Knockaert, Jasper; Verhoef, Erik T.; Rouwendal, Jan

Working Paper

Bottleneck Congestion: Differentiating the Coarse Charge

Tinbergen Institute Discussion Paper, No. 10-097/3

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Knockaert, Jasper; Verhoef, Erik T.; Rouwendal, Jan (2010) : Bottleneck Congestion: Differentiating the Coarse Charge, Tinbergen Institute Discussion Paper, No. 10-097/3, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:
https://hdl.handle.net/10419/86997

TI 2010-097/3
Tinbergen Institute Discussion Paper

Bottleneck Congestion: Differentiating the Coarse Charge

Jasper Knockaert
Erik T. Verhoef
Jan Rouwendal

VU University Amsterdam, and Tinbergen Institute.

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Bottleneck Congestion: Differentiating the Coarse Charge

Jasper Knockaert, Erik Verhoef, Jan Rouwendal
mailto:jknockaert@feweb.vu.nl

6th September 2010

Abstract

The traditional bottleneck model for road congestion promotes the implementation of a triangular, fully time varying, charge as the optimal solution for the road congestion externality. However, cognitive and technological barriers put a practical limit to the degree of differentiation real world implementations can handle. The traditional approach to accommodate for this concern has been a step toll, with the single step *coarse* charge as its simplest case.

In this paper we study how efficiency of the coarse charge can be improved by differentiating its level and timing across groups of travellers. We use the traditional bottleneck model to analyse how the coarse charge can be differentiated over two groups of travellers assuming inelastic peak-hour demand.

The results of our analysis indicate that differentiating the coarse charge across two groups of travellers considerably improves its efficiency without increasing cognitive effort and decision making costs for the individual traveller. A numeric illustration reveals a welfare gain of 69% of the first best charge, up from 53% for the generic coarse charge. This increase is similar to what is obtained by moving from the coarse charge to a generic two step toll. Once different groups have been defined, one could in fact achieve the same gains by temporal separation of drivers, for example by use of licence plate numbers.

The presented charging regime has a considerable degree of flexibility with respect to share of travellers to attribute to each scheme, which further adds to its merits in practical applicability.

1 Introduction

The bottleneck model first introduced by Vickrey (1969) has been recognised as the reference representation of peak hour road congestion. A structural definition of the model is provided by Arnott, Palma and Lindsey (1993) who illustrate its application for the assessment of a number of charging schemes. While the first best charge directly derives from the model definition, analysis of alternative schemes becomes quickly complicated as noted by Arnott, Palma and Lindsey (1990).

Subsequent literature on bottleneck charging focused on a number of extensions, including heterogeneous travellers, interaction between parallel or serial bottlenecks, the presence of untaxed alternatives and so on. An overview is provided by Arnott, Palma and Lindsey (1998); Lindsey and Verhoef (2001).

While the existing analyses of bottleneck charging are generally illuminating, barriers have been identified in implementing it straightforwardly. As noted by Arnott

et al. (1990), the implementation of a first best time varying charge is technically demanding. The fully variable characteristic of the charge probably also requires a considerable cognitive effort by the traveller. It is to be expected that effectiveness of charging reaches a limit or even decreases when search and decision costs become prohibitive as a result of too much differentiation (see for instance Norwood, 2006).

In order to meet concerns about feasibility, Arnott et al. (1990) present an optimal coarse charge which yields about half of the efficiency of the first best scheme (Arnott et al., 1993).

In this paper we present an approach that allows for an improved efficiency of bottleneck congestion charging by allowing for differentiation across groups of travellers rather than along behavioural dimensions. We show how social efficiency of the charging regime improves considerably without increasing cognitive efforts for the individual traveller.

In a first section we introduce the bottleneck model. Subsequently we discuss the first best charging scheme as well as the optimal coarse charging regime. In a next section we introduce a regime that differentiates the coarse scheme over two groups of travellers and study the merits of such a scheme. In a final section we conclude.

2 The Bottleneck model

In this section we provide a summary introduction to Vickrey's bottleneck model and discuss the no-toll equilibrium. We mainly draw from Arnott et al. (1990, 1998), we refer to the original source for the full story.

