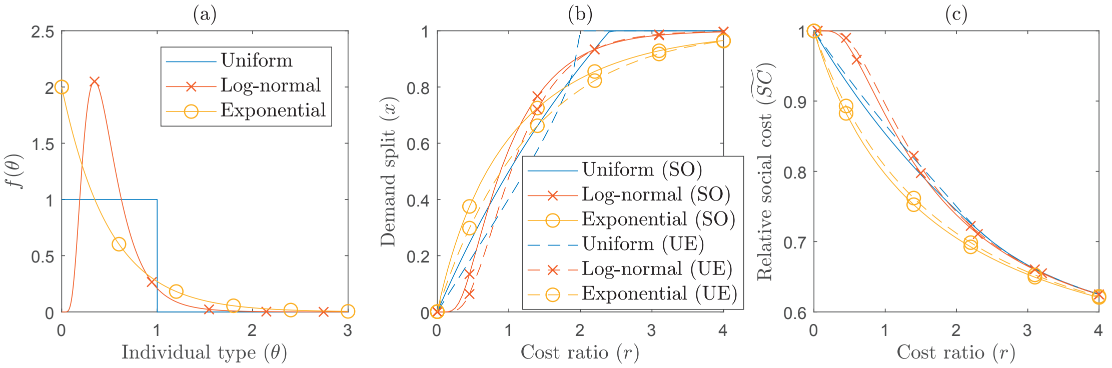

Figure 5.4 – (a) Probability density functions of the three distributions considered for the cooperation cost. The subplots (b) and (c) illustrate the socially optimal (SO) and equilibrium (UE) demand split (b) and social cost (c) for socially optimal capacity splits as functions as the cost ratio r for the three distributions considered for the cooperation cost (uniform, log-normal and exponential).

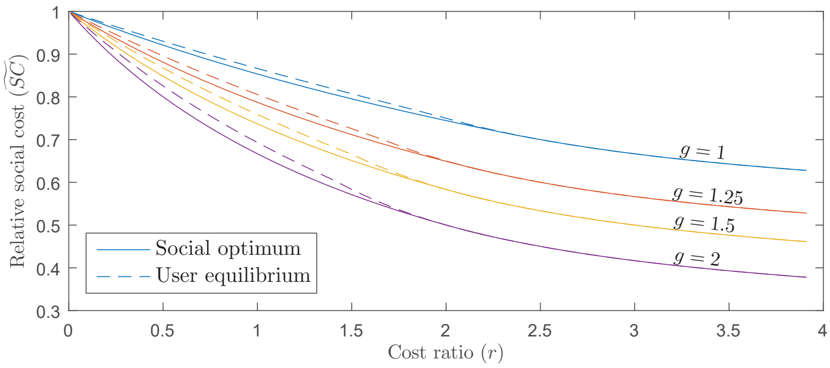

Figure 5.5 – Impact of the capacity improvement factor g on the relative social cost as a function of the cost ratio r for the social optimum and user equilibrium.

divided by two by avoiding queues, and it is further divided by g by increasing the effective capacity. We explained after the Corollary 4 that the critical user compares her own cooperation cost against her own travel time reduction when choosing her route selfishly, while she would compare it against the travel time reduction for all users if she were choosing what is socially optimal. As the difference between these two quantities increases with the impact of cooperation, the difference between equilibrium and optimum increases with the capacity improvement factor g.

5.4 Private operator

This last section investigates the compatibility of the cost-reducing scheduling service introduced above with profit-maximizing objectives. Given the current global enthusiasm for privatization, it seems in fact very likely that if such a scheduling service were to be implemented, its operation would be left to some independent organization, as it is already the case for about one third of highways in Western Europe (Verhoef, 2007). If this independent operator is not subsidized, it would need to collect revenue, most likely via a toll. As discussed in Section 5.3.2 however, there are already fewer users that are cooperative at the equilibrium than at the social optimum. Thus, it seems a priori preferable that either the private operator manages the independent route (even though this does not address the issue of the scheduling service management), or that two tolls should are applied. Based on these first considerations, the exact impact of profit-maximizing strategies is studied hereafter, first with only one toll set by a private operator on one of the two routes, and then with two tolls defined in a Stackelberg setting in which the government is leader and the private operator adjusts its toll in function of the government's toll. Stackelberg competition is classically considered in games involving several heterogeneous players, where one player is able to implement its decision first, e.g. in a spatial market competition (Wang and Ouyang, 2013; Drezner et al., 2015), when building or expanding private transportation infrastructures (Xiao et al., 2007; van den Berg and Verhoef, 2012) or, in a case more similar to ours, when a central authority concerned with the social optimum delegates the network operations to independent organizations with different objectives (Zhang et al., 2011). Stackelberg competition is believed to be one of the prevailing strategic interactions in many market situations as it often allows all players to make more benefits than, for instance, in a Nash competition (Wang et al., 2014). In the case at hand, the different types of players and the strong grip of the government on public infrastructure are strong arguments for a Stackelberg framework. Note that since the amount tolled is only a transfer of money from some individuals to others, it is not taken into account in the calculation of the social cost.

5.4.1 Profit-maximizing toll (one player only)

In this first sub-section, the private operator is the only player and it sets a toll to maximize its profit. Two cases are considered, depending on the route that is managed by the private operator.

Cooperative service managed by a private operator

Consider that a private operator is given a proportion y of the capacity and manages the cooperative service. The toll is set in order to maximize the profit, which is given by:

$$\Pi = xN\tau_{\text{pc}},$$

where τ_{pc} is the toll set by this private operator on the cooperative route ("p" stands for "private" and "c" for "cooperative"). The only user equilibria that bring some profit to the private operator are such that both routes are used (otherwise, the private operator would have either no customer, or it should pay them to use its route because the other one would have zero cost at least for some users). Thus:

$$\kappa\hat{\theta}(y) + c_{\mathrm{c}}^{\mathrm{e}} + \tau_{\mathrm{pc}} = c_{\mathrm{i}}^{\mathrm{e}}, \tag{5.16}$$

or

$$\kappa\hat{\theta}(y) + \frac{\delta N}{S}\frac{x}{2gy} + \tau_{\mathrm{pc}} = \frac{\delta N}{S}\frac{1-x}{1-y},$$

and therefore

$$\frac{\delta N}{S}\left(\frac{1+(2g-1)y}{2gy(1-y)}\right)x = \frac{\delta N}{S}\frac{1}{1-y} - \kappa\hat{\theta}(y) - \tau_{\mathrm{pc}}. \tag{5.17}$$

Since given a capacity split y, the function $\hat{\theta} \to \frac{\delta N}{S}\left(\frac{1+(2g-1)y}{2gy(1-y)}\right)x(\hat{\theta}) - \frac{\delta N}{S}\frac{1}{1-y} + \kappa\hat{\theta} + \tau_{\mathrm{pc}}$ is strictly increasing, there can be only one equilibrium for a given toll but once again, there is no closed-form expression of this equilibrium demand split.

Special case: uniform distribution With a uniform distribution of the cooperation cost, Eq. (5.17) is equivalent to

$$\hat{\theta}(y) = \frac{\frac{\delta N}{S}\frac{1}{1-y} - \tau_{\mathrm{pc}}}{\kappa + \frac{\delta N}{S}\left(\frac{1+(2g-1)y}{2gy(1-y)}\right)}.$$

The profit is now a second-order polynomial that is maximized at:

$$\tau_{\mathrm{pc}} = \frac{\delta N}{S}\frac{1}{2(1-y)}. \tag{5.18}$$

Interestingly, this toll does not depend on κ and corresponds to the average cost of travel time if all users had to use the independent route. With such a toll, the profit-maximizing demand is:

$$\begin{aligned} N\hat{\theta}^{\mathrm{pc}} &= N\frac{\frac{\delta N}{S}\frac{1}{1-y} - \frac{\delta N}{S}\frac{1}{2(1-y)}}{\kappa + \frac{\delta N}{S}\left(\frac{1+(2g-1)y}{2gy(1-y)}\right)} \\ &= N\frac{rgy}{2gy(1-y) + r(1+(2g-1)y)}, \end{aligned} \tag{5.19}$$

which is exactly half the cooperative demand when there is no toll. Note that since there were already fewer cooperative users in user equilibrium than in social optimum, this profit-maximizing strategy moves the demand split in the "wrong" direction. Thus, such a strategy where the private operator is free to choose his optimal price is not recommended for an implementation.

Independent route managed by a private operator

The same approach can be used for this symmetric case, so we just provide the final results here. The profit-maximizing toll is:

$$\tau_{\mathrm{pi}} = \frac{(2\kappa g y S + \delta N)}{4 g y S}. \tag{5.20}$$

The profit-maximizing demand is:

$$\hat{\theta}^{\mathrm{pi}} = \frac{2gy(1-y) + r(1 + (4g-1)y)}{2[2gy(1-y) + r(1 + (2g-1)y)]}.$$

Numerical applications

The social costs obtained with the profit-maximizing strategies studied in Sections 5.4.1 and 5.4.1 are represented in Fig. 5.6, together with the social costs associated to the user equilibrium with no toll and to the social optimum. When the private operator manages the independent route and has only a small percentage of the total capacity ($y \simeq 0.9$), the profit-maximizing strategy can slightly improve the user equilibrium, especially if r is small (e.g. for $r = 0.5$ and $g = 1$). Intuitively, the private operator forces a few users to become cooperative (which is overall beneficial for the society, as shown by Corollary 5) but not too many as it does not have control over a large part of the capacity and thus cannot force the users to pay extreme costs. However, this situation is not desirable since it leads to a social cost that is higher than the reference social cost. For more reasonable values of r (in the range of $0.5 - 1$), the social optimum is obtained for a capacity split that is quite balanced ($y \simeq 0.3$) and for this range of balanced capacity splits, the two profit-maximizing strategies studied always lead to poor social costs. Thus, it seems that profit-maximizing strategies are not compatible with the objective of minimizing the social cost if there is only one toll applied by the private operator. The same numerical applications were repeated with different values of g ranging between 1 and 2 but the results obtained (not shown here) were extremely similar. Note however that different governmental regulations (like taxing or price caps) could guide such a system to better situations for the social good.

