Transportation Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

On the Existence and Uniqueness of Equilibrium in the Bottleneck Model with Atomic Users

Hugo E. Silva, Robin Lindsey, André de Palma, Vincent A. C. van den Berg

informs
http://pubsonline.informs.org/journal/trsc/

TRANSPORTATION SCIENCE
Vol. 51, No. 3, August 2017, pp. 863–881
ISSN 0041-1655 (print), ISSN 1526-5447 (online)

On the Existence and Uniqueness of Equilibrium in the Bottleneck Model with Atomic Users

Hugo E. Silva,[a] Robin Lindsey,[b] André de Palma,[c] Vincent A. C. van den Berg[d,e]

[a] Departamento de Ingeniería de Transporte y Logística and Instituto de Economía, Pontificia Universidad Católica de Chile, Santiago, Chile 7820436; [b] Sauder School of Business, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada; [c] École Normale Supérieure de Cachan, Département Economie et Gestion, 94230 Cachan, France; [d] Department of Spatial Economics, VU University Amsterdam, 1081 HV Amsterdam, Netherlands; [e] Tinbergen Institute, 1082 MS Amsterdam, Netherlands

Contact: husilva@uc.cl (HES); robin.lindsey@sauder.ubc.ca (RL); andre.depalma@ens-cachan.fr (AdP); vberg@feweb.vu.nl (VACvdB)

Received: October 14, 2014
Revised: May 1, 2015; September 11, 2015
Accepted: November 25, 2015
Published Online in Articles in Advance: August 11, 2016

https://doi.org/10.1287/trsc.2016.0672

Abstract. This paper investigates the existence and uniqueness of equilibrium in the Vickrey bottleneck model when each user controls a positive fraction of total traffic. Users simultaneously choose departure schedules for their vehicle fleets. Each user internalizes the congestion cost that each of its vehicles imposes on other vehicles in its fleet. We establish three results. First, a pure strategy Nash equilibrium (PSNE) may not exist. Second, if a PSNE does exist, identical users may incur appreciably different equilibrium costs. Finally, a multiplicity of PSNE can exist in which no queuing occurs but departures begin earlier or later than in the system optimum. The order in which users depart can be suboptimal as well. Nevertheless, by internalizing self-imposed congestion costs individual users can realize much, and possibly all, of the potential cost savings from either centralized traffic control or time-varying congestion tolls.

Funding: Financial support from ERC Advanced [Grant OPTION 246969] and the Social Sciences and Humanities Research Council of Canada [Grant 435-2014-2050] is gratefully acknowledged.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/trsc.2016.0672.

Keywords: bottleneck model • large users • atomic users • existence of equilibrium • uniqueness of equilibrium

1. Introduction

Individual users of roads and other transportation facilities are usually assumed to be small in the sense that they control a negligible fraction of total traffic. However, large users are prevalent in many settings. Commercial airlines and rail companies often account for a sizable fraction of total traffic at airports and on rail networks. Postal services and major freight shippers operate large vehicle fleets that travel long distances each day. For example, FedEx handles about 150 daily flights out of Memphis International Airport and its air-cargo operations support tens of thousands of jobs. UPS operates on a similar scale out of Louisville, Kentucky (*The Economist* 2013). Major employers such as government departments and large corporations can add substantially to traffic on certain roads at peak times. Large users such as these suffer from the congestion delays their own aircraft, trains, trucks, or other vehicles impose on each other. Thus, at airports, on rail networks, on congested roads, and on other transportation infrastructure networks, one would expect large users to internalize their self-imposed delays, and therefore to make different trip-related decisions than small users controlling the same aggregate traffic.

Following the terminology of game theory we will refer to small users as *nonatomic*, and large users that control a positive fraction of traffic as *atomic*. This terminology contrasts with the terminology used in the literature on airport congestion, beginning with Daniel (1995), in which users that control a negligible fraction of traffic and treat the congestion levels as parametric are called atomistic users. Somewhat confusingly, atomistic users are therefore nonatomic, and nonatomistic users are atomic.

There are several branches of literature on congestion with atomic users. In the aviation literature, Daniel (1995) was the first to recognize that airlines with market power and large shares of total traffic could internalize the delays their aircraft impose on each other. Brueckner (2002) showed that under Cournot competition airlines fully internalize self-imposed congestion. Further contributions in this line have been made by Pels and Verhoef (2004), Brueckner (2005), Zhang and Zhang (2006), Basso and Zhang (2007), Brueckner and Van Dender (2008), and Silva and Verhoef (2013). In the context of road transportation, route-choice decisions by atomic users have been studied (e.g., Devarajan 1981; Haurie and Marcotte 1985; Marcotte 1987; Harker 1988; Catoni and Pallottino 1991; Miller, Tobin, and Friesz 1991; Yang, Zhang, and Meng 2007; Yang and Zhang 2008; Cominetti, Correa, and Stier-Moses 2009). Yang and Zhang (2008) also studied optimal pricing with atomic and nonatomic users, and He et al. (2013)

analyzed tradable credit schemes. There is also a literature in operations research and computer science on atomic congestion games (e.g., Fotakis, Kontogiannis, and Spirakis 2008; Hoefer and Skopalik 2009).

The aforementioned studies have used static models of congestion except for Daniel (1995) who uses a stochastic queuing model empirically. Other studies that use a dynamic stochastic model of congestion with atomic users include Daniel and Harback (2008) and Molnar (2013). Except for a few studies described next, traffic congestion with atomic users has not been studied with deterministic dynamic models. This is surprising because the timing of trips matters a great deal for both passenger and freight transportation, and congestion is largely a consequence of peak-period loads.

The goal of this paper is to investigate the fundamental questions of existence and uniqueness of equilibrium in trip-timing decisions with atomic users. To focus the analysis on fundamentals while minimizing mathematical complications we use Vickrey's (1969) bottleneck model. The essence of the bottleneck model is that users trade off the costs of queuing delay at the bottleneck with the costs of schedule delay (i.e., arriving earlier or later than desired). The bottleneck model has been used to study many aspects of trip-timing decisions with congestion including: congestion pricing, route choice on simple road networks, mode choice, trip chaining, parking congestion, staggered work hours, and flextime. Existence and uniqueness of equilibrium in the bottleneck model have also been established under relatively general assumptions about trip-timing preferences and heterogeneity of nonatomic users (e.g., Newell 1987; Lindsey 2004). However, little consideration has been given to atomic users in studies that use either the bottleneck model or other dynamic models.

A few studies have employed a variant of the bottleneck model in which time is discretized, and the number of users is finite, so that each user controls a positive measure of traffic (e.g., Levinson 2005; Zou and Levinson 2006; Otsubo and Rapoport 2008; Werth, Holzhauser, and Krumke 2014). However, these studies assume that each user controls only one vehicle so that self-internalization of congestion does not come into play. To our knowledge, only Daniel (2009) and Silva, Verhoef, and van den Berg (2014) have explored the scheduling decisions of atomic agents in the standard, continuous-time bottleneck model. These studies consider, in the context of aviation, a sequential competition between a Stackelberg leader with market power and a group of perfectly competitive airlines (nonatomic users). Both studies show that, when users have homogeneous preferences, nonatomic users schedule all of their flights during the peak period when passengers prefer to arrive. The Stackelberg leader schedules a fraction of its flights during the peak as well. Queuing time at the bottleneck evolves at the same rate as in the standard model of nonatomic players. The leader schedules its remaining flights earlier and later in the off-peak (and less popular) periods and limits its departure rate to bottleneck capacity so that no queue develops.

The existence of a unique equilibrium in Daniel (2009) and Silva, Verhoef, and van den Berg (2014) hinges on the sequential nature of the game they consider, and the assumption that there is only one atomic agent. By contrast, we focus in this paper on settings with two atomic users who make scheduling decisions simultaneously. The solution concept we employ is pure strategy Nash equilibrium (PSNE). We establish three major results. First, we show that a PSNE may not exist. We demonstrate this for an example featuring two identical atomic users who each control half of the total traffic. The trip-timing preferences of each vehicle in each fleet are described by parameters $\{\alpha, \beta, \gamma, t^*\}$, where α is the cost of travel time, β is the cost of schedule delay early, γ is the cost of schedule delay late, and t^* is the desired arrival time. We show that if $\gamma > \alpha$, a PSNE does not exist.

Second, for the same example, we show that if $\gamma \leq \alpha$ a multiplicity of PSNE exists in which no queuing occurs and the timing of departures is system optimal. The PSNE differ according to the departure rates of individual users and the equilibrium costs they incur. Depending on parameter values, one of the two users can incur up to three quarters of total costs. The cases $\gamma > \alpha$ and $\gamma \leq \alpha$ are both of theoretical interest, and each may be relevant in particular settings. Most empirical studies of scheduling preferences for automobile drivers have obtained estimates that satisfy $\gamma > \alpha$ (e.g., Small 1982; Wardman 2001; Asensio and Matas 2008). In addition, de Palma and Fontan (2001) list estimates from 11 studies and, including their own estimates, there are 10 cases with $\gamma > \alpha$, and two cases with $\gamma < \alpha$. By contrast, Daniel and Harback (2008) find that $\gamma < \alpha$ holds for most airlines at major U.S. airports.

Third, we consider a variant of the example in which the two users differ in their desired arrival times t^* and can have fleets of different size. We show that—independent of the relative size of α and γ—a multiplicity of PSNE can exist in which no queuing occurs but the timing of departures is not optimal. Depending on parameter values, the PSNE may begin earlier than, later than, or at the same time as the system optimum. The order in which users depart can be suboptimal as well. Nevertheless, by internalizing self-imposed congestion costs the two users realize much, and possibly all, of the potential cost savings from either centralized traffic control or time-varying congestion tolls.

These examples demonstrate that neither the existence of equilibrium nor the uniqueness of an equilibrium (if one exists) is guaranteed under conditions where a unique PSNE does exist if all of the traffic were controlled by nonatomic users. Given the central role of equilibrium models in the analysis of transportation systems, these results are troubling and highlight the need for further research.

The paper is organized as follows. Section 2 reviews the no-toll equilibrium and the system optimum in the standard bottleneck model with nonatomic users. Section 3 demonstrates the possible non-existence of PSNE, and the nonuniqueness of individual departure rates and costs where a PSNE does exist. Section 4 demonstrates the possible nonuniqueness of PSNE in the timing of departures when no queuing occurs, and the degree of inefficiency relative to the system optimum. Section 5 concludes and develops ideas for future research.

2. The Bottleneck Model with Homogeneous Nonatomic Users

The bottleneck model was developed by Vickrey (1969) and extended by Arnott, de Palma, and Lindsey (1990, 1993). It is reviewed in Arnott, de Palma, and Lindsey (1998) and de Palma and Fosgerau (2011), and the summary here is brief. In the model, all users travel from a common origin to a common destination along a single link that has a bottleneck with fixed flow capacity, s. Without loss of generality, travel times from the origin to the bottleneck and from the bottleneck to the destination are normalized to zero. If there is no queue upstream of the bottleneck, travel time through the bottleneck is also zero and departure time from the origin coincides with arrival time at the destination. If the departure rate exceeds s, a queue develops. Let $\hat{t}$ be the most recent time at which there was no queue, and $r(t)$ the aggregate departure rate from the origin at time t. The number of vehicles in the queue is then

$$Q(t) = \int_{\hat{t}}^{t} (r(u) - s)\, du.$$

Travel time through the bottleneck is $T(t) = Q(t)/s$, and a traveler who departs at time t arrives at time $t_a = t + T(t)$.

Following Small (1982) and de Palma et al. (1983), users are assumed to have a desired arrival time t^*. They incur a unit cost of $\beta > 0$ for arriving early, and a unit cost of $\gamma > 0$ for arriving late. Travel time is valued at α, with $\alpha > \beta$. The costs of schedule delay and travel time are additive so that the generalized cost of a trip, $c(t)$, is

$$c(t) = \alpha T(t) + \begin{cases} \beta(t^* - t - T(t)), & t + T(t) \le t^* \\ \gamma(t + T(t) - t^*), & t + T(t) \ge t^*. \end{cases} \tag{1}$$

The number (or measure) of users, N, is assumed to be exogenous (i.e., independent of trip cost). Each user decides when to depart from the origin by trading off schedule delay against travel delay. A PSNE is a set of departure times such that no user can benefit (i.e., reduce trip cost) by unilaterally changing departure time while taking the departure times of all other users as given.

2.1. No-Toll Equilibrium

Let superscript n denote the no-toll nonatomic PSNE, and t_s^n and t_e^n denote the start and end of the travel period. Let $\tilde{t}$ be the departure time for which a user arrives on time (i.e., $\tilde{t} + T(\tilde{t}) = t^*$). In a PSNE with no toll, $c(t)$ must be constant during the travel period $[t_s^n, t_e^n]$. Users who arrive closer to t^* must incur longer queuing delays to offset their lower schedule delay costs. The equilibrium aggregate departure rate is derived by differentiating Equation (1) and setting the derivative to zero

$$r^n(t) = \begin{cases} \dfrac{\alpha s}{\alpha - \beta}, & t \in (t_s^n, \tilde{t}) \\ \dfrac{\alpha s}{\alpha + \gamma}, & t \in (\tilde{t}, t_e^n). \end{cases} \tag{2}$$

The assumption $\alpha > \beta$ assures that the departure rate for early arrivals is positive and finite. This condition is plausible because a user who is destined to arrive early is likely to prefer arriving early to prolonging the trip by making a detour. The condition is also supported by Small's (1982) estimates for automobile commuting trips. For ease of reference we will sometimes call the departure rate for early arrivals the *early departure rate*, and the departure rate for late arrivals the *late departure rate*.