The bottleneck model is a stylised representation of traffic congestion. It assumes a group of N identical car drivers who want to arrive at their destination at time t^*. The travellers follow a single road which has a bottleneck with a fixed flow capacity s and which is otherwise uncongested. Without loss of generality for a single bottleneck, it is assumed that the drivers arrive at the bottleneck immediately after departure from their origin, and arrive at their destination immediately after leaving the bottleneck. Hence travel time is limited to waiting time (queueing time) at the bottleneck.

The limited bottleneck capacity makes it impossible for all N travellers to arrive at destination at the same desired time t^*. Those who arrive early or late face a schedule delay cost.

Generalised travel costs C of an individual are determined by queueing costs and schedule delay:[1]

$$C = \alpha(\text{travel time}) + \beta(\text{time early}) + \gamma(\text{time late}) \tag{1}$$

with α, β and γ the shadow cost of time spent waiting in the queue, schedule delay early and schedule delay late. We assume in this paper that $0 < \beta < \alpha < \gamma$ (Small, 1982). The inequality $\beta < \alpha$ is required to avoid a mass departure in the equilibrium; $\alpha < \gamma$ is not required for any such reason.

The (generalised) trip price $p(t)$ for an individual arriving at time t equals travel cost $C(t)$ plus any charge $\rho(t)$:

$$p(t) = C(t) + \rho(t) \tag{2}$$

[1] The generalised travel cost is the sum of all monetary and non-monetary costs. In the context of the bottleneck model used here it is limited to the relevant components which are waiting time costs and schedule delay costs. Other costs components like fuel costs are constant over the peak hour and not relevant for our discussion, so we leave them out for simplicity.

For an equilibrium between travellers to be achieved it is required that the generalised trip cost $p(t)$ is uniform over the peak period and higher outside this period. As schedule delay costs cannot be equal for all travellers, a variation in waiting time must compensate to allow for a user equilibrium in the absence of a charge ($\rho(t) = 0$ implies $p(t) = C(t)$). In order to satisfy this condition, a queue builds up behind a bottleneck matching the evolution in schedule delay costs (see figure 1).

In the equilibrium the first and the last traveller arriving at the bottleneck face no queue but incur the highest schedule delay costs. After the first traveller arrives, a queue starts to build up corresponding to a constant departure rate $s\alpha/(\alpha-\beta)$ that exceeds the bottleneck capacity s.

The traveller arriving at destination exactly on time t^* faces the longest waiting time but no scheduling costs. After the departure of this traveller, the departure rate drops to a level $s\alpha/(\alpha+\gamma)$ which is smaller than the bottleneck capacity s. As a result the queue length and waiting costs diminish until the last traveller departs and arrives.

Social congestion costs in the bottleneck model are equal to total travel costs TC and can be expressed as:

$$TC = NC_N = \delta\frac{N^2}{s} \tag{3}$$

with $\delta = \beta\gamma/(\beta+\gamma)$.[2]

Figure 1 indicates that half of the total travel costs TC are made up of waiting costs whereas the other half are schedule delay costs.

Social costs of peak hour travel in the bottleneck model are independent of travel time value α. The value of α only affects the length of the queue but not the actual waiting costs in the equilibrium, because delay α evolves such that it compensates for

[2]This can be verified by noting that the first (last) driver incurs a schedule delay early (late) of $(N/s)\gamma/(\beta+\gamma)$ ($(N/s)\beta/(\beta+\gamma)$). Multiplying by β (γ) gives the cost $(N/s)\delta$. The cost is equal for all drivers in equilibrium, hence the expression for TC.