Relationship between capacity and profit Let us now consider that the private operator has to pay a given amount c for each capacity unit that it rents from the government. Assuming that the private operator manages the cooperative route and that it always sets the toll to maximize its profit, the profit depends on the capacity split with the following function:

$$\Pi = N\hat{\theta}^{\mathrm{pc}}\tau_{\mathrm{pc}} - cyS.$$

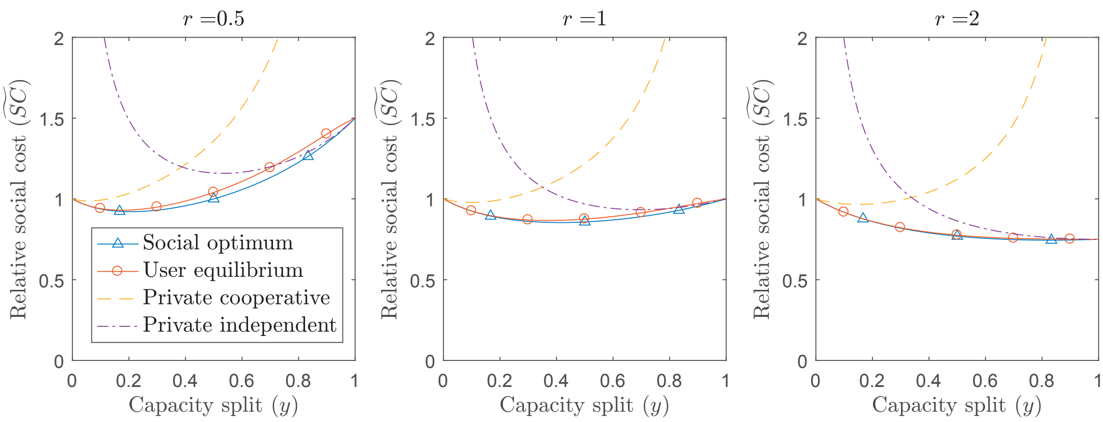

Figure 5.6 – Comparison of the social costs obtained under social optimum, user equilibrium, if the independent route is privately managed and if the cooperative route is privately managed, as functions of the capacity split and for cost ratios r equal to 0.5, 1 and 2. $g = 1$ for all figures.

Obviously, if the demand is not price-elastic and if all the capacity is given to the private operator, the users would be captive and the optimal toll would be infinite. However, for y small enough ($y \to 0$), $\Pi = \left(\frac{\delta N^2}{2S^2} - c\right) yS + o(yS)$. Thus, depending on the value of c, operating only a small part of the route might not be profitable. However, when the capacity that is privately operated gets closer to the full capacity, the situation becomes similar to a monopoly and profits dramatically increase. The situation is very similar if we consider that the private operator manages the independent route. Alternatively, one could consider that the government sets the capacity split y (or equivalently, the value of c) as a function of the toll proposed by the private operator. If the private operator manages the independent route and proposes a toll that is small enough (considering the population's cost ratio r), then the government could allocate to the private operator a capacity split that is such that the toll proposed is optimal.

5.4.2 Stackelberg equilibria

We now study the impact of a profit-maximizing strategy within a Stackelberg competition, where the government and a private company both impose a toll on one route and the capacity split is considered as given. While the government aims at minimizing the social cost (i.e. the sum of the schedule delays, congestion and cooperation costs), the private company aims at maximizing its profit. Since the government has a dominant position, we will consider that the government sets its toll first, knowing how the private company will react (we will see however that the Stackelberg equilibria obtained are also Nash equilibria).

The lack of flexibility of toll is indeed more credible for the public than for the private sector, although this does not prevent from considering some compensation for the potential profit loss of the private sector. In addition, a welfare-maximizing leader should intuitively lead to a higher welfare in the society than the private duopoly. Although this is not always the case (see for example the work of Anderson et al. (1997) with Logit demand function and without congestion effects), this result was proven by de Palma and Lindsey (2000) with an elastic demand and congestion effects for the case where the government and the private operator levy the same type of toll (either flat or fine). The case considered in this work also involves congestion effects but is analogous to having two different types of tolls: one fine toll on the cooperative route (which would be the sum of the flat toll and of the compensation mechanism) and one flat toll on the independent route. The next two subsections investigate the cases where the cooperative and independent routes are managed by the government and the private operator and vice versa.

Cooperative service managed by a private company, independent route by the government

By including an additional toll set by the government τ_{gi} on the independent route in Eq. (5.16), the equilibrium equation becomes:

$$\kappa\hat{\theta}(y) + c_{\mathrm{c}}^{\mathrm{e}} + \tau_{\mathrm{pc}} = c_{\mathrm{i}}^{\mathrm{e}} + \tau_{\mathrm{gi}}.$$

As the demand is assumed inelastic, the equilibrium demand split is uniquely determined by the difference in the two tolls, independently of their level. Thus, the calculations done with only one player (Eqs. (5.16) to (5.17)) remain valid after replacing τ_{pc} by $(\tau_{\mathrm{pc}} - \tau_{\mathrm{gi}})$. Assuming a uniform distribution for the cooperation cost, the equilibrium demand split is:

$$\hat{\theta}(y) = \frac{\frac{\delta N}{S}\frac{1}{1-y} - (\tau_{\mathrm{pc}} - \tau_{\mathrm{gi}})}{\kappa + \frac{\delta N}{S}\left(\frac{1+(2g-1)y}{2gy(1-y)}\right)}. \tag{5.21}$$

Proposition 12. *Assuming that the government is leader, for any capacity split y there exists a unique Stackelberg equilibrium. It is such that the government sets on the independent route the toll:*

$$\tau_{gi} = \frac{\kappa r}{1-y}\frac{r(1+(2g-1)y) + 3gy(1-y)}{r(1+(2g-1)y) + gy(1-y)}. \tag{5.22}$$

This equilibrium is also the social optimum and a Nash equilibrium.

Proof. The profit is a second-order polynomial maximized at

$$\tau_{\mathrm{pc}} = \frac{\tau_{\mathrm{gi}}}{2} + \frac{\delta N}{S}\frac{1}{2(1-y)}. \tag{5.23}$$

This is simply the arithmetic mean of the government toll and of the average congestion cost (schedule penalty and travel time cost) if all users had to use the independent route. By combining equations (5.21) and (5.23):

$$\begin{aligned}\hat{\theta}^{\mathrm{pc}} &= \frac{2\delta N gy + 2gyS(1-y)\tau_{\mathrm{gi}}}{2\kappa gyS(1-y) + \delta N(1+(2g-1)y)} - \\ &\quad - \frac{2gyS(1-y)}{2\kappa gyS(1-y) + \delta N(1+(2g-1)y)} \frac{\delta N + (1-y)S\tau_{\mathrm{gi}}}{2S(1-y)} \\ &= \frac{\kappa rgy + gy(1-y)\tau_{\mathrm{gi}}}{2\kappa gy(1-y) + \kappa r(1+(2g-1)y)}. \end{aligned} \tag{5.24}$$

The only decision variable is now τ_{gi} and the government should set its value such that it minimizes the social cost. Although it is a priori not necessarily feasible, if the government can set a toll such that the demand found in Eq. (5.24) is equal to the demand found in Eq. (5.9), then this toll is optimal. Mathematically, this requires that $\hat{\theta}^{\mathrm{o}} = \hat{\theta}^{\mathrm{pc}}$, so

$$\frac{2rgy}{gy(1-y) + r(1+(2g-1)y)} = \frac{\kappa rgy + gy(1-y)\tau_{\mathrm{gi}}}{2\kappa gy(1-y) + \kappa r(1+(2g-1)y)},$$

or

$$2r\left[2gy(1-y) + r(1+(2g-1)y)\right] = \left[r + (1-y)\frac{\tau_{\mathrm{gi}}}{\kappa}\right]\left[gy(1-y) + r(1+(2g-1)y)\right],$$

or

$$r(1+(2g-1)y)\left[r - (1-y)\frac{\tau_{\mathrm{gi}}}{\kappa}\right] = gy(1-y)\left[(1-y)\frac{\tau_{\mathrm{gi}}}{\kappa} - 3r\right],$$

or

$$(1-y)\frac{\tau_{\mathrm{gi}}}{\kappa}\left[r(1+(2g-1)y) + gy(1-y)\right] = r\left[3gy(1-y) + r(1+(2g-1)y)\right],$$

which is equivalent to Eq. (5.22).

Thus, the social optimum can be obtained even when a private operator manages the cooperative service. Finally, note that since the government can impose the minimum social cost, in this case Stackelberg's equilibrium is also an equilibrium in the sense of Nash. □

In the Stackelberg equilibrium studied, the objective of the government is to set its toll in such a way that the profit maximizing toll for the private operator minimizes the social cost, given this capacity split. We can actually demonstrate that by varying its own toll, the government can make any demand split profit-maximizing for the private operator and can therefore lead the system to the social optimum. Intuitively, when the private operator already has n customers, attracting the next one to its route implies reducing its toll by the sum of (i) a constant term resulting from the additional congestion cost imposed by one additional user (equal to $\frac{\delta}{2S}$) and (ii) the difference in cooperation cost between the previous critical user and the new one. This amount is independent of the current level of the toll. As it is profitable to attract this user if and only if the new amount of the toll is bigger than n times the toll reduction, the government can obtain the demand split desired simply by setting the reference toll. Note however that this mechanism can involve considerable financial transfers from the users to the private operator and the government.

In terms of comparative statics, Eq. (5.22) can be differentiated to show that the relative toll $\frac{S}{\delta N}\tau_{\text{gi}}$ increases with the capacity split y, with the capacity improvement factor g and decreases with the cost ratio r. Intuitively, if the private operator manages a greater proportion of the capacity, it is more powerful and can impose higher tolls. Similarly, if users have a lower cost ratio r, their cooperation cost is greater relatively to the congestion cost, so they are ready to pay higher tolls.

Cooperative service managed by the government, independent route by a private company

Again, this problem is symmetrical to the previous one so the details of the calculations are left to the reader. Similarly to the previous case, it is possible to obtain a social optimum if the government sets the toll to:

$$\tau_{\text{gc}} = \kappa \left(1 + r\frac{r(1+(2g-1)y) - gy((1+4g)y-1)}{2gy[r(1+(2g-1)y + gy(1-y))]}\right). \tag{5.25}$$

The relative toll $\frac{S}{\delta N}\tau_{\text{gc}}$ decreases with the capacity split y and the cost ratio r.

Numerical applications

The optimal relative tolls obtained in the Sections 5.4.2 and 5.4.2 are plotted in Fig. 5.7, together with the profit-maximizing tolls obtained in Sections 5.4.1 and 5.4.1. Note that as r increases, the tolls imposed by the government and by the private operator in Stackelberg equilibria become almost identical, which is natural since the user equilibrium coincides with the social optimum for $\kappa = 0$, i.e. $r \to \infty$ (cf. Corollary 4). Note in addition that as the private operator always sets its own toll relatively to the government's toll, the absolute value of the private toll is significantly greater for the Stackelberg equilibrium than when the government does not set any toll. The toll set on the independent route is however slightly higher than the toll on the cooperative route, in agreement with Corollary 5.