There are two further equilibrium conditions. One is that the first and last users to depart, who encounter no queue, must incur equal schedule delay costs

$$\beta(t^* - t_s^n) = \gamma(t_e^n - t^*). \tag{3}$$

The other condition is that the travel period lasts for N/s

$$t_e^n - t_s^n = \frac{N}{s}. \tag{4}$$

Together, equilibrium conditions (2)–(4) yield

$$t_s^n = t^* - \frac{\gamma}{\beta + \gamma}\frac{N}{s}, \tag{5a}$$

$$t_e^n = t^* + \frac{\beta}{\beta + \gamma}\frac{N}{s}, \tag{5b}$$

$$\tilde{t} = t^* - \frac{\beta}{\alpha}\frac{\gamma}{\beta + \gamma}\frac{N}{s}, \tag{5c}$$

$$c^n(t) = \delta\frac{N}{s}, \quad t \in [t_s^n, t_e^n] \text{ with } \delta \equiv \frac{\beta\gamma}{\beta + \gamma}. \tag{5d}$$

Total costs are

$$TC^n = \delta\frac{N^2}{s}. \tag{6}$$

2.2. System Optimum

Queuing delay at the bottleneck is a deadweight loss. The system optimum therefore avoids queuing and minimizes total schedule delay costs. The departure rate is maintained at s over a continuous time interval chosen so that the first and last users incur the same schedule delay cost. The departure period is therefore the same as in the laissez-faire PSNE (see Equations (5a) and (5b)). Using superscript o to denote the system optimum, these results are recorded for future reference as

$$r^o(t) = \begin{cases} s, & t \in (t_s^o, t_e^o), \\ 0, & \text{otherwise}, \end{cases} \tag{7a}$$

$$t_s^o = t^* - \frac{\gamma}{\beta+\gamma}\frac{N}{s}, \tag{7b}$$

$$t_e^o = t^* + \frac{\beta}{\beta+\gamma}\frac{N}{s}. \tag{7c}$$

Total system costs are only half as large as in Equation (6) for the no-toll equilibrium

$$TC^o = \frac{\delta}{2}\frac{N^2}{s} = \frac{1}{2}TC^n. \tag{8}$$

The difference between total costs in the no-toll equilibrium and system optimum, $TC^n - TC^o$, serves as an upper bound on the benefits from self-internalization of congestion by atomic agents.

3. Existence and Nonexistence of Equilibrium with Homogeneous Atomic Users

In this section we study an example featuring two identical atomic users. We show that if $\gamma > \alpha$, a PSNE in departure schedules does not exist. We then show that if $\gamma \le \alpha$, a PSNE does exist that entails no queuing and coincides with the system optimum. During early departures the two users can depart at somewhat different rates that add up to s. If $\gamma < \alpha$, their late departure rates can also differ. Moreover, with $\gamma \le \alpha$ the users can incur appreciably different total costs for their fleets. In Section 3.3, we briefly discuss how these results extend to more than two users.

Consider two atomic users, A and B. Each user controls a fleet of $N/2$ vehicles. Trip-timing preferences for each vehicle are defined by the same $\{\alpha, \beta, \gamma, t^*\}$ parameter values. Thus, the total cost of a user's fleet is simply the sum of the costs of the individual vehicles. This assumption seems realistic for delivery vans carrying merchandise or parcels to different customers. It may be inappropriate for vehicles in a military convoy or emergency vehicles traveling to an accident. Users A and B simultaneously choose departure schedules for their fleets. The schedule for user i is a departure rate function, $r_i(t) \ge 0$. The function can be seen as a distribution function of the $N/2$ vehicles over some extended time interval such as a day. This function is not restricted to be continuous, and the possibility of mass departures will be considered. Each user recognizes that dispatching a vehicle at time t may delay vehicles in its fleet that depart after t. A delay occurs if there is a queue at time t that persists when the later vehicles depart. A delay also occurs if there is no queue prior to t, but the bottleneck is at capacity so that adding a vehicle to the departure schedule at t creates a (small) queue.

As noted previously, the existence of a PSNE in this example depends on whether $\gamma > \alpha$, or $\gamma \le \alpha$. The two cases are considered in Sections 3.1 and 3.2.

3.1. Nonexistence of PSNE with $\gamma > \alpha$

When $\gamma > \alpha$, a PSNE does not exist. This result is formalized in the following proposition.

Proposition 1. *Consider two identical atomic users who each simultaneously schedule $N/2$ vehicles with trip-timing preferences for each vehicle defined by the same $\{\alpha, \beta, \gamma, t^*\}$ parameter values. If $\gamma > \alpha$, a PSNE in departure schedules does not exist.*

We prove Proposition 1 in four steps. First, we prove that a PSNE without queuing does not exist (Lemma 1). Second, we prove that mass departures cannot arise in equilibrium (Lemma 2). Hence, we can restrict attention to cases in which there is queuing, but no mass departures. Third, we show that there is a unique departure pattern with queuing such that a user cannot reduce its fleet costs by rescheduling a single vehicle (Lemma 3). Finally, we show that this departure pattern is not a PSNE because a user can reduce its fleet costs by rescheduling a positive measure of vehicles in its fleet (Lemma 4). Many such deviations are gainful. They all entail postponing departures to reduce queuing delay. We demonstrate this graphically, and then identify the most gainful deviation in which the deviant user avoids queuing altogether. Because deviation is gainful, a PSNE with queuing does not exist.

Lemma 1. *Consider two identical atomic users who each simultaneously schedule $N/2$ vehicles with trip-timing preferences for each vehicle defined by the same $\{\alpha, \beta, \gamma, t^*\}$ parameter values. If $\gamma > \alpha$, a PSNE without queuing does not exist.*

Proof. Consider a pair of departure schedules, $\{r_A(\cdot), r_B(\cdot)\}$, such that no queuing occurs. Some vehicles must arrive late since otherwise a user could reduce its fleet costs by rescheduling some vehicles to just after t^*. Consider a period (t_1, t_2) of late arrivals and assume that both users depart during this period (the case where only one user departs is considered later in the proof). The bottleneck must be used to capacity because otherwise a user could exploit the residual capacity by advancing departures for vehicles that are scheduled to depart later.

As, by assumption, in Lemma 1 there is no queue, if user i removes a vehicle from the departure schedule at time t, it saves a cost of

$$C_i^-(t) = \gamma(t - t^*). \tag{9}$$

Removing the vehicle saves the late-arrival cost incurred by the vehicle itself, but it has no effect on the rest of the fleet because there is no queue. If user i instead adds a vehicle to the departure schedule, it now causes a queue and increases its fleet costs by a marginal private cost (MPC) of

$$C_i^+(t) = \gamma(t - t^*) + \frac{\alpha + \gamma}{s} \int_t^{\bar{t}} r_i(u)\, du, \tag{10}$$

where $\bar{t}$ is the time when the queue created by the additional vehicle disappears.

The first term on the right-hand side of Equation (10) matches the right-hand side of Equation (9). The second term is the delay cost imposed on user i's other vehicles that depart from t to $\bar{t}$. Each of them suffers an increase in travel time of $1/s$ valued at α, and an increase in late arrival of $1/s$ valued at γ.

The difference between the cost saved by removing a vehicle given by Equation (9), and the cost of adding a vehicle to the same slot given by Equation (10), arises when the bottleneck is at capacity but there is no queue. Note that Equations (9) and (10) are valid only for a setting without queuing. After Lemma 2, we will consider settings with queuing.

Suppose user i advances the departure of a vehicle from t to t' where $t_1 \le t' < t \le t_2$. Then, from the assumed setting without queuing, user i's fleet costs change by

$$\begin{aligned}\Delta C_i &= -C_i^-(t) + C_i^+(t') \\ &= -\gamma(t - t^*) + \gamma(t' - t^*) + \frac{\alpha + \gamma}{s} \int_{t'}^{t} r_i(u)\, du.\end{aligned} \tag{11}$$

The queue induced by adding the vehicle at t' vanishes at t because a departure time slot opened up at t when the vehicle was removed. Let $\lambda_{t,t'}^i = \int_{t'}^t r_i(u)\, du / (s(t - t')) \in [0, 1]$ be an auxiliary variable denoting the average fraction of capacity occupied by user i during the period $[t, t']$. The change in user i's fleet costs can then be written as

$$\begin{aligned}\Delta C_i &= \frac{\alpha + \gamma}{s} \lambda_{t,t'}^i s(t - t') - \gamma(t - t') \\ &= ((\alpha + \gamma)\lambda_{t,t'}^i - \gamma)(t - t').\end{aligned}$$

For a PSNE to exist, ΔC_i must be nonnegative for both users that requires

$$\lambda_{t,t'}^i \ge \frac{\gamma}{\alpha + \gamma}, \quad t_1 \le t' < t \le t_2,\ i = A, B. \tag{12}$$

Because the bottleneck is fully utilized, $\lambda_{t,t'}^A + \lambda_{t,t'}^B = 1$. This condition is least restrictive if $\lambda_{t,t'}^A = \lambda_{t,t'}^B = 1/2$ in which case it reduces to $\gamma \le \alpha$ that is inconsistent with the assumption $\gamma > \alpha$.

Now consider the possibility that only one user, say user A, departs during (t_1, t_2). This is not a PSNE if user B departs after t_2 because user B could reduce its costs by rescheduling some of its later vehicles into (t_1, t_2). Doing so reduces their late-arrival costs, and the queue they create disappears during the time slots they vacated. Suppose user B does not depart after t_2. If $t_1 = t^*$, user B does not depart late at all. Yet this cannot be a PSNE because user B could gain by rescheduling some of its early-arriving vehicles to just after t^*, thereby reducing their schedule delay costs without imposing any delay on its other vehicles. If $t_1 > t^*$, there must exist a late arrival period (t_0, t_1) during which both users depart. Yet this case has already been considered, and shown to be inconsistent with a PSNE when $\gamma > \alpha$. Q.E.D.

To this point we have assumed that users depart at a finite rate. In theory, a user could schedule a positive measure of vehicles to depart at a given moment. In practice, this might be achieved by assembling a convoy of vehicles on a link that has a right-of-way over other links. Moreover, in the bottleneck model with nonatomic users and step tolls a PSNE may exist only if mass departures (of noncooperating vehicles) are possible (Arnott, de Palma, and Lindsey 1990; Lindsey, van den Berg, and Verhoef 2012). We now show that in the model with atomic users mass departures cannot occur in a PSNE, regardless of whether $\gamma > \alpha$ or $\gamma \le \alpha$.

Lemma 2. *Consider two identical atomic users who each simultaneously schedule $N/2$ vehicles with trip-timing preferences for each vehicle defined by the same $\{\alpha, \beta, \gamma, t^*\}$ parameter values. A PSNE cannot exhibit mass departures.*

Proof. See Online Appendix A.

We now turn to the final possibility for a PSNE with $\gamma > \alpha$ in which departure rates remain finite and queuing occurs.

Lemma 3. *Consider two identical atomic users who each simultaneously schedule $N/2$ vehicles with trip-timing preferences for each vehicle defined by the same $\{\alpha, \beta, \gamma, t^*\}$ parameter values. If $\gamma > \alpha$, there is a unique departure pattern with queuing in which a user cannot reduce its fleet costs by rescheduling a single vehicle.*

Proof. Consider a pair of departure schedules, $\{r_A(\cdot), r_B(\cdot)\}$, and let $\tilde{t}$ denote the departure time for which a user arrives on time (i.e., $\tilde{t} = t^* - T(\tilde{t})$). Assume that a queue exists during a late-departure period (t_l, t_{qe}), where t_l is an arbitrary time that satisfies $t_l > \tilde{t}$ and t_{qe} is the time when the queue disappears. Because users A and B are identical, it suffices to consider the best response of user A to $r_B(\cdot)$. The MPC to user A of scheduling a vehicle at time $t \in (t_l, t_{qe})$ is

$$C_A(t) = \alpha T(t) + \gamma(t + T(t) - t^*) + \frac{\alpha+\gamma}{s}\int_t^{t_{qe}} r_A(u)\,du. \tag{13}$$

Equation (13) has a similar interpretation to Equation (10). The first two terms on the right-hand side comprise the cost incurred by the vehicle itself, and the third term is the delay cost imposed on user A's other vehicles that depart from t to t_{qe}.