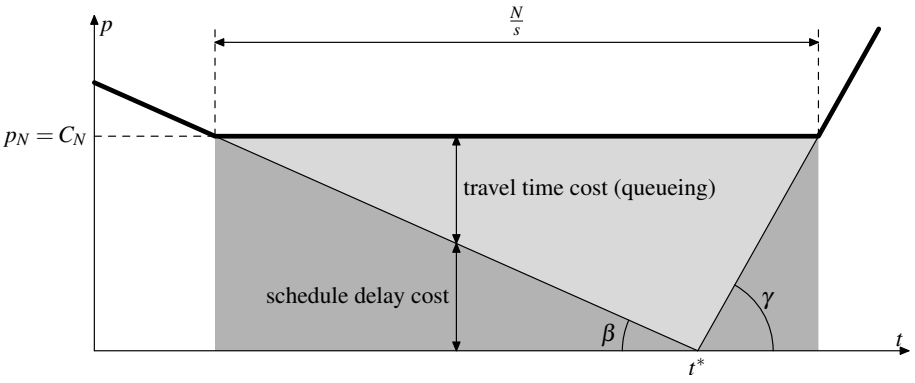

Figure 1: Generalised user cost as a function of arrival time t in the *uncharged* bottleneck model (the bold line indicates generalised user price $p(t)$ and the shaded area represents social congestion costs TC)

changes in schedule delay cost over time.

3 Different charging regimes

The literature describes the optimisation of a number of charging schemes of which we will present the first-best and the coarse charge for the case of inelastic trip demand N. The discussion here mainly draws from Arnott et al. (1990, 1998).

The *first best* charge corresponds to the situation where a time varying charge $\rho(t)$ replaces queueing time costs (see figure 2). In the equilibrium the individual traveller is faced with the same generalised price p_N as in the uncharged case; as a result the user equilibrium does not change. But through adjusting the departure rate at origin to match the bottleneck capacity s, the queue before the bottleneck disappears, which corresponds to a net welfare gain. Note that the arrival rate is unchanged compared to the uncharged regime, and remains equal to s.

The net welfare gain of the charge $\rho(t)$ corresponds to the reduction in total social costs TC compared to the uncharged scenario.[3] As can be easy derived by comparing figure 2 to 1, the welfare gain amounts to half the social cost of the peak travel. The other half are scheduling costs, which cannot be abated.

The first-best charging scheme requires a continuously varying charge, which poses some difficulties in application under real world circumstances. To address this concern, a second-best *coarse* charge has been presented by Arnott et al. (1990). The coarse charge is a fixed charge, levied over a limited time period. Arriving outside this period (before or after) is left uncharged.

The optimal coarse charge is turned on before t^* and turned off after t^*. The level ρ and timing of the charge are chosen such that (in the user equilibrium) the length of the queue is zero at the moment the charged period starts and just before the charged period ends, but at no other time during the peak hour (see figure 3).

Under the coarse charge three groups of travellers can be distinguished. A first group of N_E travellers passes through the bottleneck before the charge is turned on.

[3] Monetary transactions related to the charge are considered not to be a net welfare cost. This can be realised using a lump-sum transfer.

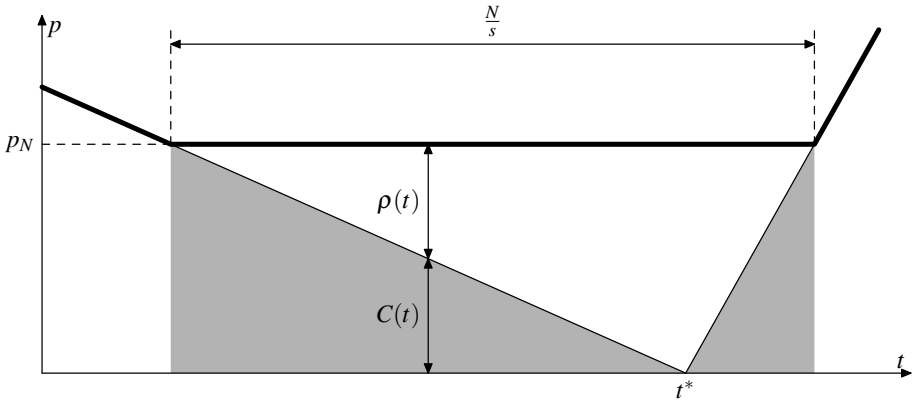

Figure 2: Generalised user cost and charge as a function of arrival time t under the *first-best* bottleneck charge (the bold line indicates generalised user price $p(t)$ and the shaded area represents social congestion costs TC)