Finally, it is of practical interest to notice that the relative tolls applied in the Stackelberg

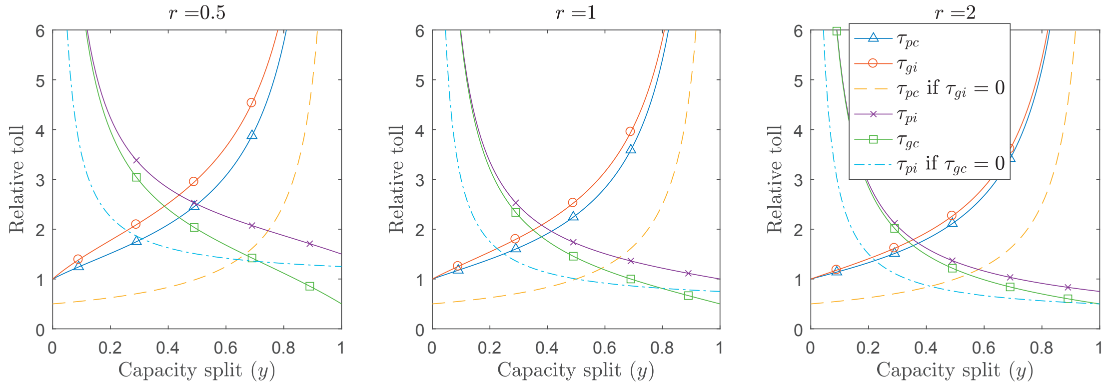

Figure 5.7 – Comparison of the tolls set by the government and by the private operator when the private operator manages the cooperative route (τ_{pc} and τ_{gi}), or when it manages the independent route (τ_{pi} and τ_{gc}), with and without a toll set by the government ($\tau_{\text{gi}} = 0$ or $\tau_{\text{gc}} = 0$). $g = 1$ for all figures.

equilibrium are systematically greater than 0.5, and often much higher[7]. In comparison, an optimal time-varying toll applied on the entire capacity as described by Vickrey (1969) allows to reach the social optimum with a maximum relative toll of exactly 0.5. Thus, a Stackelberg equilibrium with cooperation requires more money transfers and leads to a greater social cost (because of cooperation costs) than a fine toll, which makes it less acceptable and less effective. Thus, a privatization of the proposed service should be followed by governmental regulations that would lead the system close to social optimum conditions without imposing very high tolls. Again, different values of the capacity improvement factor g between 1 and 2 were tested but they led to very similar results (not included).

5.5 Conclusion

This chapter investigated the potential benefits of booking as a demand management tool in a context where some proportion of a bottleneck with constant capacity would be allocated to shared autonomous vehicles. Unlike in previous works on booking, the demand split between conventional and autonomous vehicles was made endogenous by the introduction of a cooperation cost distributed in the population. The socially optimal capacity split is then a function of this distribution. An ill-adapted capacity split may be highly counter-productive as it would create an imbalance in the network and increase the social cost.

With a socially optimal capacity split however, both the equilibrium and the socially optimal demand split reduce the social cost. The equilibrium presents several highly

[7]The assumption of an inelastic demand becomes clearly unrealistic in these cases. In practice, the tolls would be limited to more reasonable levels, but would deter some users from traveling.

desirable characteristics as it Pareto-dominates the case with no cooperation and reduces the social cost almost as much as the socially optimal demand split. While the social optimum can be decentralized with a relatively small flat toll, the independent users and some cooperative users would be worse-off (at least before redistributing the revenue of the toll). The amplitude of the social cost reduction obtained highly depends on the cooperation cost distribution and on the severity of the congestion. By offering the possibility to cooperate to the users that are naturally disposed to do so, the social cost may be reduced by about 15% for a typical peak period, even with a classic bottleneck model with constant capacity. This gain may be further multiplied by a factor of two or more if the bottleneck was previously hypercongested or if separating autonomous vehicles from conventional vehicles increases the bottleneck capacity.

Finally, the possibility of delegating the management of a route to a private operator was investigated. If prices are not restricted and if the private operator controls a significant proportion of the total capacity, the social cost may significantly increase. The socially optimal capacity split can still be obtained if the government acts as the leader of a Stackelberg competition and imposes a constant toll on the other route as well. This, however, leads to very high prices for all users. The design of suitable toll regulations should be one of the priority topics.

There are several directions in which our analysis could be extended. Besides heterogeneity in the scheduling preferences, a priority would be to model the impacts of demand elasticity and of other modes of transportation. A public transit alternative would essentially act as a source of demand elasticity for travel by autonomous or conventional vehicles. As tolling would deter some users from traveling by car, the optimum and the equilibrium might be even more similar with demand elasticity. It is unlikely however that the Stackelberg framework studied can still lead to socially optimal conditions, given the exorbitant tolls found. Nevertheless, simple profit-maximizing strategies might become more compatible with welfare maximization, as private operators should be careful not to deter too many users from using private vehicles. Finally, equity aspects should be investigated as users with low cooperation cost would benefit from this system more than others and income might be correlated with the cooperation cost. For a more general discussion on equity in transportation, we refer to Trannoy (2011).

6 Conclusion and future research

6.1 Summary of contributions

This thesis has explored several avenues related to the scheduling of trips during recurrently congested periods, generally assimilated with commutes. These avenues had received relatively little attention so far, despite their practical importance.

Chapter 3 brought the theory closer to the practice by considering stability issues with a unidirectional bottleneck model. It combined analytical insights derived for homogeneous populations in continuous time and numerical simulations of more realistic settings. On the theoretical side, we established conditions on the schedule preferences that are necessary and sufficient to guarantee the monotonicity of the utility function. This finding unveils a fundamental difference between the morning and evening commutes and explains why many morning commute simulations were reported to be unstable. Our simulations showed however that the undesirable effects of instability are largely curtailed by the presence of user heterogeneity or queue-removing pricing strategies. We concluded that real world commutes may still be reasonably approximated by their equilibrium under favorable circumstances, at least in terms of congestion cost. Other indicators such as total schedule penalties are subject to important biases.

Chapter 4 contributed to extend the field of applications of the literature by considering an entirely different type of congestion, closer to large urban areas than to isolated highways. Among other advantages, this approach allows for trip length heterogeneity. We then characterized some of the properties of the equilibrium. It is shown in particular that if users have continuously distributed characteristics, the network accumulation at equilibrium is a continuous function of time. With $\alpha - \beta - \gamma$ preferences and under certain conditions, a partial FIFO pattern emerges among early and late users. This FIFO pattern is strict only within families of users having heterogeneous trip lengths and identical preferences, or vice versa. Finally, the well-established flow-maximizing pricing strategy is proven to be sub-optimal when departure time choice is considered.

Chapter 4 also identified many differences between isotropic and unidirectional congestion. The major difference is perhaps in the shape of the cumulative outflow. Unlike the bottleneck model, an isotropic area allows for strong temporal variations in the trip completion rate, which are further amplified by trip length heterogeneity. Since users with short trips are less sensitive to congestion, they are more likely to travel at times with severe congestion. As a consequence, the decomposition of social cost into delays and schedule penalties greatly differs from the one observed with a bottleneck model.

Finally, Chapter 5 proposed an original approach of the bottleneck problem based on the optional participation to a cooperative service of slot-reservation, which benefits from dedicated right of way. Users then choose whether to participate in that cooperative program by weighing the congestion cost saving it offers against their personal cooperation cost (accounting for the inconvenience associated to reserving a slot and traveling at that time). We explained how such a scheme would be advantageously implemented together with car-sharing and autonomous vehicles, mostly for reasons of capacity gains, simplified enforcement, and cooperation cost reduction.

The consequences on welfare were then evaluated depending on the regime (laissez-faire, welfare- or profit-maximizing) and on the capacity split. Provided some users have low cooperation costs, there exists a capacity split such that the laissez-faire equilibrium Pareto-dominates the case without a cooperative service. The demand split between the two services can be made socially optimal by applying a constant toll on the non-cooperative lanes, but it only marginally improves the benefits of cooperation. Although the Pareto-improvement result may not hold for every single user in the case of richer heterogeneity, it remains a key advantage of letting users choose. Profit-maximizing strategies however turn out to be hardly compatible with welfare maximization.

6.2 Practical implications

The thesis has important practical implications concerning cost-benefit analyses and the design of congestion alleviating measures. Cost-benefit analyses require comparing welfare in different situations. Welfare however includes many components, which are sometimes difficult to estimate. When considering road congestion, travel time and schedule penalties represent two major components of the congestion cost. Since schedule penalties cannot be directly measured, one may be tempted to infer them from the measured travel time, by applying a proportionality factor derived theoretically. This thesis reveals that doing so would typically entail several biases. Chapter 3 showed that instability is rather common with unidirectional flows and that it typically results in a larger proportion of schedule penalties in the total congestion cost. On the contrary, Chapter 4 revealed that the relative weight of schedule penalties is typically smaller with isotropic models, although it strongly depends on congestion severity. In a different vein, Chapter 5 enhanced another aspect of cost-benefit analyses by recognizing that schemes based on mobility credits would entail an important cooperation cost.

This thesis also provides useful guidance for the design of congestion alleviating measures. The measure that is currently the most widely supported by the academic community is certainly congestion pricing. With unidirectional congestion, the socially optimal toll is known for different types of schedule preferences and heterogeneity. Yet, it had rarely been studied from a stability point of view. Chapter 3 shows that besides corresponding to a more efficient equilibrium, pricing also greatly stabilizes the system and therefore reduces travel time variability. Chapter 3 also suggests that when the total demand is not precisely known, it is preferable to overestimate the population size when designing the toll than to underestimate it. With isotropic congestion however, we still know relatively little concerning the socially optimal toll. Although it was not our priority, Chapter 4 contributed to the body of knowledge by showing that with heterogeneous trip lengths, the social optimum may exhibit hypercongestion.

Finally, the scheme proposed in Chapter 5 would fit particularly well with the currently emerging trends toward vehicle automation and car-sharing. Although future mobility services will probably not follow our framework precisely, it is important to raise awareness about the cost of competition in departure time choice and to show that smart designs can significantly reduce that cost.

6.3 Future research

Many avenues for future research have already been outlined in the conclusions of the Chapters 3, 4 and 5. This section provides further details about two specific directions that are particularly urgent in our opinion.