User A could reschedule a vehicle from t to another time $t' \in (t_l, t_{qe})$. This would leave t_{qe} unchanged because the additional queuing time caused by inserting the vehicle at t' is offset by the reduction in queuing time as a result of removing the vehicle at t. Equation (13) therefore holds if t is replaced by any $t' \in (t_l, t_{qe})$. Hence, a necessary condition for $r_A(\cdot)$ to be a best response to $r_B(\cdot)$ is that $C_A(t)$ in Equation (13) is constant during the interval (t_l, t_{qe}). Differentiating Equation (13) with respect to t, and setting the derivative to zero, yields

$$\frac{\partial C_A(t)}{\partial t} = \gamma + (\alpha+\gamma)\frac{\partial T(t)}{\partial t} - \frac{\alpha+\gamma}{s} r_A(t) = 0.$$

Using the relationship $\partial T(t)/\partial t = (r_A(t) + r_B(t) - s)/s$, this condition simplifies to

$$r_B(t) = \frac{\alpha s}{\alpha+\gamma}, \quad t \in (t_l, t_{qe}). \tag{14}$$

According to Equation (14), user A is willing to schedule a vehicle for late arrival when there is a queue if, and only if, user B is departing at exactly the rate $\alpha s/(\alpha+\gamma)$. This is the equilibrium aggregate departure rate for the model with nonatomic users (cf. Equation (2)). Equation (14) also holds for user B with $r_A(t)$ in place of $r_B(t)$. The aggregate departure rate during an interval of late arrivals must therefore be $r(t) = r_A(t) + r_B(t) = 2\alpha s/(\alpha+\gamma)$. With $\gamma > \alpha$, $r(t) < s$ and the queue must be shrinking for late arrivals. Consequently, a queue must exist at time $\tilde{t}$ and it is possible to set $t_l = \tilde{t}$, where $\tilde{t} + T(\tilde{t}) = t^*$. This in turn implies

$$r_A(t) = r_B(t) = \frac{\alpha s}{\alpha+\gamma}, \quad t \in (\tilde{t}, t_{qe}).$$

Since a queue exists at time $\tilde{t}$, it must have been built up during a period of early arrivals before $\tilde{t}$. Let t_q be the time at which queuing begins. The MPC to user A of scheduling a vehicle at any time $t \in (t_q, \tilde{t})$ is

$$C_A(t) = \alpha T(t) + \beta(t^* - t - T(t)) + \frac{\alpha-\beta}{s} \cdot \int_t^{\tilde{t}} r_A(u)\,du + \frac{\alpha+\gamma}{s}\int_{\tilde{t}}^{t_{qe}} r_A(u)\,du. \tag{15}$$

Again, the first two terms on the right-hand side of Equation (15) comprise the cost borne by the vehicle itself. The third term is the cost imposed on user A's other vehicles that depart after t but still arrive early. Each of them suffers an increase in travel time of $1/s$ valued at α, and benefits from a reduction in early arrival of $1/s$ valued at β. The last term in Equation (15) is the cost imposed on user A's other vehicles that arrive late.

A necessary condition for $r_A(\cdot)$ to be a best response to $r_B(\cdot)$ is for $C_A(t)$ to be constant during $(t_q, \tilde{t})$. Setting the derivative of $C_A(t)$ to zero, one obtains a counterpart to Equation (14)

$$r_B(t) = \frac{\alpha s}{\alpha-\beta}, \quad t \in (t_q, \tilde{t}). \tag{16}$$

An analogous necessary condition applies for user B. Hence, the aggregate early departure rate must be $r(t) = r_A(t) + r_B(t) = 2\alpha s/(\alpha-\beta)$.

In summary, the unique departure rate of the candidate PSNE during the full period of queuing is

$$r_A(t) = r_B(t) = \frac{r(t)}{2} = \begin{cases} \dfrac{\alpha s}{\alpha-\beta}, & t \in (t_q, \tilde{t}) \\ \dfrac{\alpha s}{\alpha+\gamma}, & t \in (\tilde{t}, t_{qe}), \end{cases} \tag{17}$$

where

$$\tilde{t} + T(\tilde{t}) = t^*. \tag{18}$$

Because queuing begins at t_q, and ends at t_{qe}, cumulative departures during the period (t_q, t_{qe}) match cumulative arrivals

$$\int_{t_q}^{t_{qe}} (r(u) - s)\,du = 0. \tag{19}$$

Equations (17)–(19) define evolution of the queue for the candidate PSNE with queuing.

By Lemma 1, no vehicles can depart without queuing after t_{qe}, so departures end at time $t_e = t_{qe}$. However, the cost of a vehicle trip at the beginning of the queuing period, $c(t_q)$, is less than the cost at the end of the period, $c(t_e)$, because the trip at t_q imposes a private delay cost on subsequent vehicles whereas the trip at t_e does not (if $c(t_q) = c(t_e)$, $C^+(t_q) > C^+(t_e)$ would hold and the user equilibrium condition would be violated). Departures must therefore occur during some time interval $[t_s, t_q)$ preceding t_q. This interval is defined by two conditions. First, vehicle trip cost must be the same at t_s and t_e since otherwise a user could reschedule vehicles from the time with higher cost to the time with lower cost and reduce its overall fleet costs without causing any queuing. Second, the full departure period $[t_s, t_e]$ must be long enough for all N vehicles to pass the bottleneck. Equations (3) and (4) for the nonatomic PSNE therefore hold for the candidate PSNE

$$\beta(t^* - t_s) = \gamma(t_e - t^*), \tag{20}$$

$$t_e - t_s = \frac{N}{s}. \tag{21}$$

The two users each control $N/2$ vehicles and schedule the same number of vehicles during the queuing period. Therefore, they must also schedule the same number during $[t_s, t_q]$

$$\int_{t_s}^{t_q} r_A(u)\,du = \int_{t_s}^{t_q} r_B(u)\,du, \tag{22}$$

where

$$r_A(u) + r_B(u) = s, \quad u \in (t_s, t_q). \tag{23}$$

The early departure schedule defined by Equations (22) and (23) is consistent with a PSNE as far as trip timing by individual vehicles. The first vehicle scheduled at t_s creates the same MPC as all vehicles scheduled during $[t_q, t_e]$. Vehicles scheduled during (t_s, t_q) create a lower MPC, so that rescheduling them to any time outside (t_s, t_q) would increase total fleet costs. Rescheduling any vehicle into (t_s, t_q) would also increase fleet costs because it would impose a queuing delay on all vehicles departing later until t_e, and would therefore create a higher MPC than a vehicle departing at t_q. Q.E.D.

Together, Equations (17)–(23) define the candidate PSNE with queuing. Using superscript c to denote this candidate, the critical times are

$$t_s^c = t^* - \frac{\gamma}{\beta+\gamma}\frac{N}{s}, \tag{24a}$$

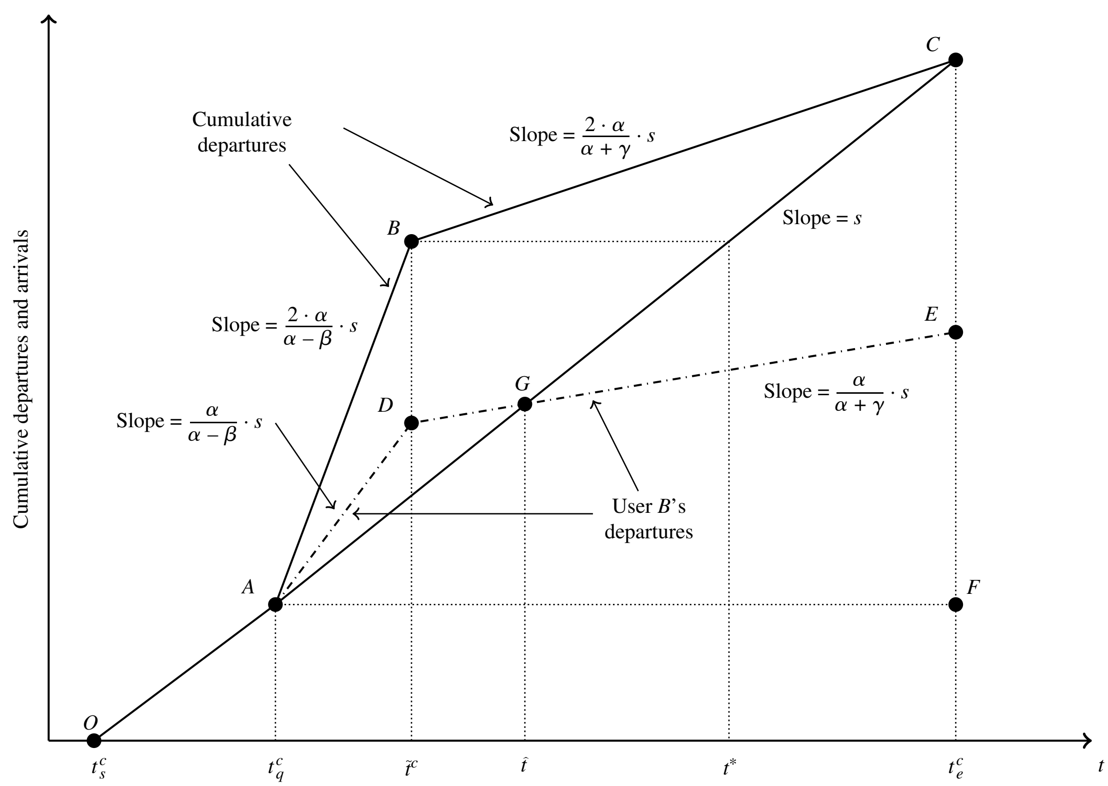
$$t_e^c = t^* + \frac{\beta}{\beta+\gamma}\frac{N}{s}, \tag{24b}$$

$$\tilde{t}^c = t^* - \frac{\beta(\gamma-\alpha)}{2\alpha(\beta+\gamma)}\frac{N}{s}, \tag{24c}$$

$$t_q^c = t^* - \frac{\beta(\gamma-\alpha)}{(\alpha+\beta)(\beta+\gamma)}\frac{N}{s}. \tag{24d}$$

The candidate PSNE is depicted in Figure 1. Cumulative departures during the whole travel period are shown by the piecewise linear schedule $OABC$. During the first interval (t_s^c, t_q^c), the two users depart early at rates consistent with Equation (22). During the remaining interval (t_q^c, t_e^c), user B contributes to the cumulative departures the portion between schedule $ADGE$ and the horizontal line AF. User A contributes the equally big portion between schedule ABC and schedule $ADGE$. During the interval $(t_q^c, \tilde{t}^c)$, each user departs early at rate $\alpha s/(\alpha - \beta)$, and during $(\tilde{t}^c, t_e^c)$ each user departs late at rate $\alpha s/(\alpha+\gamma)$. At time $\hat{t}$, each user's cumulative departures over the interval $(t_q^c, \hat{t})$ match the cumulative bottleneck throughput over the same interval.

Figure 1. Candidate PSNE with $\gamma > \alpha$ and Queuing



We now show that the candidate PSNE just derived is not a PSNE because either user can reduce its fleet costs by rescheduling a positive fraction of its vehicles. This result is formalized as:

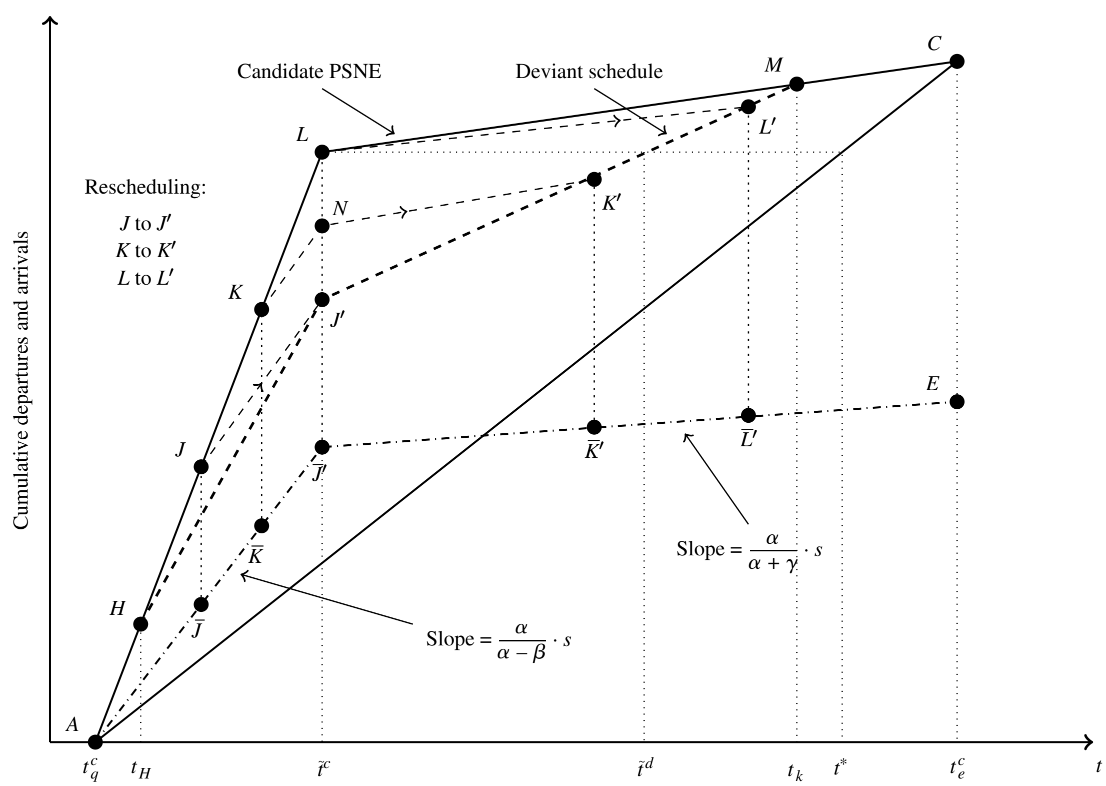
Lemma 4. *Consider two identical atomic users who each simultaneously schedule $N/2$ vehicles with trip-timing preferences for each vehicle defined by the same $\{\alpha, \beta, \gamma, t^*\}$ parameter values. If $\gamma > \alpha$, the unique candidate departure pattern with queuing in which users cannot reduce costs by rescheduling a single vehicle is not a PSNE with respect to rescheduling a positive fraction of the fleet.*

Proof. Figure 2 shows the portion of the candidate PSNE in Figure 1 during which queuing occurs. Points A, C, and E have the same labels as in Figure 1. Suppose user A deviates from the candidate PSNE by departing at a rate below $\alpha/(\alpha-\beta)s$ during the interval $(t_H, \tilde{t}^c)$, and a rate above $\alpha/(\alpha+\gamma)s$ during the interval $(\tilde{t}^c, t_K)$ so that at time t_K cumulative departures have caught up to the candidate PSNE departures. The cumulative departures curve with the deviant schedule is shown by the thick dashed line passing through points H, J', K', L', and M. It lies to the right of the candidate PSNE during the interval (t_H, t_K) so that the queue is shorter, and the on-time departure time is postponed from $\tilde{t}^c$ to $\tilde{t}^d$.