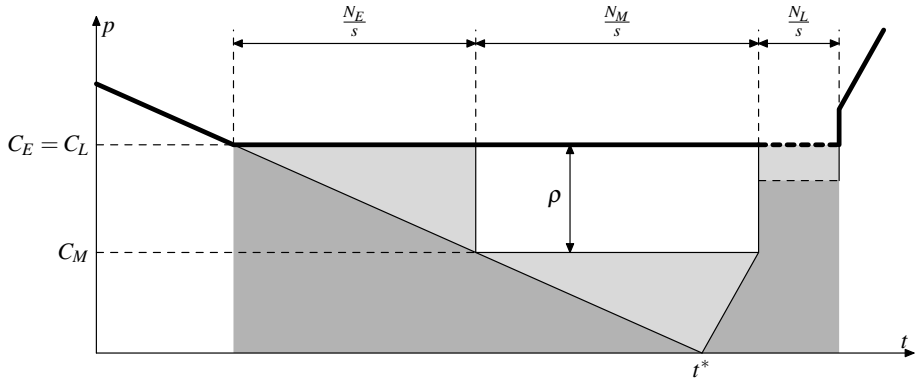

Figure 3: Generalised user cost and charge as a function of arrival time t under the *coarse* bottleneck charge (the bold line indicates generalised user price $p(t)$ and the shaded area represents social congestion costs TC)

After the first traveller enters the bottleneck a queue builds up corresponding to a constant departure rate $s\alpha/(\alpha-\beta)$ exceeding the bottleneck capacity s. After the last traveller of N_E departs, there are no departures until the charge is turned on. In the equilibrium the travellers of group N_E face a uniform generalised travel price p_E:

$$p_E = C_E = C_M + \beta\frac{N_E}{s} \tag{4}$$

After the queue length is reduced back again to zero, the charge ρ is turned on. A second group of travellers N_M passes through the bottleneck during the charged period. After the first traveller of N_M enters the bottleneck, a queue starts to build up as a result of a constant departure rate $s\alpha/(\alpha-\beta)$ exceeding the bottleneck capacity s. The queue reaches maximum length at the departure time of the traveller arriving at destination at t^*. After this traveller departs the constant departure rate drops to a level $s\alpha/(\alpha+\gamma)$. The queue shortens correspondingly to reach zero length again when the last traveller of N_M travels through the bottleneck. Travellers of group N_M incur uniform generalised travel costs that correspond to uncharged bottleneck congestion with uniform travel time and schedule delay cost C_M plus a uniform charge ρ:

$$p_M = C_M + \rho = \delta\frac{N_M}{s} + \rho \tag{5}$$

Immediately after the charge is lifted, all travellers of the last group N_L depart concurrently, and pass through the bottleneck in random order. This mass departure is needed in equilibrium to make the expected cost for a traveller in this group equal to that of the last tolled traveller, who faces a positive toll and a zero travel delay. In figure 3 we represent the *expected* travel costs C_L, equal to the expected price p_L, for each of these travellers with dashed lines:

$$p_L = C_L = C_M + \frac{\alpha+\gamma}{2}\frac{N_L}{s} \tag{6}$$

Note that the first traveller in group L faces a lower realisation of cost than p_L, and the last one a higher realisation. It is the randomness in passage times that equates the expected costs across individuals.

The user equilibrium across the different groups requires that $p_E = p_L = p_M$. From this condition and the expressions for generalised travel prices p_E (4) and p_L (6) follows the size of the groups N_E, N_L and ultimately N_M as a function of ρ.

Optimising total social costs TC delivers us the optimal coarse charge level ρ:[4]

$$TC = N_E C_E + N_M C_M + N_L C_L \tag{7}$$

$$\frac{dTC}{d\rho} = 0 \Rightarrow \rho = \frac{\delta}{2}\frac{N}{s} \tag{8}$$

The timing of the charge follows from the definition of the scheme and the group sizes.