6.3.1 Stability in isotropic environments

One of the next research priorities is to analyze the stability of departure time choice with multi-directional flows. Intuitively, directionality matters because it affects the structure of externalities. Indeed, uni-directional congestion mechanisms are normally expected to satisfy the "causality" criterion, which requires that externalities only propagate forward in time (i.e. users can only delay those departing later). Such a constraint does not apply to multi-directional flows, where users can even mutually delay each other. Too see this, consider for instance a two-way road, where intersections are such that additional vehicles cause delays in both directions (for instance because of left turns, or roundabouts). Two users traveling in opposite directions who cross in the middle of the road would afterwards be delayed by the additional congestion the other user left behind her.

Another important characteristic of multi-directional networks is hypercongestion. Although some unidirectional models theoretically allow for hypercongestion (e.g. via a link travel time function (Mahmassani and Herman, 1984) or via a bottleneck where capacity depends on queue length (Yang and Huang, 1997)), it is not clear whether such phenomena exist in real-world unidirectional settings. With multiple directions however, hypercongestion is very natural. If we consider for instance a ring road, the average flow on the road first increases with the number of vehicles, then decreases and eventually reaches zero at the jam accumulation. Such situations also naturally emerge in grids, when queues in one direction create queues in another, which create queues in still another, etc. until the last queue blocks the first. Gridlock situations are admittedly rare in practice, but hypercongestion has been repeatedly observed at the zone level (Geroliminis and Daganzo, 2008; Loder et al., 2017). The challenge is now to determine how these specificities of multidirectional set-ups influence stability. The analysis of the literature provides some valuable hints. de Palma (2000) for instance reported that "the adjustment processes converge under much milder conditions (on the values of the parameters) as the size of the network increases". Yet, these effects need to be better explained and documented.

The isotropic model used in Chapter 4 seems to be a good candidate mechanism to explain these effects because it exhibits the desirable features of real world multi-directional settings (hypercongestion and externalities propagating also backward in time), while remaining quite general and independent of the exact network topology. Even with this simplification however, the equilibrium turns out to be significantly more difficult to analyze than with a unidirectional bottleneck. One major difficulty is the absence of a general and strict relation between the orders of departures and arrivals. Pure LIFO and FIFO sorting patterns emerge with some types of schedule preferences (as shown in Fosgerau (2015) and in Chapter 4), but the real world is likely to exhibit a combination of both LIFO and FIFO regimes.

Consider for instance the situation illustrated in Fig. 6.1, obtained after running a simulation for 200 days with an isotropic region and 4000 discrete agents having schedule preferences $u_o(t) = \alpha$ and $u_d(t) = \frac{2+\gamma-\beta}{2} + \arctan(4(t-t^*))\frac{\gamma+\beta}{\pi}$, with $\alpha = 1$, $\beta = 0.5$, $\gamma = 2$.[1] Trip lengths l were uniformly between 0 and 1 and users were divided (independently of l) in 10 groups of equal size, each having a different value of t^* in the set $\{8.05, 8.15, ...8.95\}$. The ratio of demand to capacity was set to 4, such that at capacity, the peak period would last 4 h ($N/n_{\mathrm{cr}} \simeq 3.56$). If we focus on the group with the largest t^* (in dark red), users with short trips (approximately smaller than 0.6 distance units) follow a LIFO pattern (those with longer trips start earlier and arrive later), while those with longer trips are sorted in a FIFO pattern (those with longer trips start later and arrive later). The LIFO pattern naturally prevails in the middle of the peak, when schedule preferences vary the most rapidly and speed reaches a plateau.

Fig. 6.1 also illustrates another complexity of the equilibrium: the yellow, orange and red groups (with intermediate t^*) not only combine FIFO and LIFO patterns, but they

[1] The code used to generate this figure is available at https://github.com/raplam/departureTimeChoice.

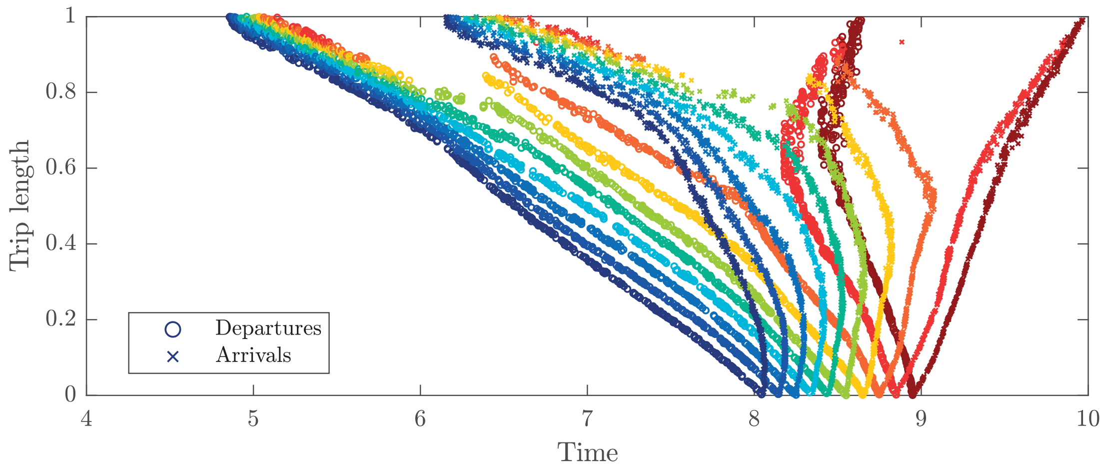

Figure 6.1 – Departure and arrival times with uniformly distributed trip lengths and 10 values of t^* (smooth schedule preferences)

also include two disconnected FIFO periods. If we consider the departure or arrival times as functions of trip length for one of these groups only, we see a discontinuity, occurring for trip lengths around 0.9 (slightly smaller for the yellow group, larger for the red group). These two FIFO periods can be quite far away: for the red group, one occurs during the onset and the other during the offset. These sorting patterns need to be better understood before analyzing the stability properties of the equilibrium because they strongly influence the structure of externalities.

6.3.2 Other types of road usage separations

The interaction between road usage separation and departure time choice represents another important research avenue that has been largely overlooked. In Chapter 5 we considered the case of a permanent capacity split. In theory, this idea could be extended to deal with other types of congestion mechanisms. In practice however, the situation is quite different in multi-directional set-ups because the capacity cannot easily be divided. Some streets could admittedly be reserved (similarly to what is done for public transit in many cities) but it would be extremely difficult to create a fully connected subnetwork that provides access to all parts of the city and yet does not interact with general traffic. Thus, it seems difficult to fully segregate vehicles in urban environments, at least spatially.

An interesting alternative would be time segregation. Segregation could either be fully time-based (the capacity is allocated to only one class of vehicles at a time), or hybrid (spatial but time-dependent), with temporary reserved lanes. This idea is very similar to the concept of *fast lane*, proposed by Fosgerau (2011). Note that time-dependent capacity allocation can reduce queuing, even without cooperation. Indeed, as the amount of queuing per passenger reflects the variability in schedule penalty in the allocated

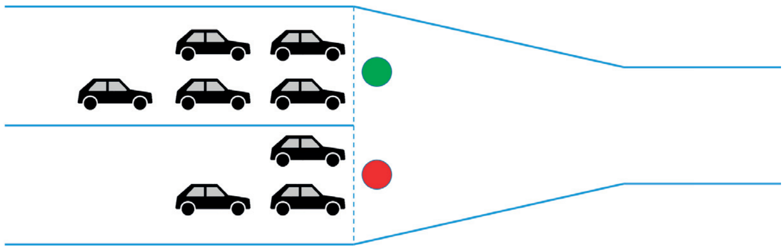

Figure 6.2 – Illustration of a bottleneck divided between two populations

period, the queueing times can be significantly reduced if the allocated period only covers a small range of schedule penalties. Fosgerau (2011) showed that with homogeneous users, such a scheme is equivalent to a coarse toll and produces a Pareto-improvement.

However, it is not clear yet to which extent "fast lanes" might be implemented in a fair and efficient manner in the real world. With homogeneous users, the optimum would require as many slots as users. Yet, such a solution can hardly be applied with heterogeneous users for several reasons. First, users' scheduling preferences are generally unknown and ignoring them could be counterproductive and increase the social cost. Second, users with different access rights arriving at the same time should go to different queues (see Fig. 6.2 for an example with two populations), thereby limiting the number of user classes that can be considered in practice. Third, creating almost individual categories also raises equity and privacy-related issues commonly associated with non-anonymous pricing.

In a joint work presented at the 2017 edition of the hEART conference (Lamotte et al., 2017), we considered only two categories based on vehicle type (autonomous and conventional vehicles) but introduced discrete intra-group heterogeneity in value of time and/or scheduling preferences. We showed that in such a context, the socially optimal access restrictions apply to the entire bottleneck capacity, such that the road capacity is always fully allocated to one vehicle type at a time (bang-bang control). The number of time windows allocated per group depends on the amount of heterogeneity in schedule preferences. With strong heterogeneity, it might be socially optimal to allocate multiple non-contiguous time periods to the same category of vehicles.

This was only a preliminary investigation in a very simple setting. In practice, the time segregation should probably not be strict, to allow for errors in planning or emergency trips. One way to circumvent this issue would be to keep some capacity open to all at all times. Another may be to allow vehicles to travel out of their allocated period by paying a fine.

A An approach of sorting with the MFD relying only on exogenous assumptions

This Appendix details an approach to sorting that only relies on exogenous assumptions and represents an alternative to the approach presented in Chapter 4. The main assumption is that desired arrival times are distributed in a compact subset of the time axis. Then, we show that accumulation cannot be decreasing before this time interval and cannot be increasing after it (Proposition 14). We refer to this result as the 'single predominant peak". Accumulation can be constant during some time period only if nobody departs or arrives during that time period (Lemma 6). Altogether, these results form a basis to derive a sorting result that is valid among users traveling exclusively out of the interval where desired arrival times are distributed (Proposition 15).

A.0.1 Preliminary results

While the concept is quite simple, the proofs are made more tedious by the utilization of a continuum of users. A drawback of this approach is that even with the full knowledge of the cumulative numbers of departures and arrivals $D(t)$ and $A(t)$, the departure and arrival *rate* functions are only determined up to a set of measure zero, so that one cannot be sure that there are actually users departing or arriving at any specific time. Fortunately, Proposition 13 allows us to circumvent this issue. Before stating it, we need to introduce some notations and a definition.

Let $\mathcal{Z} = \{(\alpha, \beta, \gamma, l, t^*) \in \left(\mathbb{R}^{*+}\right)^4 \times \mathbb{R}, \beta < \alpha\}$ denote the set of admissible characteristics for users with $\alpha - \beta - \gamma$ preferences. The condition $\beta < \alpha$ ensures that users arriving early do not remain in their vehicle until their desired arrival time t^*. Let also $C(z, t_a)$ denote the function from $\mathcal{Z} \times \mathbb{R}$ to $\mathbb{R}$ that maps a user with characteristics z and arrival time t_a to her experienced cost. Note that since the travel time function $\tau(t_\mathrm{a}, l)$ is continuous, the cost function $C(z, t_\mathrm{a})$ is continuous as well.