Assume without loss of generality that user A schedules its vehicles in the same order as in the candidate PSNE. When the vehicle originally scheduled to depart at point J departs, $J\bar{J}$ other vehicles in user A's fleet have already left since time t_q^c (recall that the thick dash-dot line passing through points A, $\bar{J}'$, and E depicts user B's cumulative departures). In the deviant schedule this vehicle therefore departs at point J' where distance $J'\bar{J}'$ equals distance $J\bar{J}$. Similarly, the vehicle originally scheduled to depart at point K now departs at point K' where distance $K'\bar{K}'$ equals distance $K\bar{K}$. The vehicle originally scheduled to depart at point L now departs at point L' where distance $L'\bar{L}'$ equals distance $L\bar{J}'$. Retiming of departures for other vehicles can be determined in the same way. We will show that all vehicles originally scheduled to depart after point J and before point L incur lower costs with the deviant schedule than in the candidate PSNE. All of user A's other vehicles experience no change in costs so that user A's total fleet costs fall.

First consider the vehicle that is rescheduled from point J to point J'. The dashed line from J to J' runs parallel to the segment $\bar{J}\bar{J}'$ of user B's cumulative departures curve. The slope of this curve, $\alpha/(\alpha-\beta)s$, matches the equilibrium departure rate for nonatomic users (cf. Equation (2)), for which travel costs of individual vehicles remain constant over time. The rescheduled vehicle thus incurs no change in cost as the decrease in the early arrival cost offsets the increase in queuing cost. For the same reason, all vehicles that were scheduled to depart between H and J incur the same cost as they did before.

Figure 2. Deviation from Candidate PSNE



Next consider the vehicle that is rescheduled from point L to point L'. The dashed line from L to L' runs parallel to the segment $\bar{J}'\bar{L}'$ of user B's cumulative departures curve. Because the slope of this curve, $\alpha/(\alpha+\gamma)s$, again matches the equilibrium departure rate for nonatomic users, the vehicle incurs the same cost as it did at point L. This time it does so by reducing the queuing cost and increasing the late arrival cost by the same amount. Similarly, all vehicles originally scheduled to depart after L incur the same cost as before.

Finally, consider the vehicle originally scheduled to depart at point K which is representative of the vehicles scheduled after J and before L. This vehicle is rescheduled to point K'. The shift from K to K' can be divided into two parts: from K to N, and from N to K'. The shift from K to N leaves the vehicle's cost unchanged since the dashed line from K to N runs parallel to segment $\bar{K}\bar{J}'$ of user B's cumulative departures curve. However, the shift from N to K' causes the vehicle's cost to drop because segment NK' has a slope $\alpha/(\alpha+\gamma)s$, whereas the vehicle still arrives early when it departs at K'. The vehicle benefits both from a reduction in early arrival time and a reduction in queuing delay. For the same reason, all vehicles originally scheduled between points J and L experience a drop in cost.

User A can gain from deviation because user B's departure rate is not constant, but rather drops when vehicles start arriving late in the candidate PSNE. User A takes advantage of this by postponing departures to the period when user B's departure rate is lower. Note that user A would not benefit if all of the vehicles that deviate arrived early both before and after the deviation. Similarly, user A would not benefit if all of these vehicles arrived late both before and after deviation.

A further point to note is that the gain from deviation increase more than proportionally with the number of vehicles that deviate. The candidate PSNE is constructed so that rescheduling a single vehicle is not gainful at all. If user A were to reschedule vehicles within a short interval spanning $\tilde{t}^c$, few vehicles would benefit and the benefit per vehicle would be small because queuing delay would drop only marginally. Rescheduling more vehicles would increase both the number of vehicles that benefit and the benefit per vehicle. In fact, user A's most profitable deviation strategy is to avoid queuing altogether by rescheduling all its vehicles from period $(t_q^c, \hat{t})$ to period $(\hat{t}, t_e^c)$ where $\hat{t}$ is shown in Figure 1. User A stops departing until the queue created by user B has vanished at time $\hat{t}$, and then maintain a departure rate of $\gamma s/(\alpha+\gamma)$ during $(\hat{t}, t_e^c)$. Since user B departs at rate $\alpha s/(\alpha+\gamma)$ during $(\hat{t}, t_e^c)$, the total departure rate during this period equals s and the bottleneck is fully utilized without queuing. Moreover, since user B keeps the bottleneck fully utilized during $(t_q^c, \hat{t})$, all vehicles in both fleets complete their trips by t_e^c. Consequently, with the deviant schedule, all of user A's vehicles can complete their trips within the same time period as in the candidate PSNE, but without queuing.

Another way to see why deviation is gainful is to reconsider the marginal private cost function with queuing in Equation (15)

$$C_A(t) = \alpha T(t) + \beta(t^* - t - T(t)) + \frac{\alpha-\beta}{s}\int_t^{\tilde{t}} r_A(u)\,du + \frac{\alpha+\gamma}{s}\int_{\tilde{t}}^{t_{qe}} r_A(u)\,du.$$

The candidate PSNE is derived while treating $C_A(t)$ as given. Nevertheless, $C_A(t)$ clearly depends on user A's choice of departure schedule. The derivatives with respect to $T(t)$ and $\tilde{t}$ are $\partial C_A(t)/\partial T(t) = \alpha - \beta > 0$, and $\partial C_A(t)/\partial \tilde{t} = -((\beta+\gamma)/s) r_A(\tilde{t}) < 0$, respectively. This suggests that user A can lower its fleet costs by reducing queuing delay, $T(t)$, and postponing the time at which vehicles start arriving late, $\tilde{t}$. Both observations are consistent with the demonstration in Figure 2. Q.E.D.

Proposition 1 establishes that a PSNE does not exist when $\gamma > \alpha$. As noted in Section 1, most empirical studies of scheduling preferences for automobile drivers have obtained estimates that satisfy this inequality. However, there is little evidence either on the trip-timing preferences of users that may control large shares of road traffic (e.g., freight shippers), or on preferences for travel by other modes of transportation. Daniel and Harback (2008), without making explicit the role of the passengers' valuation of time, estimate that $\gamma < \alpha$ holds for many U.S. airlines. Thus, it is of interest to study the existence and nature of PSNE when $\gamma \le \alpha$.

3.2. Existence and Nature of PSNE with $\gamma \le \alpha$

In this section we establish two results for the case $\gamma \le \alpha$. First, we show that there exists a unique PSNE in the aggregate departure schedule that coincides with the system optimum (Proposition 2). Second, we show that the two users' individual departure rates are not uniquely defined in the PSNE (Proposition 3), and the users can incur different fleet costs (Section 3.2.3).

3.2.1. Existence of PSNE with $\gamma \le \alpha$.

Proposition 2. *Consider two identical atomic users who each simultaneously schedule $N/2$ vehicles with trip-timing preferences for each vehicle defined by the same $\{\alpha, \beta, \gamma, t^*\}$ parameter values. If $\gamma \le \alpha$, multiple PSNE exist. All of them have an aggregate departure schedule that coincides with the system optimum (i.e., $r_A(t) + r_B(t) = s$, $t \in (t_s^o, t_e^o)$). During the late-arrival period, each user's equilibrium departure rate is bounded in the range $r_i(t) \in [(\gamma/(\alpha+\gamma))s, (\alpha/(\alpha+\gamma))s]$, $t \in (t^*, t_e^o)$, $i = A, B$. During the early-arrival period, each user's equilibrium departure rate is bounded next:*

$r_i(t) \geq r_E$, $t \in (t_s^o, t^*)$, $i = A, B$, *where* $r_E \equiv ((2\alpha(\beta+\gamma) + \gamma(\beta+2\gamma) - \sqrt{\beta^2\gamma^2 + 4\alpha(\beta+\gamma)^2(\alpha+\gamma)})/(\gamma(\alpha+\gamma))) \cdot (s/2) \in (0, (\gamma/(\alpha+\gamma))s)$.

Proof. The proof entails establishing four results: (1) A PSNE with queuing does not exist. (2) All PSNE without queuing must coincide with the system-optimal departure pattern given by Equations (7a)–(7c). (3) Given the lower bound on individual departure rates for late arrivals, neither user can gain by rescheduling a single vehicle. (4) Given the lower bounds on individual departure rates for early and late arrivals, neither user can gain by rescheduling part of its fleet.

Result 1. A PSNE with queuing does not exist.

By Lemma 2, mass departures cannot be part of a PSNE. Any candidate PSNE with queuing must satisfy conditions (17)–(19). With $\gamma \leq \alpha$, these conditions cannot all be satisfied since the aggregate early departure rate exceeds capacity, and the aggregate late departure rate is no less than capacity. Hence any queue cannot dissipate while users are departing, which is inconsistent with a PSNE.

Result 2. The system-optimal departure pattern is the only possible PSNE in the aggregate departure schedule.

Given Result 1, in equilibrium, the aggregate departure rate cannot exceed the bottleneck capacity when $\gamma \leq \alpha$. Therefore, equilibrium departures must occur at rate s over a connected time interval, otherwise either user could reduce its fleet costs by rescheduling vehicles into "gaps" in the departure schedule when bottleneck capacity is not fully utilized. The departure period, $[t_s, t_e]$, must be as given by Equations (7b) and (7c) since otherwise trip costs at t_s and t_e would differ, and at least one user could reduce its fleet costs by rescheduling vehicles from the higher-cost endpoint to the lower-cost endpoint.

Result 3. A user cannot gain by rescheduling a single vehicle if each user's departure rate satisfies the conditions in Proposition 2.

In the candidate PSNE with a system-optimal aggregate departure pattern, there is no queuing but the bottleneck is used to capacity. As in the proof of Lemma 1, it is therefore necessary to distinguish between the cost saved by removing a vehicle from the departure schedule (which does not affect other vehicles' costs) and the cost of adding a vehicle (which creates a queue unless the vehicle is added at t_e^o). The respective costs are

$$C_i^-(t) = \begin{cases} \beta(t^*-t), & t \in [t_s^o, t^*] \\ \gamma(t-t^*), & t \in [t^*, t_e^o], \end{cases}$$

$$C_i^+(t) = \begin{cases} \beta(t^*-t) + \dfrac{\alpha-\beta}{s}\displaystyle\int_t^{t^*} r_i(u)\,du \\ \quad + \dfrac{\alpha+\gamma}{s}\displaystyle\int_{t^*}^{t_e^o} r_i(u)\,du, & t \in [t_s^o, t^*], \\ \gamma(t-t^*) + \dfrac{\alpha+\gamma}{s}\displaystyle\int_t^{t_e^o} r_i(u)\,du, & t \in [t^*, t_e^o]. \end{cases}$$

A vehicle can be rescheduled in four ways: (i) late to late, (ii) late to early, (iii) early to late, and (iv) early to early. Consider each possibility in turn.

(i) *Rescheduling late to late*: Rescheduling a late vehicle to a later time is never beneficial because the vehicle's trip cost increases, and other vehicles do not gain. Suppose a vehicle is rescheduled earlier from t to t' where $t^* \leq t' < t$. The change in fleet costs is given by Equation (11)

$$\begin{aligned} \Delta C_i &= -C_i^-(t) + C_i^+(t') \\ &= -\gamma(t-t') + \frac{\alpha+\gamma}{s}\int_{t'}^{t} r_i(u)\,du \stackrel{s}{=} \lambda_{t',t}^i - \frac{\gamma}{\alpha+\gamma}, \end{aligned}$$

where $\stackrel{s}{=}$ means identical in sign. Given $r_i(t) \geq \gamma s/(\alpha+\gamma)$ for $t \in (t^*, t_e^o)$, $\lambda_{t',t}^i \geq \gamma/(\alpha+\gamma)$, $\Delta C_i \geq 0$, and the deviation is not beneficial.