Note that for the second best coarse charge it is sufficient that the charging scheme applies to drivers traveling outside the charging period. It is a bit paradoxical that the N_M drivers travelling during the charging period can be exempt from the charging scheme without affecting the equilibrium.

Ultimately, the coarse charge user equilibrium could even be reproduced without any charging scheme by barring a dedicated group of $N_E + N_L$ drivers from travelling during the charging period and allowing the remainder of the drivers (N_M) to travel in that period without paying any charge.[5]

One could imagine an alternating scheme where drivers are attributed to one of both groups on a day-by-day basis to implement such a scheme. Licence plate numbers could be used for assigning individuals to groups, and would in fact allow for refinement over more classes than just two. When the time intervals thus assigned to the different groups alternate over working days, all users would in the longer run benefit from such policies, at least when demand is perfectly inelastic, so that a strict Pareto improvement is achieved.

By defining separate groups, tolls are no longer needed to keep drivers out of the central time intervals, while the efficiency gains are identical to those achieved by step tolls designed for the same number of intervals. This fact, combined with the presumably higher acceptability of the policy because of the absence of pricing, may make it an attractive possibility for congestion management in practice. Shen and Zhang (in press) make similar observations for temporal separation of travellers through the use of ramp-metering for highways with multiple on-ramps.

4 The differentiated coarse charge

In this section we propose a scheme for a coarse charge which is differentiated over two groups of travellers. We elaborate on the case where $0 < \beta < \alpha < \gamma$ as in the regimes discussed in the previous sections. We optimise our scheme with respect to level and timing of the charges and share of the travellers attributed to each scheme.

We will initiate the discussion under a setting where travellers are exogenously attributed to one of the two groups. Subsequently we will study what happens if the group choice is endogenous.

The scheme we present is inspired by the optimal coarse charge discussed in the previous section. A first group of travellers N_1 is confronted with a coarse charge ρ_1

[4] A full derivation is given in Arnott et al. (1990).

[5] It turns out that a comparable result was derived simultaneously, independent of ours, by Fosgerau (2010), in the context of a bottleneck with parallel queues and a different formulation of schedule delay costs.

and a second group with a coarse charge ρ_2. For the discussion here we arbitrarily choose $\rho_2 > \rho_1$.

The *first group* of travellers N_1 faces the generalised price evolution illustrated in figure 4. Before the charge ρ_1 is levied a group of travellers N_{1E} pass through the bottleneck from t^s_{1E} onwards. The first traveller of N_{1E} faces no waiting costs but only schedule delay. After the first traveller travels through the bottleneck, a queue builds up corresponding to a constant departure rate $s\alpha/(\alpha-\beta)$ exceeding the bottleneck capacity s. After the last traveller of N_{1E} departs, the departure rate drops to zero until this traveller arrives at destination at t^e_{1E}, and the queue length is reduced back to zero just before the charge ρ_1 is turned on. The first traveller paying the toll will have a zero travel delay and a cost C_M. The first traveller at t^s_{1E} has also a zero travel delay and an additional schedule delay cost $\beta N_{1E}/s$. The generalised travel price p_{1E} is then equal to the uniform sum of waiting and scheduling costs C_{1E}:

$$p_{1E} = C_{1E} = C_{1M} + \beta \frac{N_{1E}}{s} \tag{9}$$

After the charge ρ_1 is turned on, another group of travellers N_{1M} passes through the bottleneck (together with travellers of group N_{2M}, see below), with the first and the last traveller facing no queue and the traveller arriving at t^* facing the highest queueing cost but no schedule delay penalty. Travellers of group N_{1M} incur uniform generalised travel costs $C_{1M}(=C_{2M})$ that correspond to uncharged bottleneck congestion for a peak with $N_{1M}+N_{2M}$ travellers, plus a uniform charge ρ_1:

$$p_{1M} = C_{1M} + \rho_1 = \delta \frac{N_{1M}+N_{2M}}{s} + \rho_1 \tag{10}$$

Immediately after the charge ρ_1 is turned off, there is a mass departure of N_{1L} travellers who travel through the bottleneck in random order. The expected (or average) generalised price p_{1L} for this group is:

$$p_{1L} = C_{1L} = C_{1M} + \frac{\alpha+\gamma}{2} \frac{N_{1L}}{s} \tag{11}$$

The dynamic user equilibrium between the subgroups of N_1 requires a uniform generalised trip price: $p_{1E} = p_{1M} = p_{1L}$. From this condition and the expressions for