Definition 5. *A function $h : \mathbb{R} \to \mathbb{R}$ is said to be locally constant at $\tilde{t} \in \mathbb{R}$ if there exists $\epsilon > 0$, such that $h(t)$ is constant on $[\tilde{t} - \epsilon, \tilde{t} + \epsilon]$.*

Proposition 13. *Consider a system governed by a speed-MFD and a population of users whose characteristics are continuously distributed on a compact subspace Z of $\mathcal{Z}$. Consider an equilibrium realization of the morning commute. If the cumulative number of departures (resp. arrivals) is not locally constant at some $t \in \mathbb{R}$, then t is an equilibrium departure (resp. arrival) time for some user with characteristics $z \in Z$.*

Proof. If $D(t)$ (resp. $A(t)$) is not locally constant, one can find for all $n \in \mathbb{N}$ a user that arrives (at equilibrium) in the interval $t_n \in [t - 1/n, t + 1/n]$. Denote z_n and t_n her characteristics and arrival time. Since all terms of $(z_n)_{n\in\mathbb{N}}$ are in the compact space Z, the Bolzano-Weierstrass theorem ensures that there exists a subsequence of $(z_n)_{n\in\mathbb{N}}$ that converges. Let z denote the limit of this subsequence. Because Z is closed, z belongs to Z. The continuity of the function $C(z, t)$ ensures that t is an equilibrium departure (resp. arrival) time for a user characterized by z. □

Note that there might not be any user with characteristics z in the population but it does not matter. Note also that the statement "D is not locally constant at t" is equivalent to "there exists $m \in (0, N]$, such that $t = \inf\{u \in \mathbb{R}, D(u) \geq m\}$". The time t is often abusively referred to as the departure time of the m^{th} user to depart, although as we have discussed, there might not be any user in the population who departs at t.

A.0.2 The single predominant peak result

We are now ready to define our assumptions and the single predominant peak result.

Assumption 5 (Distributed $\alpha - \beta - \gamma$ preferences)**.**

1. *All users have $\alpha - \beta - \gamma$ preferences, with coefficients belonging to a compact subspace Z of $\mathcal{Z}$. In particular, t^* belongs to an interval $[\underline{t^*}, \bar{t^*}]$.*
2. *Desired arrival times (t^*) are continuously distributed.*
3. *$\frac{\beta}{\alpha}$ and $\frac{\gamma}{\alpha}$ are bounded away from 0 for all users.*
4. *Trip length and trip length conditioned on the desired arrival time are continuously distributed variables.*

Note that Assumption 5 implies Assumption 3. The main additional requirements of Assumption 5 are that all users have $\alpha - \beta - \gamma$ preferences and that the users' characteristics are distributed on a compact space. Requirement 3 is purely technical and quite reasonable.

Proposition 14 (Single predominant peak)**.** *Consider a system governed by a speed-MFD and a population satisfying Assumption 5. At equilibrium, accumulation is weakly increasing for all $t < \underline{t^*}$ and weakly decreasing for all $t > \bar{t^*}$.*

Proof of Proposition 14. Requirement 3 of Assumption 5 allows us to define $\epsilon > 0$ such that $\frac{\alpha}{\alpha-\beta} > 1+\epsilon$ and $\frac{\gamma+\alpha}{\alpha} > 1+\epsilon$ for all users. Proposition 2 implies that between the departure and the arrival of an early (resp. late) user, speed must be divided (resp. multiplied) by at least $1+\epsilon$. Then, the proof proceeds by contradiction: we demonstrate that an accumulation decrease before $\underline{t}^*$ or an increase after $\bar{t}^*$ implies the existence of a sequence of users whose speed at departure or arrival tends to infinity, which is not possible given the speed-MFD considered.

We detail the proof that $n(t)$ is weakly increasing for $t < \underline{t}^*$ and then explain how it can be adapted to show that $n(t)$ is weakly decreasing for $t > \bar{t}^*$. Suppose there exist $t_1 < t_2 < \underline{t}^*$ such that $n(t_1) > n(t_2)$ and let $u \in (v(t_1), v(t_2)) \cap (v(t_2)/(1+\epsilon), v(t_2))$. Since Assumption 5 implies Assumption 3, Proposition 1 guarantees the continuity of accumulation and speed. The intermediate value theorem ensures the existence of a time $t_1'' \in [t_1, t_2]$ such that $v(t_1'') = u$. Define then $A = \{t \in [t_1'', t_2], n(t) = n(t_1'')\}$. A is clearly bounded, not empty and since $n(t)$ is continuous, A is closed and admits a largest element, that we denote t_1'. Similarly, let t_2' be the smallest element of $\{t \in [t_1', t_2], n(t) = n(t_2)\}$. The interval $[t_1', t_2']$ constructed is such that for all t inside it, $n(t)$ belongs to $[n(t_2'), n(t_1')]$.

The definition of t_1' also implies that for all $\xi > 0$, there exists $t \in [t_1', t_1' + \xi]$ such that $n(t) < n(t_1')$, hence $A(t) > A(t_1')$, i.e. A is not locally constant at t_1'. By applying Proposition 13, there exists an element $z = (\alpha, \beta, \gamma, l, t^*) \in \mathcal{Z}$ that would be at equilibrium by arriving at t_1'. Hence:

$$\begin{aligned}
C(z, t_1') \leq C(z, t_2') &\Leftrightarrow \alpha\tau(t_1', l) + \beta(t^* - t_1') \leq \alpha\tau(t_2', l) + \beta(t^* - t_2') \\
&\Leftrightarrow \alpha(\tau(t_2', l) - \tau(t_1', l)) \geq \beta(t_2' - t_1') \\
&\Leftrightarrow \alpha \int_{t_1'}^{t_2'} \frac{\partial \tau}{\partial u}(u, l)\, \mathrm{d}u \geq \beta(t_2' - t_1') \\
&\Leftrightarrow \alpha \int_{t_1'}^{t_2'} 1 - \frac{v(u)}{v(t_\mathrm{d}(u, l))}\, \mathrm{d}u \geq \beta(t_2' - t_1') \quad \text{(using Eq. (4.2))} \\
&\Leftrightarrow \int_{t_1'}^{t_2'} f(u)\, \mathrm{d}u \geq 0, \qquad \text{(A.1)}
\end{aligned}$$

where $f(t) = 1 - \frac{\beta}{\alpha} - \frac{v(t)}{v(t_\mathrm{d}(t,l))}$.

Proposition 2 implies that $v(t_1')/v(t_\mathrm{d}(t_1', l)) = 1 - \beta/\alpha$, i.e. $f(t_1') = 0$. Since $v(u) \geq v(t_1')$ for all $u \in [t_1', t_2']$ and $v(u) > v(t_1')$ in a neighborhood around t_2', then $v(t)$ also has to be strictly greater than $v(t_\mathrm{d}(t_1', l))$ for some $t \in [t_\mathrm{d}(t_1', l), t_\mathrm{d}(t_2', l)]$ for Eq. (A.1) to be satisfied. We ignore a priori whether t_1' is greater or smaller than $t_\mathrm{d}(t_2', l)$ but we do know that for all $t \in [t_1', t_2']$, $v(t) \leq v(t_2') < (1+\epsilon)v(t_1') < v(t_\mathrm{d}(t_1', l))$. Hence, there exists a time $s_2 \in (t_\mathrm{d}(t_1', l), t_1')$ such that $n(s_2) < n(t_\mathrm{d}(t_1', l))$.

To conclude, one could repeat the same process iteratively with the pair $(s_1 = t_\mathrm{d}(t_1', l), s_2)$ replacing the pair (t_1, t_2). Since $v(t_\mathrm{d}(t_1', l)) > (1+\epsilon)v(t_1') > v(t_1)$, the speed of the left-hand term of the pair tends to infinity. This is not possible with the speed-MFD considered and we can reject the initial assumption.

The reasoning to prove that accumulation cannot be increasing after $\bar{t^*}$ is very similar. Consider an hypothetical pair $\bar{t^*} < t_1 < t_2$, such that $n(t_1) < n(t_2)$. By following the same process as above, one can find an interval $[t'_1, t'_2]$ such that for all t inside it, $n(t)$ belongs to $[n(t'_1), n(t'_2)]$, $n(t_1) < n(t'_1)$, $v(t'_2) > v(t'_1)/(1+\epsilon)$ and $D(t)$ is not locally constant at t'_2. We then need to define the function $t_{\mathrm{a}}(t, l)$ which associates to a departure time t and a trip length l the corresponding arrival time. Clearly, $t_{\mathrm{a}}(t_{\mathrm{d}}(t, l), l) = t$ holds for all $t \in \mathbb{R}$. Then, $C(z, t_{\mathrm{a}}(t'_1, l)) \geq C(z, t_{\mathrm{a}}(t'_2, l))$ is equivalent to $\int_{t_{\mathrm{a}}(t'_1,l)}^{t_{\mathrm{a}}(t'_2,l)} f(u)\,\mathrm{d}u \leq 0$, where $f(t) = 1 + \frac{\gamma}{\alpha} - \frac{v(u)}{v(t_{\mathrm{d}}(u,l))}$. Again $f(t_{\mathrm{a}}(t'_2, l)) = 0$ and since $v(t_{\mathrm{d}}(u, l)) \geq v(t'_2)$ for all $u \in [t_{\mathrm{a}}(t'_1, l), t_{\mathrm{a}}(t'_2, l)]$ and $v(t_{\mathrm{d}}(u, l)) > v(t'_2)$ in a neighborhood around $t_{\mathrm{a}}(t'_1, l)$, then $v(t)$ also has to be strictly greater than $v(t_{\mathrm{a}}(t'_2, l))$ for some $t \in [t_{\mathrm{a}}(t'_1, l), t_{\mathrm{a}}(t'_2, l)]$. We ignore a priori whether t'_2 is greater or smaller than $t_{\mathrm{a}}(t'_1, l)$ but we do know that for all $t \in [t'_1, t'_2], v(t) \leq v(t'_1) < (1+\epsilon)v(t'_2) < v(t_{\mathrm{a}}(t'_2, l))$. Hence, there exists a time $s_1 \in (t'_2, t_{\mathrm{a}}(t'_2, l))$ such that $n(s_1) < n(t_{\mathrm{a}}(t'_2, l))$. Note finally that $v(t_{\mathrm{a}}(t'_2, l)) > (1+\epsilon)v(t'_2) > v(t_2)$. The same process can be repeated iteratively with the pair $(s_1, s_2 = t_{\mathrm{a}}(t'_2, l))$ replacing (t_1, t_2) and the speed of the right-hand term tends to infinity. □

Proposition 14 states essentially that out of the range where desired arrival times are distributed, accumulation does not exhibit local extrema: it is weakly increasing before the lower bound of desired arrival times and weakly decreasing after the upper bound. Intuitively, if desired arrival time is close to being homogeneous, then accumulation would essentially consist of single peak. While this supports the idea that single congestion peaks should be quite common as well with MFD dynamics, this result is weaker than the result of Smith (1984a) and Daganzo (1985) for the bottleneck with constant capacity.