(ii) *Rescheduling late to early*: Rescheduling a late vehicle to an early time is clearly inferior to rescheduling it to t^* because the vehicle incurs an early-arrival cost and creates a queue for a longer period. Yet rescheduling it to t^* is not beneficial as per case (i).

(iii) *Rescheduling early to late*: The best option in this case is to reschedule a vehicle from t_s^o. However, the gain is the same as for rescheduling a vehicle from t_e^o, and this is not beneficial as per case (i). Rescheduling early to late therefore cannot be beneficial.

(iv) *Rescheduling early to early*: The best option in this case is to reschedule a vehicle from t_s^o to t^*, but this, too, is not beneficial for the same reason as in case (iii).

This establishes that the candidate PSNE in Proposition 2 is robust to deviations in which a single vehicle is rescheduled.

Result 4. A user cannot gain by rescheduling a positive measure of its fleet.

If user A reschedules vehicles to depart outside $[t_s^o, t_e^o]$, its fleet costs necessarily increase since the rescheduled vehicles experience greater schedule delay costs without benefiting the rest of the fleet. If user A instead reschedules vehicles to depart inside $[t_s^o, t_e^o]$, queuing occurs. The optimal departure rate in the presence of a queue was derived in Section 3.1 and, although with $\gamma \leq \alpha$ a PSNE cannot exhibit queuing, we need to check if a deviation from the candidate to a setting with queuing is cost reducing. For early arrivals, user A is willing to depart at a positive and finite rate only if condition (16) is satisfied; i.e., $r_B(t) = \alpha s/(\alpha-\beta) > s$. Because user B departs at a rate less than

s in the candidate PSNE, user A is better off scheduling the vehicles later. For late arrivals, user A is willing to depart at a positive and finite rate only if condition (14) is satisfied so that $r_B(t) = \alpha s/(\alpha + \gamma)$. Proposition 2 stipulates that $r_B(t) \le \alpha s/(\alpha+\gamma)$. If $r_B(t) = \alpha s/(\alpha+\gamma)$, and therefore $r_A(t) = \gamma s/(\alpha+\gamma)$, user A is indifferent between departing or not so that the deviation does not reduce its fleet costs. If $r_B(t) < \alpha s/(\alpha+\gamma)$, user A is better off scheduling the vehicles later. The only remaining possibility for gainful deviation is one that involves mass departures. In Online Appendix B we prove that a deviation with mass departures is not gainful if the lower bounds on early and late departure rates stated in Proposition 2 are both satisfied. Q.E.D.

The intuition for Proposition 2 is as follows. The only way for user A to gainfully deviate from the candidate PSNE by rescheduling a single vehicle is to advance its departure during the late arrival period. Doing so reduces the vehicle's late-arrival cost, but imposes queuing delay on user A's vehicles that depart later. The trade-off is not worthwhile if enough vehicles in A's fleet have yet to depart, and queuing is sufficiently costly relative to late arrival. The lower bound on the late departure rate stated in Proposition 2 assures this condition is met. If user A deviates by rescheduling vehicles in mass departures and imposes queuing delays on its other vehicles, the same argument holds. The trade-off is not worthwhile because the lower bound on the late departure rate ensures that there are enough vehicles in the fleet yet to depart that will be negatively affected.

To avoid delaying other vehicles in its fleet, user A must advance departures for vehicles that participate in a mass. Because vehicles in the mass suffer queuing delay, user A can benefit from such a deviation only if the vehicles' schedule delay costs are reduced enough. This requires that the vehicles were departing late over a period longer than the time they take to pass the bottleneck when in the mass. This, in turn, is possible only if user B occupies a large enough share of bottleneck capacity during the candidate PSNE. The lower bound on the late departure rate identified in Proposition 2 ensures that this condition is not met.

To see why a lower bound on the early departure rate is also required, suppose that user A does not depart during some time interval (t, t^*). User A can then reschedule its vehicles departing in some interval (t^*, t') by launching them in a mass at a time $t_m \in (t, t^*)$. None of its fleet departing after t' will be delayed by the mass. If t_m is chosen to minimize the total schedule delay costs of vehicles in the mass, their total costs will fall. However, if user A is scheduling enough departures during (t_m, t^*), the deviation will not be gainful since the mass departure either imposes queuing delays on its other vehicles, or it has to include vehicles that were arriving early, which will suffer higher early arrival costs and queuing costs. If the early departure rate is high enough, such a mass departure is unfavorable. Proposition 2 identifies a minimum early departure rate to guarantee this when the late departure rate is held fixed at its minimum value: $\gamma s/(\alpha+\gamma)$.

3.2.2. Nonuniqueness of PSNE with $\gamma \le \alpha$. Although the aggregate departure schedule in the PSNE described in Proposition 2 is unique when $\gamma \le \alpha$, many pairs of departure schedules $\{r_A(\cdot), r_B(\cdot)\}$ are consistent with the aggregate pattern. Thus, the PSNE is not unique in terms of individual departure rates. This result is formalized in the following proposition:

Proposition 3. *Consider two identical atomic users who each simultaneously schedule $N/2$ vehicles with trip-timing preferences for each vehicle defined by the same $\{\alpha, \beta, \gamma, t^*\}$ parameter values. During the late departure period $t \in (t^*, t_e^o)$, the two users depart at the same rate $r_A(t) = r_B(t) = s/2$ if $\alpha = \gamma$, but they can depart at different rates if $\gamma < \alpha$. During the early departure period $t \in (t_s^o, t^*)$, a continuum of departure schedules $\{r_A(\cdot), r_B(\cdot)\}$ are consistent with the PSNE.*

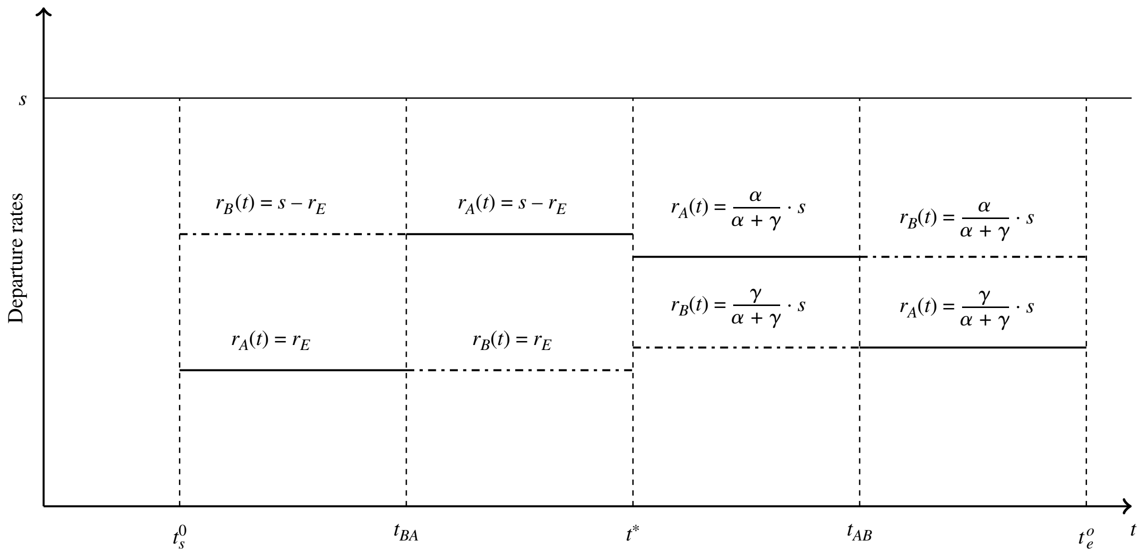
Proof. By Proposition 2, for $t \in (t^*, t_e^o)$, $\gamma/(\alpha+\gamma) \le r_A(t)/s \le \alpha/(\alpha+\gamma)$ and $r_B(t) = s - r_A(t)$. If $\gamma = \alpha$, $\gamma/(\alpha+\gamma) = 1/2$ and therefore $r_A(t) = r_B(t) = s/2$. If $\gamma < \alpha$, there is a continuum of pairs of $\{r_A(t), r_B(t)\}$ that satisfy the equilibrium condition while assuring that all vehicles depart during the interval $[t_s^o, t_e^o]$. For $t \in (t_s^o, t^*)$, the constraint $r_i(t) \ge r_E$ is less strict because $r_E < s/2$. Any pair of departure schedules $\{r_A(\cdot), r_B(\cdot)\}$ that satisfies $r_E \le r_A(t) \le s - r_E$ and $r_B(t) = s - r_A(t)$ satisfies the aggregate PSNE condition. Q.E.D.

The nonuniqueness of individual departure rates in a PSNE implies that, unlike in the model for identical nonatomic users, users can experience different fleet costs in a PSNE. This prospect is examined further in Section 3.2.3.

3.2.3. Asymmetric Equilibrium Costs with $\gamma \le \alpha$. Clearly, if users A and B depart at the same rate throughout the departure period they incur the same fleet costs. In addition, there are many asymmetric PSNE departure schedules that result in the same average schedule delay costs for the two users and hence the same fleet costs. However, there are also many asymmetric departure schedules that result in different fleet costs.

Figure 3 depicts an illustrative example of a PSNE in which user A incurs lower fleet costs than user B. In the example, user A's departures are concentrated near t^* so that user A has lower average schedule delay costs than B. User A's departure rate is shown by solid lines and user B's by thick dashed lines. During the interval (t_s^o, t_{BA}), user A departs at the minimum rate r_E defined in Proposition 2. Throughout the next interval

Figure 3. A PSNE with Asymmetric Costs: $\gamma \leq \alpha$, No Queuing



(t_{BA}, t^*), user B departs at rate r_E and user A at the complementary rate $s - r_E$. During the first part (t^*, t_{AB}) of the late-arrival period, user A departs at the maximum rate consistent with a PSNE for user B, $\alpha s/(\alpha+\gamma)$. During the remaining part (t_{AB}, t_e^o) of the departure period, user A departs at the minimum rate consistent with a PSNE for itself, $\gamma s/(\alpha+\gamma)$. The transition times t_{BA} and t_{AB} are such that each user dispatches a total of $N/2$ vehicles. Expressed in terms of user A's fleet, the requisite condition is

$$r_E(t_{BA} - t_s^o) + (s - r_E)(t^* - t_{BA}) + \frac{\alpha}{\alpha+\gamma} s (t_{AB} - t^*) + \frac{\gamma}{\alpha+\gamma} s (t_e^o - t_{AB}) = \frac{N}{2}. \tag{25}$$

Total costs in the PSNE, TC^o, are a given; i.e., independent of users A's and B's individual departure rates. The maximum difference between the users' costs, $TC_B - TC_A$, can therefore be found by minimizing TC_A with respect to t_{BA} and t_{AB}. As described in Online Appendix C, user A's costs as a fraction of total costs work out to

$$f \equiv \frac{TC_A}{TC^o} = \frac{3+z}{4} - \frac{1}{2}\sqrt{\frac{\alpha}{\alpha+\gamma} + z^2},$$

where $z \equiv \beta\gamma/((\alpha+\gamma)(\beta+\gamma))$. The fraction f depends on α, β, and γ only through the ratios $\beta/\alpha < 1$ and $\gamma/\alpha \leq 1$. It is readily shown that f is a monotonically increasing function of β/α and γ/α. At the upper limits with $\beta/\alpha \cong 1$ and $\gamma/\alpha = 1$, $f \cong 0.4535$. As β/α and γ/α approach their lower limits of 0, f approaches 1/4.

This example illustrates that although there exists a unique PSNE in terms of the aggregate departure rate, individual user departure rates can differ substantially and so can their costs. If the unit schedule delay cost parameters, β and γ, are small compared with the cost of travel time, α, one user's fleet costs can be as little as one-third the other user's costs. This is because one user can concentrate its departures around t^* without the other user wanting to reschedule its fleet because the gain from reducing schedule delay costs would be outweighed by the high costs of queuing delay. This suggests that equity of access to a bottleneck can be an issue with atomic users.

3.3. Extension to Multiple Users

The analysis in this section can be generalized to $m > 2$ homogeneous users. Propositions 1 and 2 can be extended in a straightforward manner by following similar lines of reasoning. For example, condition (12) for existence of a PSNE without queuing, $\lambda_{t,t'}^i \geq \gamma/(\alpha+\gamma)$, still holds. In the least restrictive case in which all m users depart at equal rates, this condition implies that $1/m \geq \gamma/(\alpha+\gamma)$, or $(m-1)\gamma \leq \alpha$. The intuition is similar to the case with $m = 2$. In a symmetric candidate PSNE, each user departs at rate s/m and contributes a fraction $1/m$ of the traffic. If user i reschedules one vehicle ΔT units of time earlier, it reduces the vehicle's late arrival cost by $\gamma\Delta T$. The vehicle imposes a delay of $1/s$ on the $(s/m)\Delta T$ of user i's other vehicles that depart during the time interval. The unit cost of the delay is $\alpha + \gamma$. User i benefits from rescheduling the vehicle unless $\gamma\Delta T \leq (\alpha+\gamma)\Delta T/m$, which is equivalent to $(m-1)\gamma \leq \alpha$.