A.0.3 Sorting result

Lemma 6. *Consider a system governed by a speed-MFD and a population satisfying Assumption 5. Let* $t \in \mathbb{R}\backslash[\underline{t^*}, \bar{t^*}]$.[1] *If accumulation is locally constant at time* t *at equilibrium, then the cumulative departure and cumulative arrival functions are also locally constant at* t.

Proof of Lemma 6. This proof follows the same approach as the proof of Proposition 14: we show by contradiction that the existence of a time t that violates Lemma 6 implies the existence of a sequence of speeds that tends to infinity. We only prove the result for the early case as the late case is very similar.

[1] The notation $A\backslash B$ denotes the complement of B with respect to A, i.e. the set of elements in A but not in B.

Let us first define $\epsilon > 0$ such that $\frac{\alpha}{\alpha-\beta} > 1+\epsilon$ and $\frac{\gamma+\alpha}{\alpha} > 1+\epsilon$ for all users. Assume that there exists $\tilde{t} < \underline{t}^*$ and $\xi > 0$ such that accumulation is constant on $[\tilde{t}-\xi, \tilde{t}+\xi]$. Take ξ small enough to ensure that $\tilde{t}+\xi < \underline{t}^*$. Assume (by contradiction) that the cumulative departure function D or the cumulative arrival function A is not locally constant at $\tilde{t}$. Since $n(t) = D(t) - A(t)$ for all time t, this actually implies that both D and A are not locally constant at $\tilde{t}$.

For any interval I, let $D_I(t)$ denote the cumulative departure function of the users arriving during I. Let M denote the measure of users arriving during $[\tilde{t}-\xi/2, \tilde{t}+\xi/2]$. Choose any $z \in (0, M]$. For any small neighborhood around the time $\tilde{t}' = \min\{t \in \mathbb{R}, D_{[\tilde{t}-\xi/2, \tilde{t}+\xi/2]} = z\}$, there is a positive measure of users that departed in this neighborhood and arrived in $[\tilde{t}-\xi/2, \tilde{t}+\xi/2]$.

We now show that accumulation is locally constant at $\tilde{t}'$. Let $\xi' = \xi v\left(\tilde{t}\right)/v_{\mathrm{f}}$, where v_{f} is the free-flow speed. Let l and t_{a} denote the trip length and arrival time of some user departing in $[\tilde{t}'-\xi'/3, \tilde{t}'+\xi'/3]$ and arriving in $[\tilde{t}-\xi/2, \tilde{t}+\xi/2]$. On one hand, user equilibrium requires that for all $t \in [\tilde{t}-\xi, \tilde{t}+\xi]$, $C_l(t_{\mathrm{a}}) \leq C_l(t)$. By applying the same reasoning as in the proof of Proposition 14, this is equivalent to $\int_{t_{\mathrm{a}}}^{t} f(u)\,\mathrm{d}u \geq 0$, where the function $f(t)$ is defined as in Eq. (A.1). On the other hand, since $v(u)$ is constant between t_{a} and t and $v(t_{\mathrm{d}}(u,l))$ is weakly decreasing (Proposition 14), $f(t)$ is also weakly decreasing between t_{a} and t. Since this is valid for all $t \in [\tilde{t}-\xi, \tilde{t}+\xi]$ and Proposition 2 requires that $f(t_{\mathrm{a}}) = 0$, $f(t)$ is non-negative on $[\tilde{t}-\xi, t_{\mathrm{a}}]$ and non-positive on $[t_{\mathrm{a}}, \tilde{t}+\xi]$.

Together, these two results imply that $f(t) = 0$ for all $t \in [\tilde{t}-\xi, \tilde{t}+\xi]$, i.e. speed is constant on $[t_{\mathrm{d}}(\tilde{t}-\xi, l), t_{\mathrm{d}}(\tilde{t}+\xi, l)]$. By construction, $\tilde{t}'$ belongs to the interior of $[t_{\mathrm{d}}(\tilde{t}-\xi, l), t_{\mathrm{d}}(\tilde{t}+\xi, l)]$. Indeed, as $\int_{t_{\mathrm{d}}(\tilde{t}-\xi,l)}^{\tilde{t}-\xi} v(u)\,\mathrm{d}u$ and $\int_{t_{\mathrm{d}}(t_{\mathrm{a}},l)}^{t_{\mathrm{a}}} v(u)\,\mathrm{d}u$ must both be equal to l, $\int_{t_{\mathrm{d}}(\tilde{t}-\xi,l)}^{t_{\mathrm{d}}(t_{\mathrm{a}},l)} v(u)\,\mathrm{d}u = \int_{\tilde{t}-\xi}^{t_{\mathrm{a}}} v(u)\,\mathrm{d}u$. Since $\int_{t_{\mathrm{d}}(\tilde{t}-\xi,l)}^{t_{\mathrm{d}}(t_{\mathrm{a}},l)} v(u)\,\mathrm{d}u \leq \left(t_{\mathrm{d}}(t_{\mathrm{a}},l) - t_{\mathrm{d}}(\tilde{t}-\xi,l)\right) v_{\mathrm{f}}$ and $\int_{\tilde{t}-\xi}^{t_{\mathrm{a}}} v(u)\,\mathrm{d}u = \left(t_{\mathrm{a}} - (\tilde{t}-\xi)\right) v(\tilde{t}) \geq \frac{\xi}{2} v(\tilde{t})$, we obtain $t_{\mathrm{d}}(\tilde{t}-\xi, l) \leq t_{\mathrm{d}}(t_{\mathrm{a}}, l) - \frac{\xi v(\tilde{t})}{2 v_{\mathrm{f}}}$. As $t_{\mathrm{d}}(t_{\mathrm{a}}, l) \leq \tilde{t}' + \xi'/3$, we have shown that $t_{\mathrm{d}}(\tilde{t}-\xi, l) \leq \tilde{t}' + \xi'/3 - \xi'/2$, i.e. $t_{\mathrm{d}}(\tilde{t}-\xi, l) < \tilde{t}'$. Similarly, $t_{\mathrm{d}}(\tilde{t}+\xi, l) \geq t_{\mathrm{d}}(t_{\mathrm{a}}, l) + \frac{\xi v(\tilde{t})}{2 v_{\mathrm{f}}} \geq \tilde{t}' - \xi'/3 + \xi'/2 > \tilde{t}'$. Hence accumulation is locally constant at $\tilde{t}'$.

To conclude, one could repeat the same process iteratively, going back in time from $\tilde{t}$ to $\tilde{t}'$, then from $\tilde{t}'$ to $\tilde{t}^{(2)}$, etc. Since there are early users that depart close to $\tilde{t}'$ and arrive close to $\tilde{t}$ and accumulation is locally constant in the neighborhoods of $\tilde{t}'$ and $\tilde{t}$, Proposition 2 implies that speed is multiplied by a factor of at least $(1+\epsilon)$ at every step, and hence tends to infinity.

The proof for the case $\tilde{t} > \bar{t}^*$ is extremely similar, except that one should go forward in time, as in the proof of Proposition 14. □

Proposition 15. *Consider a system governed by a speed-MFD and a population of users satisfying Assumption 5. Consider two users with schedule preferences* $(\alpha_1, \beta_1, \gamma_1)$ *and*

$(\alpha_2, \beta_2, \gamma_2)$ *arriving before* $\underline{t}^*$ *(resp. after* $\bar{t}^*$*) at equilibrium and departing at times where the cumulative departure function is not locally constant. If* $\frac{\beta_1}{\alpha_1} = \frac{\beta_2}{\alpha_2}$ *(resp.* $\frac{\gamma_1}{\alpha_1} = \frac{\gamma_2}{\alpha_2}$*), then the user that arrives the earliest (resp. latest) has the longest trip length.*

The proof is very similar to the proof of Proposition 3 but makes use of the Proposition 14 and Lemma 6 to conclude from Eq. (4.8).

Bibliography

Anderson, S. P., de Palma, A., Thisse, J.-F., 1997. Privatization and efficiency in a differentiated industry. European Economic Review 41 (9), 1635–1654.

Arnott, R., 2013. A bathtub model of downtown traffic congestion. Journal of Urban Economics 76, 110–121.

Arnott, R., de Palma, A., Lindsey, R., 1988. Schedule delay and departure time decisions with heterogeneous commuters. Transportation Research Record 1197, 56–67.

Arnott, R., de Palma, A., Lindsey, R., 1990. Economics of a bottleneck. Journal of urban Economics 27 (1), 111–130.

Arnott, R., Kokoza, A., Naji, M., 2016. Equilibrium traffic dynamics in a bathtub model: A special case. Economics of Transportation 7, 38–52.

Bhat, C. R., 1995. A heteroscedastic extreme value model of intercity travel mode choice. Transportation Research Part B: Methodological 29 (6), 471–483.

Braid, R. M., 1996. Peak-load pricing of a transportation route with an unpriced substitute. Journal of Urban Economics 40 (2), 179–197.

Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. Transportation Research Record: Journal of the Transportation Research Board (2124), 127–136.

Carrion, C., Levinson, D., 2012. Value of travel time reliability: A review of current evidence. Transportation Research Part A: Policy and Practice 46 (4), 720–741.

Daganzo, C. F., 1985. The uniqueness of a time-dependent equilibrium distribution of arrivals at a single bottleneck. Transportation science 19 (1), 29–37.

Daganzo, C. F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transportation Research Part B: Methodological 41 (1), 49–62.

Daganzo, C. F., Lehe, L. J., 2015. Distance-dependent congestion pricing for downtown zones. Transportation Research Part B: Methodological 75 (0), 89–99.

De Feijter, R., Evers, J. J., Lodewijks, G., 2004. Improving travel-time reliability by the use of trip booking. Intelligent Transportation Systems, IEEE Transactions on 5 (4), 288–292.

de Palma, A., 2000. Solution and stability for a simple dynamic bottleneck model. In: Filar, J. A., Gaitsgory, V., Mizukami, K. (Eds.), Advances in Dynamic Games and Applications. Springer, pp. 405–425.

de Palma, A., Lindsey, R., 2000. Private toll roads: Competition under various ownership regimes. The Annals of Regional Science 34 (1), 13–35.

de Palma, A., Lindsey, R., 2002. Comparison of morning and evening commutes in the vickrey bottleneck model. Transportation Research Record: Journal of the Transportation Research Board 1807, 26–33.

de Palma, A., Marchal, F., 2002. Real cases applications of the fully dynamic metropolis tool-box: An advocacy for large-scale mesoscopic transportation systems. Networks and Spatial Economics 2 (4), 347–369.

de Palma, A., Marchal, F., Nesterov, Y., 1997. Metropolis: Modular system for dynamic traffic simulation. Transportation Research Record: Journal of the Transportation Research Board 1607, 178–184.

Drezner, T., Drezner, Z., Kalczynski, P., 2015. A leader-follower model for discrete competitive facility location. Computers & Operations Research 64, 51–59.

Dupuis, P., Nagurney, A., 1993. Dynamical systems and variational inequalities. Annals of Operations Research 44 (1), 7–42.

Edara, P., Teodorović, D., 2008. Model of an advance-booking system for highway trips. Transportation Research Part C: Emerging Technologies 16 (1), 36–53.

Fajardo, D., Au, T.-C., Waller, S. T., Stone, P., Yang, C. Y. D., 2012. Automated intersection control: Performance of a future innovation versus current traffic signal control. Transportation Research Record: Journal of the Transportation Research Board 2259, 223–232.

Fargier, P.-H., 1983. Effects of the choice of departure time on road traffic congestion. theoretical approach. In: Proceedings of the 8th International Symposium on Transportation and Traffic Theory. pp. 223–263.

Fosgerau, M., 2009. The marginal social cost of headway for a scheduled service. Transportation Research Part B: Methodological 43 (8-9), 813–820.

Fosgerau, M., 2010. On the relation between the mean and variance of delay in dynamic queues with random capacity and demand. Journal of Economic Dynamics and Control 34 (4), 598–603.

Fosgerau, M., 2011. How a fast lane may replace a congestion toll. Transportation Research Part B: Methodological 45 (6), 845–851.

Fosgerau, M., 2015. Congestion in the bathtub. Economics of Transportation 4 (4), 241–255.

Fosgerau, M., de Palma, A., 2012. Congestion in a city with a central bottleneck. Journal of Urban Economics 71 (3), 269–277.

Fosgerau, M., Karlström, A., 2010. The value of reliability. Transportation Research Part B: Methodological 44 (1), 38–49.

Fosgerau, M., Small, K. A., 2013. Hypercongestion in downtown metropolis. Journal of Urban Economics 76 (0), 122–134.

Geroliminis, N., Daganzo, C. F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B: Methodological 42 (9), 759–770.

Geroliminis, N., Levinson, D. M., 2009. Cordon pricing consistent with the physics of overcrowding. In: Lam, W. H. K., Wong, S. C., Lo, H. K. (Eds.), Transportation and Traffic Theory 2009: Golden Jubilee. Springer US, pp. 219–240.

Godfrey, J. W., 1969. The mechanism of a road network. Traffic Engineering and Control 11 (7), 323–327.

Guo, R.-Y., Yang, H., Huang, H.-J., 2018. Are we really solving the dynamic traffic equilibrium problem with a departure time choice? Transportation Science 52 (3), 603–620.

Hall, J. D., 2018. Pareto improvements from lexus lanes: The effects of pricing a portion of the lanes on congested highways. Journal of Public Economics 158, 113–125.

Henderson, J., 1981. The economics of staggered work hours. Journal of Urban Economics 9 (3), 349–364.

Hendrickson, C., Plank, E., 1984. The flexibility of departure times for work trips. Transportation Research Part A: General 18 (1), 25–36.

Herman, R., Prigogine, I., 1979. A two-fluid approach to town traffic. Science 204 (4389), 148–151.

Hjorth, K., Börjesson, M., Engelson, L., Fosgerau, M., 2015. Estimating exponential scheduling preferences. Transportation Research Part B: Methodological 81, Part 1, 230–251.

Hofbauer, J., Sandholm, W. H., 2009. Stable games and their dynamics. Journal of Economic Theory 144 (4), 1665–1693.

Hurwicz, L., Reiter, S., 2006. Designing economic mechanisms. Cambridge University Press.

International Transport Forum, 2015a. Automated and autonomous driving-regulation under uncertainty. Tech. rep., OECD.

International Transport Forum, 2015b. Urban mobility system upgrade: How shared self-driving cars could change city traffic. Tech. rep., OECD.

Iryo, T., 2008. An analysis of instability in a departure time choice problem. Journal of Advanced Transportation 42 (3), 333–358.

Iryo, T., Yoshii, T., 2007. Equivalent optimization problem for finding equilibrium in the bottleneck model with departure time choices. In: 4th IMA International Conference on Mathematics in Transport. pp. 231–244.

Jamson, A. H., Merat, N., Carsten, O. M., Lai, F. C., 2013. Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. Transportation Research Part C: Emerging Technologies 30, 116–125.

Ji, Y., Luo, J., Geroliminis, N., 2014. Empirical observations of congestion propagation and dynamic partitioning with probe data for large-scale systems. Transportation Research Record: Journal of the Transportation Research Board 2422, 1–11.

Kouvelas, A., Saeedmanesh, M., Geroliminis, N., 2017. Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. Transportation Research Part B: Methodological 96, 26–45.

Lamotte, R., de Palma, A., Geroliminis, N., 2017. The morning commute problem with temporary access restrictions for conventional and autonomous vehicles. In: hEART 2017: 6th Symposium of the European Association for Research in Transportation. URL https://infoscience.epfl.ch/record/253689

Lamotte, R., Geroliminis, N., 2016. The morning commute in urban areas: Insights from theory and simulation. In: Transportation Research Board 95th Annual Meeting.

Lamotte, R., Geroliminis, N., 2017. The morning commute in urban areas with heterogeneous trip lengths. Transportation Research Part B: Methodological (In press). URL https://doi.org/10.1016/j.trb.2017.08.023

Le Vine, S., Zolfaghari, A., Polak, J., 2015. Autonomous cars: The tension between occupant experience and intersection capacity. Transportation Research Part C: Emerging Technologies 52, 1–14.

Li, Z., Hensher, D. A., Rose, J. M., 2010. Willingness to pay for travel time reliability in passenger transport: A review and some new empirical evidence. Transportation Research Part E: Logistics and Transportation Review 46 (3), 384–403.

Lindsey, R., 2004. Existence, uniqueness, and trip cost function properties of user equilibrium in the bottleneck model with multiple user classes. Transportation Science 38 (3), 293–314.

Liu, L. N., McDonald, J. F., 1999. Economic efficiency of second-best congestion pricing schemes in urban highway systems. Transportation Research Part B: Methodological 33 (3), 157–188.

Liu, W., Yang, H., Yin, Y., 2015. Efficiency of a highway use reservation system for morning commute. Transportation Research Part C: Emerging Technologies 56, 293–308.

Loder, A., Ambühl, L., Menendez, M., Axhausen, K. W., 2017. Empirics of multi-modal traffic networks – Using the 3D macroscopic fundamental diagram. Transportation Research Part C: Emerging Technologies 82, 88–101.

Mahmassani, H., Herman, R., 1984. Dynamic user equilibrium departure time and route choice on idealized traffic arterials. Transportation Science 18 (4), 362–384.

Mahmassani, H., Williams, J. C., Herman, R., 1984. Investigation of network-level traffic flow relationships: some simulation results. Transportation Research Record: Journal of the Transportation Research Board 971, 121–130.

Mc Breen, J., Jensen, P., Marchal, F., 2006. An agent based simulation model of traffic congestion. In: ATT06: Proceedings of the 4th workshop on Agents in Traffic and Transportation. pp. 43–49.

Mounce, R., 2006. Convergence in a continuous dynamic queueing model for traffic networks. Transportation Research Part B: Methodological 40 (9), 779–791.

Newell, C., 1982. Applications of queueing theory. Vol. 4.

Newell, G. F., 1987. The morning commute for nonidentical travelers. Transportation Science 21 (2), 74–88.

Parthasarathi, P., Srivastava, A., Geroliminis, N., Levinson, D., 2011. The importance of being early. Transportation 38 (2), 227–247.

Peeta, S., Mahmassani, H. S., 1995. System optimal and user equilibrium time-dependent traffic assignment in congested networks. Annals of Operations Research 60 (1), 81–113.

Qian, X., Gregoire, J., Moutarde, F., De La Fortelle, A., 2014. Priority-based coordination of autonomous and legacy vehicles at intersection. In: IEEE 17th International Conference on Intelligent Transportation Systems (ITSC). pp. 1166–1171.

Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. Transportation Research Part B: Methodological 74, 1–19.

Rassenti, S. J., Smith, V. L., Bulfin, R. L., 1982. A combinatorial auction mechanism for airport time slot allocation. The Bell Journal of Economics 13 (2), 402–417.

Sandholm, W. H., 2010a. Local stability under evolutionary game dynamics. Theoretical Economics 5, 27–50.

Sandholm, W. H., 2010b. Population games and evolutionary dynamics. MIT press.

Sandholm, W. H., 2015. Chapter 13 - population games and deterministic evolutionary dynamics. In: Young, H. P., Zamir, S. (Eds.), Handbook of Game Theory with Economic Applications. Vol. 4. Elsevier, pp. 703–778.

Schepperle, H., Böhm, K., Forster, S., 2008. Traffic management based on negotiations between vehicles - a feasibility demonstration using agents. In: Ninth Workshop on Agent Mediated Electronic Commerce (AMEC IX), Honolulu, Hawaii, USA. pp. 90–104.

Skabardonis, A., Varaiya, P., Petty, K., 2003. Measuring recurrent and nonrecurrent traffic congestion. Transportation Research Record: Journal of the Transportation Research Board 1856, 118–124.

Small, K. A., 1982. The scheduling of consumer activities: Work trips. The American Economic Review 72 (3), 467–479.

Small, K. A., Chu, X., 2003. Hypercongestion. Journal of Transport Economics and Policy 37 (3), 319–352.

Small, K. A., Verhoef, E. T., 2007. The economics of urban transportation. Routledge.

Smith, M. J., 1979. The existence, uniqueness and stability of traffic equilibria. Transportation Research Part B: Methodological 13 (4), 295–304.

Smith, M. J., 1984a. The existence of a time-dependent equilibrium distribution of arrivals at a single bottleneck. Transportation Science 18 (4), 385–394.