The nonexistence of a PSNE with queuing can also be established using the same approach for $m > 2$ as for $m = 2$. Candidate equilibrium departure rates for other users during early and late departures are still given by Equations (16) and (14), respectively. The total departure rate is therefore $m\alpha s/((m-1)(\alpha-\beta))$ for early departures, and $m\alpha s/((m-1)(\alpha+\gamma))$ for late departures. The candidate equilibrium is well defined and unique, and it approaches the nonatomic PSNE as $m \to \infty$. Nevertheless, the candidate equilibrium breaks down because any user can reduce its costs by rescheduling a portion of its fleet. Using a diagram similar to Figure 2, it is straightforward to show that any user can gainfully deviate by postponing departures for a portion of its fleet during an interval spanning the on-time departure time $\tilde{t}^c$.

Clearly, the condition $(m-1)\gamma \le \alpha$ becomes more stringent as m increases so that a PSNE without queuing becomes progressively less plausible. Indeed, given any fixed values of $\gamma > 0$ and $\alpha > 0$, condition $(m-1)\gamma \le \alpha$ necessarily fails if m is large enough. Thus, as $m \to \infty$ the bottleneck model with atomic users does not converge in behavior to the nonatomic model which has a unique PSNE. The reason for this divergence is that even for large values of m, users in the atomic model take into account how they affect queuing times. The absence of PSNE occurs because users can do marginal and "large changes" (where they change the departures of a positive fraction of the fleet); no matter how diminutive the users are, these two options exist. Only with a continuum of users are larger changes not possible. The gains from deviation per user decrease in magnitude as m rises, but they remain positive. By contrast, in the standard bottleneck model users treat queuing times as given.

In this section we have shown that a PSNE may not exist when symmetric atomic users schedule vehicles simultaneously. In particular, if there are two identical users and $\gamma > \alpha$, a PSNE does not exist. By contrast, if $\gamma \le \alpha$, multiple PSNE exist and in all of them no queuing occurs. We now turn to the case where users are not identical and study the existence and uniqueness of PSNE.

4. Nonuniqueness of Equilibrium with Heterogeneous Atomic Users

In this section we consider a modified version of the model in which users A and B differ in their preferred arrival times and can have fleets of different size. Arnott, de Palma, and Lindsey (1987) studied this setting for the case of nonatomic users, and we draw on some of their results. We show that, regardless of the relative size of parameters α and γ, a multiplicity of PSNE without queuing can exist with two atomic users if preferred arrival times differ within a certain range. The PSNE differ both in the timing of aggregate departures and total costs. Departures can begin earlier than, later than, or at the same time as in the system optimum. Thus, a PSNE can be inefficient even if there is no queuing.

Suppose user i, $i = A, B$, has a fleet of N_i vehicles, each with a preferred arrival time of t_i^*. Assume without loss of generality that $t_B^* > t_A^*$. To begin, assume $t_B^* \gg t_A^*$ so that the users can schedule their fleets at individually optimal time windows that do not overlap. Let t_{is} and t_{ie} denote the first and last departure times for user i, and let superscript d denote the PSNE with disjoint arrival times. Each user departs at a rate s during the interval that minimizes the total schedule delay costs of its fleet

$$(t_{As}^d, t_{Ae}^d) = \left(t_A^* - \frac{\gamma}{\beta+\gamma}\frac{N_A}{s}, t_A^* + \frac{\beta}{\beta+\gamma}\frac{N_A}{s}\right), \tag{26a}$$

$$(t_{Bs}^d, t_{Be}^d) = \left(t_B^* - \frac{\gamma}{\beta+\gamma}\frac{N_B}{s}, t_B^* + \frac{\beta}{\beta+\gamma}\frac{N_B}{s}\right). \tag{26b}$$

To rule out this uninteresting case we hereafter assume $t_{Bs}^d < t_{Ae}^d$, which is equivalent to assuming that

$$t_B^* - t_A^* < \frac{\beta}{\beta+\gamma}\frac{N_A}{s} + \frac{\gamma}{\beta+\gamma}\frac{N_B}{s}. \tag{27}$$

Define the auxiliary variable

$$x \equiv \frac{\beta}{\beta+\gamma}\frac{N_A}{s} + \frac{\gamma}{\beta+\gamma}\frac{N_B}{s} - (t_B^* - t_A^*) > 0. \tag{28}$$

Variable x measures how much the users' preferred departure schedules overlap, and thus the degree of "conflict" between them.

In addition to condition (27) we assume

$$t_B^* - t_A^* > \left\| \frac{\gamma}{\beta+\gamma}\frac{N_B}{s} - \frac{\beta}{\beta+\gamma}\frac{N_A}{s} \right\|. \tag{29}$$

Condition (29) assures that in the system optimum, derived in Section 4.1, some of user A's fleet arrives late and some of user B's fleet arrives early.

4.1. System Optimum

In the system optimum, vehicles depart at an aggregate rate of s over a connected time interval. In general, the optimal individual departure rates, $r_A^*(t)$ and $r_B^*(t)$, are not unique. However, given condition (29), there is a unique optimum in which all of user A's fleet is dispatched before any of user B's fleet. Some of user A's fleet arrives late, and some of user B's fleet arrives early. Let t_{As} be the time at which user A starts to depart. User A departs over the interval $(t_{As}, t_{As} + N_A/s)$, and user B departs over the interval $(t_{As} + N_A/s, t_{As} + (N_A + N_B)/s)$. Total costs are

$$\begin{aligned} TC = {} & \frac{\beta s}{2}(t_A^* - t_{As})^2 + \frac{\gamma s}{2}\left(t_{As} + \frac{N_A}{s} - t_A^*\right)^2 \\ & + \frac{\beta s}{2}\left(t_B^* - \left(t_{As} + \frac{N_A}{s}\right)\right)^2 \\ & + \frac{\gamma s}{2}\left(t_{As} + \frac{N_A + N_B}{s} - t_B^*\right)^2. \end{aligned} \tag{30}$$

The first-order condition for minimizing TC with respect to t_{As} yields

$$t_{As}^{o} = \frac{t_A^* + t_B^*}{2} - \frac{\gamma}{2(\beta+\gamma)}\frac{N_A + N_B}{s} - \frac{1}{2}\frac{N_A}{s}, \tag{31}$$

where superscript o again refers to the system optimum. Given condition (29), $t_A^* < t_{Ae}^o < t_B^*$.

Substituting Equation (31) into Equation (30), total costs in the system optimum can be written as

$$TC^o = \frac{\delta}{2}\frac{N_A^2 + N_B^2}{s} + \frac{\beta+\gamma}{4}sx^2, \tag{32}$$

where x is defined in Equation (28). The first term in Equation (32) equals total costs if condition (27) did not hold and the two users traveled in disjoint intervals. The second term in Equation (32) is the additional costs incurred because of overlap in the users' preferred arrival times. This term is an increasing, quadratic function of the degree of conflict, x.

4.2. No-Toll Equilibrium with Heterogeneous Nonatomic Users

Before considering the PSNE with users A and B, we briefly discuss the analogous no-toll PSNE with heterogeneous nonatomic users. In this case there is a measure N_A of nonatomic users with a preferred arrival time t_A^*, and a measure N_B of nonatomic users with a preferred arrival time t_B^*. Arnott, de Palma, and Lindsey (1987) show that when conditions (27) and (29) both hold, the equilibrium queuing pattern has two peaks. The first peak corresponds to on-time arrival for users with preferred arrival time t_A^*, and the second peak to on-time arrival for users with preferred arrival time t_B^*. Total costs are

$$TC^n = \delta\frac{N_A^2 + N_B^2}{s} + \frac{\beta N_A + \gamma N_B}{2}x. \tag{33}$$

The first term in (33) gives total costs if the equilibrium departure schedules of the two user groups do not conflict. This term is twice the first term in (32), just as total costs with homogeneous users in Equation (6) are twice the total costs in the system optimum in Equation (8). The second term in (33) is a linear function of x. This contrasts with the second term in (32), which is a quadratic function of x. Thus, a small degree of conflict between the two user groups raises total costs disproportionately more in the no-toll equilibrium than in the system optimum. In Section 4.3, we use the difference between total costs in the no-toll equilibrium and system optimum as a metric to assess the efficiency achieved from self-internalization of congestion by atomic users.

4.3. PSNE

A PSNE without queuing exists if $t_{Ae} < t_B^*$ since the two users then do not arrive late at the same time and do not compete to complete their trips as soon as possible. Indeed, a continuum of PSNE exists. To see this, assume that both users depart in the period $[t_A^*, t_{Ae}]$ and capacity is fully used. Suppose user A advances a vehicle's departure from t_2 to t_1, where $t_A^* \le t_1 < t_2 \le t_{Ae}$. User A's fleet costs change by

$$\begin{aligned}\Delta C_A &= -C_A^-(t_2) + C_A^+(t_1)\\ &= -\gamma(t_2 - t_A^*) + \gamma(t_1 - t_A^*) + \frac{\alpha+\gamma}{s}\int_{t_1}^{t_2} r_A(u)\,du\\ &= ((\alpha+\gamma)\lambda_{t_1,t_2}^A - \gamma)(t_2 - t_1).\end{aligned}$$

User A does not benefit from the rescheduling if

$$\lambda_{t_1,t_2}^A \ge \frac{\gamma}{\alpha+\gamma}.$$

This condition is satisfied for any choice of t_1 and t_2 as long as

$$r_A(t) \in \left[\frac{\gamma}{\alpha+\gamma}s, s\right], \quad t \in [t_A^*, t_{Ae}]. \tag{34}$$

User B's departure rate is then $r_B(t) = s - r_A(t) \in [0, \alpha s/(\alpha+\gamma)]$. There is no restriction on $r_B(t)$ because user B arrives early throughout $[t_A^*, t_{Ae}]$. Thus, any departure profile satisfying Equation (34) is consistent with a PSNE. However, different departure profiles generally result in different total costs.

An exhaustive treatment of all PSNE in this example would be tedious. Attention is limited to two illustrative cases. In Case 1, user A schedules its entire fleet during the same interval it would choose if user B did not exist. Hence, $r_A(t) = s$ during the interval (t_{As}^d, t_{Ae}^d) given by Equation (26a). User B departs at rate s during the ensuing interval (t_{Ae}^d, t_{Be}). In Case 2, user A departs at rate s during the interval (t_{As}, t_A^*). During the next interval (t_A^*, t_{Ae}), user A departs at the minimum rate consistent with Equation (34), $r_A(t) = \gamma s/(\alpha+\gamma)$. User B departs at the complementary rate $r_B(t) = \alpha s/(\alpha+\gamma)$. During the final interval (t_{Ae}, t_{Be}), user B departs at rate s. Cases 1 and 2 are now examined in turn.

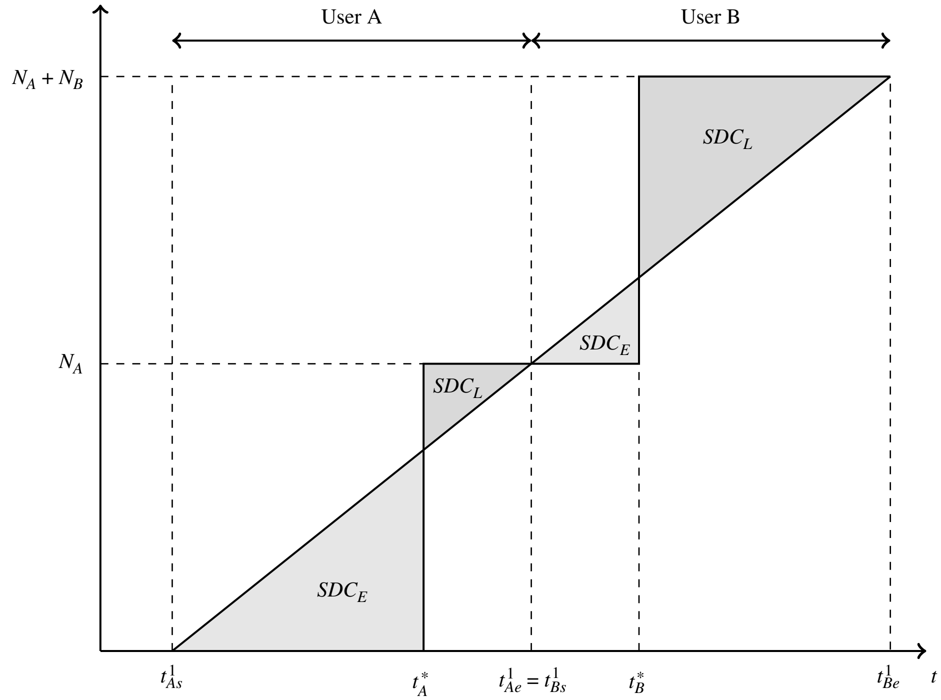
4.3.1. PSNE for Case 1. Figure 4 depicts a PSNE conforming with Case 1, using the following parameter values: $N_A = 1$, $N_B = 1$, $s = 1$, $\alpha = 2$, $\beta = 1$, $\gamma = 2$, $t_A^* = 0$, and $t_B^* = 2/3$. Because all of user A's fleet departs before any of user B's fleet

$$t_{Ae}^1 - t_{As}^1 = \frac{N_A}{s}, \tag{35}$$

where superscript 1 denotes Case 1. User A can schedule a vehicle either just before t_{As}^1 or just after t_{Ae}^1 without imposing a congestion delay on other vehicles in its fleet. The first and last vehicles must therefore incur equal schedule delay costs

$$\beta(t_A^* - t_{As}^1) = \gamma(t_{Ae}^1 - t_A^*). \tag{36}$$

Figure 4. PSNE for Case 1 with $t_B^* > t_A^*$, $N_A = 1$, $N_B = 1$, $s = 1$, $\alpha = 2$, $\beta = 1$, $\gamma = 2$, $t_A^* = 0$, and $t_B^* = 2/3$