Smith, M. J., 1984b. The stability of a dynamic model of traffic assignment: An application of a method of lyapunov. Transportation Science 18 (3), 245–252.

Smith, M. J., Ghali, M., 1990. The dynamics of traffic assignment and traffic control: A theoretical study. Transportation Research Part B: Methodological 24 (6), 409–422.

Surowiecki, J., 2014. In praise of efficient price gouging. MIT Technology Review.

Tisato, P., 1992. User cost minimisation and transport subsidy. Economics Letters 39 (2), 241–247.

Trannoy, A., 2011. Equity dimensions of transport policy. In: de Palma, A., Lindsey, R., Quinet, E., Vickerman, R. (Eds.), A Handbook of Transport Economics. pp. 624–649.

Tseng, Y.-Y., Verhoef, E. T., 2008. Value of time by time of day: A stated-preference study. Transportation Research Part B: Methodological 42 (7-8), 607–618.

van den Berg, V., Verhoef, E. T., 2011. Winning or losing from dynamic bottleneck congestion pricing?: The distributional effects of road pricing with heterogeneity in values of time and schedule delay. Journal of Public Economics 95 (7–8), 983–992.

van den Berg, V. A. C., Verhoef, E. T., 2012. Is the travel time of private roads too short, too long, or just right? Transportation Research Part B: Methodological 46 (8), 971–983.

van den Berg, V. A. C., Verhoef, E. T., 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.

van Wageningen-Kessels, F., Van Lint, H., Vuik, K., Hoogendoorn, S., 2015. Genealogy of traffic flow models. EURO Journal on Transportation and Logistics 4 (4), 445–473.

Varaiya, P., 1993. Smart cars on smart roads: Problems of control. IEEE Transactions on Automatic Control 38, 195–207.

Verhoef, E. T., 2001. An integrated dynamic model of road traffic congestion based on simple car-following theory: Exploring hypercongestion. Journal of Urban Economics 49 (3), 505–542.

Verhoef, E. T., 2007. Second-best road pricing through highway franchising. Journal of Urban Economics 62 (2), 337–361, essays in Honor of Kenneth A. Small.

Verhoef, E. T., Small, K. A., 2004. Product differentiation on roads: Constrained congestion pricing with heterogeneous users. Journal of Transport Economics and Policy 38 (1), 127–156.

Vickrey, W. S., 1961. Counterspeculation, auctions, and competitive sealed tenders. The Journal of Finance 16 (1), 8–37.

Vickrey, W. S., 1969. Congestion Theory and Transport Investment. American Economic Review 59 (2), 251–60.

Vickrey, W. S., 1973. Pricing, metering, and efficiently using urban transportation facilities. Highway Research Record 476, 36–48.

Wada, K., Akamatsu, T., 2013. A hybrid implementation mechanism of tradable network permits system which obviates path enumeration: An auction mechanism with day-to-day capacity control. Transportation Research Part E: Logistics and Transportation Review 60, 94–112.

Wang, H., Meng, Q., Zhang, X., 2014. Game-theoretical models for competition analysis in a new emerging liner container shipping market. Transportation Research Part B: Methodological 70, 201–227.

Wang, X., Ouyang, Y., 2013. A continuum approximation approach to competitive facility location design under facility disruption risks. Transportation Research Part B: Methodological 50, 90–103.

Watling, D., 1999. Stability of the stochastic equilibrium assignment problem: a dynamical systems approach. Transportation Research Part B: Methodological 33 (4), 281–312.

Watling, D., Hazelton, M. L., 2003. The dynamics and equilibria of day-to-day assignment models. Networks and Spatial Economics 3 (3), 349–370.

Wong, J.-T., 1997. Basic concepts for a system for advance booking for highway use. Transport Policy 4 (2), 109–114.

Xiao, F., Yang, H., Han, D., 2007. Competition and efficiency of private toll roads. Transportation Research Part B: Methodological 41 (3), 292–308.

Xiao, Y., Coulombel, N., de Palma, A., 2017. The valuation of travel time reliability: does congestion matter? Transportation Research Part B: Methodological 97, 113–141.

Yang, H., Huang, H.-J., 1997. Analysis of the time-varying pricing of a bottleneck with elastic demand using optimal control theory. Transportation Research Part B: Methodological 31 (6), 425–440.

Yildirimoglu, M., Limniati, Y., Geroliminis, N., 2015. Investigating empirical implications of hysteresis in day-to-day travel time variability. Transportation Research Part C: Emerging Technologies 55, 340–350.

Zhang, X., Zhang, H., Huang, H.-J., Sun, L., Tang, T.-Q., 2011. Competitive, cooperative and Stackelberg congestion pricing for multiple regions in transportation networks. Transportmetrica 7 (4), 297–320.

Raphaël Lamotte
Rue des artisans 8,
1026 Echandens,
Switzerland

Phone : +41 21 693 53 97
Website: `http://luts.epfl.ch/`
`https://github.com/raplam`
E-mail : `raphael.lamotte@epfl.ch`

Education

2014–2018 **PhD** Laboratory of Urban Transportation Systems, EPFL, Switzerland.
Thesis title: Congestion and departure time choice equilibrium in urban road networks.

2012–2014 **M.A.Sc.** Concordia University, Canada.
Major: transportation engineering. *Early graduation award.*

2010–2012 **M.Sc.Eng.** Ecole Centrale de Lyon, France.
Multidisciplinary engineering. *With high honors.*

2008–2010 **French *Classe préparatoire*** Lycée Champollion, France.
Major: Mathematics, Physics and Computer science

Research projects

2017–2018 **A city-level observatory of mobility,** Swisscom AG.
We assessed the potential of cellphone traces for monitoring and managing mobility at the city scale. A larger follow-up project is under discussion.

2014–2018 **A real-time traffic control strategy for Geneva,** Canton of Geneva, DGT.
The team I coordinated analyzed data related to road and public transit operations. We found important technical malfunctions and proposed an adaptive congestion-management scheme for the city centre.

2014–2018 **METAFERW (ERC grant):** Modeling and controlling traffic congestion and propagation in large-scale urban multimodal networks.

2012–2014 **Safety of HOV lanes,** Transports Quebec, McGill University.

Publications

Papers in international journal (published)

Lamotte, R., de Palma, A. and Geroliminis, N., 2017. On the use of reservation-based autonomous vehicles for demand management. Transportation Research Part B: Methodological, 99, pp.205-227.

Lamotte, R. and Geroliminis, N., 2017. The morning commute in urban areas with heterogeneous trip lengths. Transportation Research Part B: Methodological (In Press). *Presented at the 22nd International Symposium on Transportation and Traffic theory (ISTTT), Chicago, Illinois.*

Journal papers in preparation

Murashkin, M., Lamotte, R. and Geroliminis, N. Dynamic Modeling of Trip Completion Rate in Urban Areas with MFD Representations.

Lamotte, R. and Geroliminis, N. Scheduling trips through a bottleneck: monotonicity and stability.

Book chapter

Lamotte, R., Meyer, R., Geroliminis, N., 2017. La congestion en ville : une approche globale. In: Big data et politiques publiques dans les transports. Eds: de Palma, A. and Dantan, S., Economica.

Papers in conference proceedings

Lamotte, R., Murashkin, M., Kouvelas, A., Geroliminis, N., 2018. Dynamic Modeling of Trip Completion Rate in Urban Areas with MFD Representations. Presented at the 97th Annual Meeting of the Transportation Research Board, Washington, DC.

Lamotte, R., Geroliminis, N., 2018. (In)stability of departure time choice with the bottleneck model. Presented at the 18th Swiss Transport Research Conference (STRC), Ascona, Switzerland.

Lamotte, R., Geroliminis, N., 2017. The morning commute in homogeneous cities. Presented at the 17th Swiss Transport Research Conference (STRC), Ascona, Switzerland.

Lamotte, R., Geroliminis, N., 2016. The morning commute in urban areas: Insights from theory and simulation. Presented at the 95th Annual Meeting of the Transportation Research Board, Washington, DC.

Lamotte, R., Kouvelas, A., Geroliminis, N., 2016. Estimation of macroscopic traffic variables in cities with big but sparse multi-sensor data. Presented at the 16th Swiss Transport Research Conference (STRC), Ascona, Switzerland.

Lamotte, R., Geroliminis, N., 2015. Dynamic Traffic Modeling: Approximating the equilibrium for peak periods in urban areas. Presented at the 15th Swiss Transport Research Conference (STRC), Ascona, Switzerland.

Lamotte, R. and Alecsandru, C., 2014. Fast multi-objective optimization for continuous network design problems based on Gaussian process models. Presented at the 93rd Annual Meeting of the Transportation Research Board, Washington, DC.

Podium presentations

- ISTTT (International Symposium on Transportation and Traffic Theory): 2017 (Chicago, USA).
- hEART (European Association for Research in Transportation) : 2015 (Copenhagen, Denmark), 2017 (Haifa, Israel), 2018 (Athens, Greece).
- IFORS (International Federation of Operational Research Societies): 2017 (Quebec City, Canada).
- ITEA (International Transportation Economics Association): 2015 (Oslo, Norway), 2017 (Barcelona, Spain).

Teaching activities

Lecturer & T. A. **Transportation Systems Engineering** (Bachelor, 3rd year)
Lead assistant from 2015 to 2018, 2nd assistant in 2014. I acted as substitute lecturer for lessons on transit signal priority; public transit operations; route and departure time choice.

Global Issues: Mobility (Bachelor, 1st year)
Lead assistant in 2015.

MOOC Creation of an online course on "Traffic flow modeling and ITS", taught together with Prof. Geroliminis (EPFL) and Dr. Kouvelas (ETHZ). I was responsible for the module on "Equilibria in transportation".

Supervision Makni, M., 2017 (Master thesis),
Zaradicka, M., 2016 (Master thesis),
6 semester projects.

Reviewing

- Economics of Transportation
- Journal of Advanced Transportation
- Transportation Research Part B: Methodological
- Transportation Research Part C: Emerging technologies
- Transportation Science
- Transportation Research Record (TRR)
- IEEE Intelligent Transportation Systems Society Conference
- International Symposium on Transportation and Traffic Theory (ISTTT)

Popular science

- Comment se forment les bouchons, CQFD, la 1ère, May 3, 2017 (in French).
- Les algorithmes contre les bouchons, Nouvo, RTS, September 15, 2015 (in French).

Skills

Programming I mostly used MATLAB during my PhD but I was exposed to various languages during my bachelor and masters, including C++, Python, PHP, HTML, CSS, MySQL and Caml.

Languages English: fluent | French: native | German: basic (formerly B1).