Together, Equations (35) and (36) imply

$$t_{As}^1 = t_A^* - \frac{\gamma}{\beta+\gamma}\frac{N_A}{s}, \tag{37a}$$

$$t_{Ae}^1 = t_A^* + \frac{\beta}{\beta+\gamma}\frac{N_A}{s}. \tag{37b}$$

As noted previously, the departure interval (t_{As}^1, t_{Ae}^1) coincides with the interval (t_{As}^d, t_{Ae}^d) given by Equation (26a). User B departs in a connected time interval immediately after user A. Hence,

$$t_{Bs}^1 = t_{Ae}^1 = t_A^* + \frac{\beta}{\beta+\gamma}\frac{N_A}{s}, \tag{38a}$$

$$t_{Be}^1 = t_A^* + \frac{\beta}{\beta+\gamma}\frac{N_A}{s} + \frac{N_B}{s}. \tag{38b}$$

A notable feature of the departure schedule given by Equations (37a)–(38b) is that it does not depend on t_B^*. To be consistent with the assumption that some of user B's fleet arrives early we require $t_{Ae}^1 < t_B^*$, or

$$t_B^* - t_A^* > \frac{\beta}{\beta+\gamma}\frac{N_A}{s}. \tag{39}$$

Case 1 is a PSNE if neither user has an incentive to deviate from it. User A clearly has no incentive to deviate because its fleet costs are the minimum possible. Two conditions must be satisfied for user B. First, user B cannot gain by rescheduling its last vehicle from t_{Be}^1 to some time $\hat{t}$ between t_{As}^1 and t_{Ae}^1. This condition is satisfied since rescheduling would impose queuing delay on all vehicles that depart after $\hat{t}$ without reducing schedule delay costs for user B's fleet.

Second, user B cannot gain by rescheduling its last vehicle from t_{Be}^1 to just before the beginning of the travel period at t_{As}^1. The condition is $\beta(t_B^* - t_{As}^1) > \gamma(t_{Be}^1 - t_B^*)$, which reduces to

$$t_B^* - t_A^* \geq \frac{\gamma}{\beta+\gamma}\frac{N_B}{s}. \tag{40}$$

Conditions (39) and (40) together guarantee that condition (29) is satisfied.

The timing of departures in Case 1 and the system optimum can be compared using Equations (37a) and (31)

$$t_{As}^1 - t_{As}^o = \frac{1}{2}\left(\frac{\beta}{\beta+\gamma}\frac{N_A}{s} + \frac{\gamma}{\beta+\gamma}\frac{N_B}{s} - (t_B^* - t_A^*)\right) = \frac{x}{2}.$$

Because $x > 0$, $t_{As}^1 > t_{As}^o$: departures in Case 1 begin later than in the system optimum. To see why, note that while early and late arrivals are balanced optimally for user A, user B's arrivals begin inefficiently late because user A occupies the bottleneck until t_{Ae}^1 and effectively squeezes user B out. As shown in Online Appendix D, total costs in Case 1 can be written as

$$\begin{aligned} TC^1 &= \frac{\delta}{2}\frac{N_A^2 + N_B^2}{s} + \frac{\beta+\gamma}{2}sx^2 \\ &= TC^o + \frac{\beta+\gamma}{4}sx^2. \end{aligned} \tag{41}$$

Total costs are an increasing quadratic function of the degree of conflict, x. This is similar to the system optimum (cf. Equation (32)), but unlike the nonatomic equilibrium where the dependence is linear (cf. Equation (33)). The last expression in Equation (41), $(\beta+\gamma)\cdot sx^2/4$, measures inefficiency of the PSNE in Case 1. This inefficiency can be expressed as a price of anarchy, PA, using the ratio

$$PA = \frac{TC^1 - TC^o}{TC^o}. \tag{42}$$

The price of anarchy is bounded by 0. As shown in Online Appendix D, it is bounded above by 1/5. Thus, despite the fact that there is no queuing in the PSNE for Case 1, total costs can exceed the system-optimal level by up to 20%.

Efficiency of the PSNE can also be measured relative to the nonatomic equilibrium using the index

$$w \equiv \frac{TC^n - TC^1}{TC^n - TC^o}. \tag{43}$$

The numerator of (43) is the reduction in total costs achieved by self-internalization of congestion by the atomic users. The denominator is the reduction in total costs realized at the system optimum. Index w reaches its maximum value of 1 when $x = 0$. Online Appendix D establishes that it is bounded below by 6/7. Self-internalization of congestion by the atomic users independently therefore achieves most of the potential benefits from coordinating departure times centrally.

4.3.2. PSNE for Case 2. Figure 5 depicts an example of a PSNE in Case 2 using the same parameter values as for Figure 4. During the initial interval, (t_{As}^2, t_A^*), user A departs at rate s. User B begins to depart immediately after t_A^*, and during the interval (t_A^*, t_{Ae}^2) the two users depart simultaneously with an aggregate rate of s. User A departs at rate $r_A(t) = \gamma s/(\alpha+\gamma)$, and user B departs at rate $r_B(t) = \alpha s/(\alpha+\gamma)$. During the final interval, (t_{Ae}^2, t_{Be}^2), user B departs at rate s.

The PSNE for Case 2 is solved in Online Appendix D. The three transition times work out to

$$t_{As}^2 = t_A^* - \frac{\alpha+\gamma}{\alpha+\beta+\gamma}\frac{N_A}{s}, \tag{44a}$$

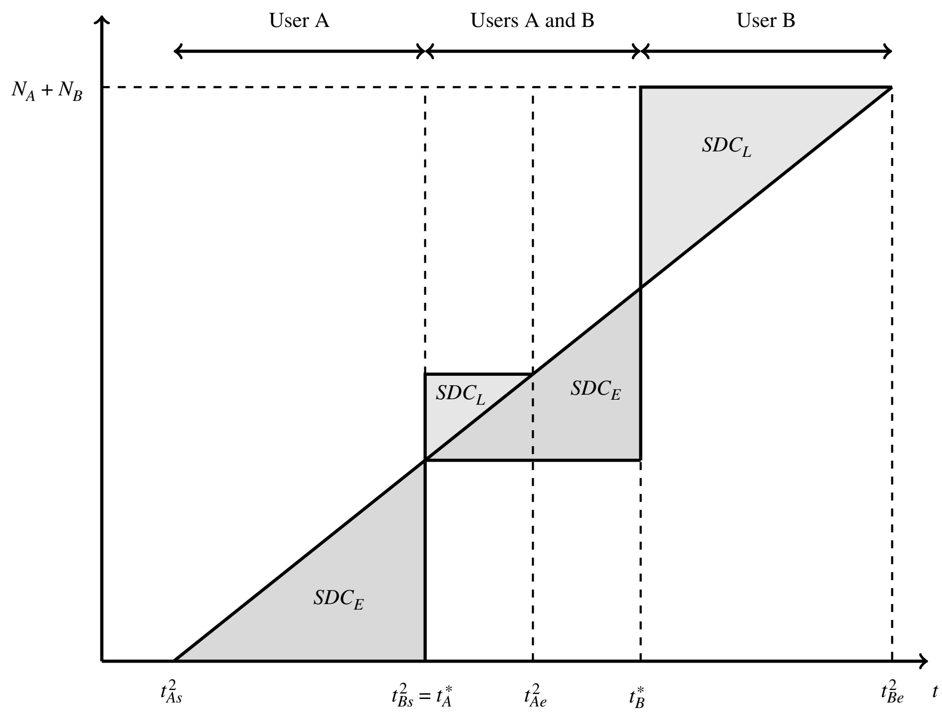
$$t_{Ae}^2 = t_A^* + \frac{\beta(\alpha+\gamma)}{\gamma(\alpha+\beta+\gamma)}\frac{N_A}{s}, \tag{44b}$$

$$t_{Be}^2 = t_A^* + \frac{\beta}{\alpha+\beta+\gamma}\frac{N_A}{s} + \frac{N_B}{s}. \tag{44c}$$

Similar to Case 1, the timing of the PSNE given by Equations (44a)–(44c) does not depend on t_B^*. As shown in Online Appendix D, departures in the PSNE of Case 2 can begin either earlier or later than in the system optimum. As t_B^* decreases, the system optimum begins earlier (see Equation (31)), whereas the PSNE does not (see Equation (44a)). Hence, as user B's preferred schedule moves closer to user A's and the conflict between them grows, the PSNE in Case 2 shifts from beginning too early to beginning too late.

Unlike for Case 1, there is no simple formula for total costs in Case 2. Also unlike Case 1, total costs do

Figure 5. PSNE for Case 2 with $t_B^* > t_A^*$, $N_A = 1$, $N_B = 1$, $s = 1$, $\alpha = 2$, $\beta = 1$, $\gamma = 2$, $t_A^* = 0$, and $t_B^* = 2/3$



not reach the system-optimal minimum value at the point where t^2_{As} coincides with t^o_{As}. Thus, the price of anarchy is always positive. This is because both users depart during the interval (t^*_A, t^2_{Ae}), with user A arriving late and user B arriving early. To see that simultaneous departure is inefficient, consider a vehicle of user B arriving at t_1 and a vehicle of user A arriving at $t_2 > t_1$. If the two vehicles exchanged slots, user A's vehicle would arrive less late and user B's vehicle would arrive less early. The two users might agree to such an exchange if they cooperated, but the switch cannot occur in a Nash equilibrium.

We conclude this section with a numerical example featuring the same parameter values as those used to construct Figures 4 and 5: $N_A = 1$, $N_B = 1$, $s = 1$, $\alpha = 2$, $\beta = 1$, $\gamma = 2$, $t^*_A = 0$, and $t^*_B = 2/3$. Departures and arrivals last for two hours. The system optimum begins at $t^o_{As} = -5/6$, and total costs are $TC^o = 3/4$. The PSNE for Case 1 begins at $t^1_{As} = -2/3$, and total costs are $TC^1 = 5/6$ which is 11.1% higher than TC^o. User A's total costs are 1/3, and user B's total costs are 1/2. User B's total costs are 50% higher than user A's total costs even though the two users have the same size of fleet and the same unit costs α, β, and γ. The PSNE for Case 2 begins at $t^2_{As} = -4/5$, and total costs are $TC^2 = 0.81$ which is 8.4% higher than TC^o but 2.4% lower than TC^1. User A's total costs are 2/5, and user B's total costs are 0.413. User B's total costs are just 3.3% higher than user A's total costs so that the PSNE in Case 2 is both more efficient and more equitable than in Case 1. Finally, note that if user B's desired arrival time is increased from $t^*_B = 2/3$ to $t^*_B = 4/5$, t^o_{As} increases to $-23/30$ while t^2_{As} remains at $-4/5$. The PSNE in Case 2 then begins later than the system optimum.

5. Conclusion

In this paper we have explored the existence and uniqueness of PSNE in the bottleneck model with atomic users. We consider a simple case featuring two users with piecewise linear and independent trip-timing preferences for each vehicle in their fleets. We show that if users are identical, and $\gamma > \alpha$, a PSNE does not exist. By contrast, if $\gamma \leq \alpha$, multiple PSNE do exist in which no queuing occurs. The aggregate departure profile is unique. However, there exists a continuum of departure-time schedules for the two users that differ in the share of total costs borne by each user. We also consider a setting in which the two users differ in their preferred arrival times, t^*_A and t^*_B. For a range of values of $t^*_B - t^*_A$, there exists a continuum of PSNE that differ in the timing of trips, total costs, and the relative burden of costs borne by each user. Table 1 summarizes the main results of the paper.

The potential nonexistence of PSNE with atomic users is, we believe, the most significant of the results. In the standard bottleneck model a PSNE exists under relatively general assumptions. As Lindsey (2004) shows, an equilibrium fails to exist only if users' trip cost functions have discontinuities of a particular sort. It can fail if the schedule delay cost function is not upper semicontinuous (e.g., if there is a discrete penalty for arriving after a particular time). It can also fail to exist if the cost of travel time takes an upward jump above a particular trip duration. In the model with nonatomic users nonexistence is a problem even with smooth preferences because atomic users face conflicting incentives. On one hand they prefer to spread out departures of their fleets to avoid self-imposed queuing delays. If there is only one user, this results in a system-optimal departure schedule with no queuing. Yet when other users are present, an atomic user has an incentive to force vehicles into the departure stream near the peak to reduce its fleet's schedule delay costs. A user's best response can be to preempt other users by scheduling a mass departure near t^* and hogging bottleneck capacity. Yet mass departures are inconsistent with PSNE, as we have shown. In a sense, nonexistence of PSNE is likely to be more prevalent with atomic users than nonatomic users because atomic users have more degrees of freedom. A nonatomic user decides only when to schedule a single vehicle departure. An atomic user has an uncountable infinite number of departure-time decisions to make. Consequently, it is more difficult to find a departure profile for each fleet such that no user can beneficially deviate by rescheduling its vehicles in any way.

The out-of-equilibrium behavior in the bottleneck model bears some resemblance to firm behavior in

Table 1. Summary of Results

Main case	Subcase	Existence	Inefficiency	Comments
Homogeneous users	$\gamma > \alpha$	No PSNE	—	—
	$\gamma \leq \alpha$	Multiple PSNE	0%	One user can incur up to 75% of the total costs.
Heterogeneous users in t^* and fleet size	$t^*_B > t^*_A$	Multiple PSNE	$\leq 20\%$	There is no queuing. Departures may start earlier or later than optimal.

Notes. The inefficiency is measured as the price of anarchy (see Equation (42)). The results for heterogeneous users are valid only for the range of $t^*_B - t^*_A$ described by Equations (27), (39), and (40).

one-dimensional (Hotelling) spatial competition models (see Anderson, de Palma, and Thisse 1992, Chapter 8). In location-choice games with parametric prices, firms leapfrog each other to gain market share. The outcome depends on the number of firms. With two or four firms a unique PSNE exists. Yet with five firms there are multiple PSNE, and with three firms a PSNE does not exist. In simultaneous location- and price-choice games with two firms, a PSNE does not exist. Firms face conflicting location-choice incentives that are broadly similar to the incentives of atomic users picking departure times in the bottleneck model. Moving closer to the center (analogous to departing close to t^*) gains firms market share, but moving away softens price competition. The Hotelling model (Hotelling 1929) is similar to the bottleneck model insofar as both settings feature atomic players. However, the Hotelling model is simpler because each player controls only one outlet or plant, whereas in the bottleneck model each player controls a continuum of vehicles.

Preliminary analysis suggests that the difference between the standard bottleneck model and the model with atomic users remains as the number of atomic users increases because each user still recognizes how its scheduling choices affect queuing delays. Thus, simple congestion-prone systems with atomic users can exhibit fundamentally different dynamic behavior than with nonatomic users. Dynamic systems with nonatomic users have been well studied in the literature using the bottleneck model as well as various flow-congestion models. Yet, to our knowledge, this paper is the first to examine the existence and uniqueness of equilibrium in dynamic systems with atomic users that make scheduling decisions simultaneously.

Our analysis is exploratory and can be extended in various directions. One obvious priority is to determine how the bottleneck model can be modified to restore existence of equilibrium without restricting parameter values. One possibility is to consider both atomic and nonatomic users scheduling vehicles simultaneously. Although this setting may be more realistic in some markets, our preliminary analysis indicates that the presence of nonatomic agents does not restore existence of PSNE. Indeed, the conditions for existence of a PSNE are more stringent in the setting considered here even if atomic users control only a small share of traffic.

Another option to restore equilibrium is to consider mixed-strategy Nash equilibrium. However, since the strategy space in our setting allows discontinuous functions that may take a positive value only at a finite number of points, there is no guarantee that a Nash equilibrium in mixed strategies exists. It does not appear that the strategy space satisfies the sufficient conditions for existence established in theorems that consider infinite strategy spaces (e.g., a compact Hausdorff space as described in Reny 1999). A discrete version of the bottleneck model with finite time steps and discrete vehicles is guaranteed to have a Nash equilibrium in mixed strategies because the strategy space is finite. Such an extension is beyond the scope of this paper. Moreover, studies that have used the discrete bottleneck model have found that a Nash equilibrium in pure strategies may not exist even if each player controls only one vehicle. Deriving pure or mixed strategy equilibria is also computationally-demanding for more than a few players. Adopting the discrete version of the bottleneck model therefore does not appear to be a promising approach.

A third option is to consider equilibrium under centralized pricing. In our setting, charging the first-best tolls for the model with nonatomic users restores equilibrium, but it is an open question whether a PSNE exists under more realistic step-tolling. Further possibilities for restoring equilibrium include: sequential decision making, consideration of repeated games, heterogeneity in unit values of time and schedule delay, and a more general specification of scheduling preferences.

Another direction for future research is to investigate more fully the benefits of coordination or other forms of cooperation. Self-internalization of congestion externalities by an atomic agent is one form of coordination, and its potential for welfare gains is apparent from Section 4. Cooperation between atomic agents can also be beneficial as shown by the examples in which the price of anarchy is positive. Major freight shippers might cooperate by harmonizing their delivery schedules although doing so without triggering concerns about illegal collusion might be tricky. Governments can also implement or encourage cooperative behavior. One example is staggered work hours and flextime programs that entail coordination on target arrival times (i.e., t^*). Staggered work hours were introduced in the 1970s in Washington, D.C., and Ottawa, Ontario, Canada. Gutiérrez-i-Puigarnau and Van Ommeren (2012) report that one in three German firms uses staggered working hours. Mun and Yonekawa (2006) note that in Japan, in 1998, 8% of employees worked at a firm with flexible working hours, and in the United States, in 1994–1997, less than 6% of employees had a formal flexible working arrangement, but 28% of all full-time workers varied their working times to some degree.

Acknowledgments

The authors are grateful to the associate editor and two anonymous referees for very helpful suggestions. The authors also thank Erik Verhoef for his constructive and incisive comments that have significantly improved the paper. The authors would also like to thank participants at the Kuhmo Nectar Conference on Transportation Economics 2013 at Northwestern University. Any errors are the authors.

References

Anderson SP, de Palma A, Thisse JF (1992) *Discrete Choice Theory of Product Differentiation* (MIT Press, Cambridge, MA).

Arnott R, de Palma A, Lindsey R (1987) Bottleneck congestion with elastic demand. Discussion Paper 690, Institute for Economic Research, Queen's University, Kingston, Ontario, Canada.

Arnott R, de Palma A, Lindsey R (1990) Economics of a bottleneck. *J. Urban Econom.* 27(1):111–130.

Arnott R, de Palma A, Lindsey R (1993) A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *Amer. Econom. Rev.* 83(1):161–179.

Arnott R, de Palma A, Lindsey R (1998) Recent developments in the bottleneck model. Button KJ, Verhoef ET, eds. *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility* (Edward Elgar, Aldershot, UK), 79–110.

Asensio J, Matas A (2008) Commuters valuation of travel time variability. *Transportation Res. Part E: Logist. Transportation Rev.* 44(6):1074–1085.

Basso LJ, Zhang A (2007) Congestible facility rivalry in vertical structures. *J. Urban Econom.* 61(2):218–237.

Brueckner JK (2002) Airport congestion when carriers have market power. *Amer. Econom. Rev.* 92(5):1357–1375.

Brueckner JK (2005) Internalization of airport congestion: A network analysis. *Internat. J. Indust. Organ.* 23(7–8):599–614.

Brueckner JK, Van Dender K (2008) Atomistic congestion tolls at concentrated airports? Seeking a unified view in the internalization debate. *J. Urban Econom.* 64(2):288–295.

Catoni S, Pallottino S (1991) Technical note—Traffic equilibrium paradoxes. *Transportation Sci.* 25(3):240–244.

Cominetti R, Correa JR, Stier-Moses NE (2009) The impact of oligopolistic competition in networks. *Oper. Res.* 57(6):1421–1437.

Daniel JI (1995) Congestion pricing and capacity of large hub airports: A bottleneck model with stochastic queues. *Econometrica* 63(2):327–370.

Daniel JI (2009) The deterministic bottleneck model with non-atomic traffic. Working Paper 09–08, Department of Economics, University of Delaware, Newark.

Daniel JI, Harback KT (2008) (When) do hub airlines internalize their self-imposed congestion delays? *J. Urban Econom.* 63(2):583–612.

de Palma A, Fontan C (2001) Eléments d'analyse de la composante horaire des déplacements: le cas de la région ile-de-france. *Les Cahiers Scientifiques du Transport* 39:55–86.

de Palma A, Fosgerau M (2011) Dynamic traffic modeling. de Palma A, Lindsey R, Quinet E, Vickerman R, eds. *Handbook in Transport Economics* (Edward Elgar, Cheltenham, UK), 188–212.

de Palma A, Ben-Akiva M, Lefevre C, Litinas N (1983) Stochastic equilibrium model of peak period traffic congestion. *Transportation Sci.* 17(4):430–453.

Devarajan S (1981) A note of network equilibrium and noncooperative games. *Transportation Res. Part B: Methodological* 15(6): 421–426.

Economist, The (2013) Air cargo: Cabin fever. (November 2), http://www.economist.com/news/business/21588923-fedex-and-ups-have-turned-memphis-and-louisville-aerotropolises-cabin-fever.

Fotakis D, Kontogiannis S, Spirakis P (2008) Atomic congestion games among coalitions. *ACM Trans. Algorithms* (*TALG*) 4(4): 1–27.

Gutiérrez-i-Puigarnau E, Van Ommeren JN (2012) Start time and worker compensation implications for staggered-hours programmes. *J. Transport Econom. Policy* 46(2):205–220.

Harker PT (1988) Multiple equilibrium behaviors on networks. *Transportation Sci.* 22(1):39–46.

Haurie A, Marcotte P (1985) On the relationship between Nash–Cournot and Wardrop equilibria. *Networks* 15(3):295–308.

He F, Yin Y, Shirmohammadi N, Nie YM (2013) Tradable credit schemes on networks with mixed equilibrium behaviors. *Transportation Res. Part B: Methodological* 57:47–65.

Hoefer M, Skopalik A (2009) Altruism in atomic congestion games. Fiat A, Sanders P, eds. *Algorithms—ESA 2009*, Lecture Notes Comput. Sci., Vol. 5757 (Springer-Verlag, Berlin Heidelberg), 179–189.

Hotelling H (1929) Stability in competition. *Economic J.* 39(153):41–57.

Levinson D (2005) Micro-foundations of congestion and pricing: A game theory perspective. *Transportation Res. Part A: Policy Practice* 39(7):691–704.

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

Lindsey R, van den Berg VAC, Verhoef ET (2012) Step tolling with bottleneck queuing congestion. *J. Urban Econom.* 72(1):46–59.

Marcotte P (1987) Algorithms for the network oligopoly problem. *J. Oper. Res. Soc.* 38(11):1051–1065.

Miller TC, Tobin RL, Friesz TL (1991) Stackelberg games on a network with Cournot–Nash oligopolistic competitors. *J. Regional Sci.* 31(4):435–454.

Molnar A (2013) Congesting the commons: A test for strategic congestion externalities in the airline industry. Working paper, Stanford University, Stanford, CA.

Mun S, Yonekawa M (2006) Flextime, traffic congestion and urban productivity. *J. Transport Econom. Policy* (*JTEP*) 40(3):329–358.

Newell GF (1987) The morning commute for nonidentical travelers. *Transportation Sci.* 21(2):74–88.

Otsubo H, Rapoport A (2008) Vickrey's model of traffic congestion discretized. *Transportation Res. Part B: Methodological* 42(10): 873–889.

Pels E, Verhoef ET (2004) The economics of airport congestion pricing. *J. Urban Econom.* 55(2):257–277.

Reny PJ (1999) On the existence of pure and mixed strategy Nash equilibria in discontinuous games. *Econometrica* 67(5): 1029–1056.

Silva HE, Verhoef ET (2013) Optimal pricing of flights and passengers at congested airports and the efficiency of atomistic charges. *J. Public Econom.* 106:1–13.

Silva HE, Verhoef ET, van den Berg VAC (2014) Airlines' strategic interactions and airport pricing in a dynamic bottleneck model of congestion. *J. Urban Econom.* 80:13–27.

Small KA (1982) The scheduling of consumer activities: Work trips. *Amer. Econom. Rev.* 72(3):467–479.

Vickrey WS (1969) Congestion theory and transport investment. *Amer. Econom. Rev.* 59(2):251–260.

Wardman M (2001) A review of British evidence on time and service quality valuations. *Transportation Res. Part E: Logist. Transportation Rev.* 37(2):107–128.

Werth TL, Holzhauser M, Krumke SO (2014) Atomic routing in a deterministic queuing model. *Oper. Res. Perspectives* 1(1):18–41.

Yang H, Zhang X (2008) Existence of anonymous link tolls for system optimum on networks with mixed equilibrium behaviors. *Transportation Res. Part B: Methodological* 42(2):99–112.

Yang H, Zhang X, Meng Q (2007) Stackelberg games and multiple equilibrium behaviors on networks. *Transportation Res. Part B: Methodological* 41(8):841–861.

Zhang A, Zhang Y (2006) Airport capacity and congestion when carriers have market power. *J. Urban Econom.* 60(2):229–247.

Zou X, Levinson D (2006) A multi-agent congestion and pricing model. *Transportmetrica* 2(3):237–249.