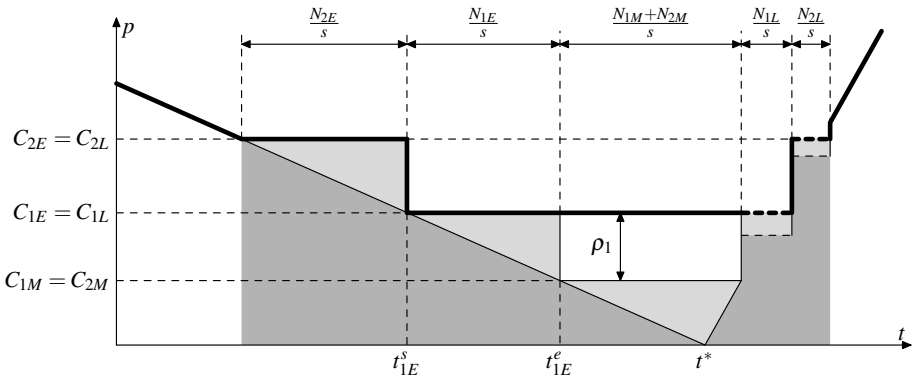

Figure 4: Generalised user cost and charge ρ_1 as a function of arrival time t under the *differentiated coarse* bottleneck charge (the bold line indicates generalised user price $p(t)$ of group N_1 and the shaded area represents total social congestion costs TC)

generalised trip cost p_{1E} (9) and p_{1L} (11) follows the size of the groups as a function of ρ_1:

$$p_{1E} = p_{1M} \Rightarrow N_{1E} = \frac{\rho_1 s}{\beta} \tag{12}$$

$$p_{1L} = p_{1M} \Rightarrow N_{1L} = \frac{2\rho_1 s}{\alpha + \gamma} \tag{13}$$

The *second group* of travellers N_2 face a charge ρ_2 that is levied during the entire period over which travellers of the first group travel. The generalised travel price for this group is presented in figure 5. As the charge ρ_2 is larger than ρ_1, no travellers of the second group travel together with travellers of N_{1E} and N_{1L}.

A first batch of travellers N_{2E} passes through the bottleneck before N_{1E}. The first traveller of N_{2E} faces no waiting costs but only schedule delay. After the first traveller travels through the bottleneck, a queue builds up corresponding to a constant departure rate $s\alpha/(\alpha-\beta)$ exceeding the bottleneck capacity s. After the last traveller of N_{2E} departs, the departure rate drops to zero until this traveller arrives at destination and queue length is reduced back to zero just before the charge ρ_2 is turned on (after which travellers of N_{1E} start using the bottleneck). The generalised travel price p_{2E} is equal to the uniform sum of waiting and scheduling costs C_{2E}:

$$p_{2E} = C_{2E} = C_{2M} + \beta \frac{N_{1E} + N_{2E}}{s} \tag{14}$$

In the middle of the peak a batch N_{2M} passes through the bottleneck over the same time period as (and together with) N_{1M}. They face generalised travel costs $C_{2M} = C_{1M}$ plus the charge ρ_2:

$$p_{2M} = C_{2M} + \rho_2 = \delta \frac{N_{1M} + N_{2M}}{s} + \rho_2 \tag{15}$$

After the last traveller of N_{1L} arrives, the charge ρ_2 is turned off and immediately a mass of N_{2L} travellers depart. This group travels through the bottleneck in random order. The expected generalised price p_{2L} for this group is:

$$p_{2L} = C_{2L} = C_{2M} + \gamma \frac{N_{1L}}{s} + \frac{\alpha + \gamma}{2} \frac{N_{2L}}{s} \tag{16}$$

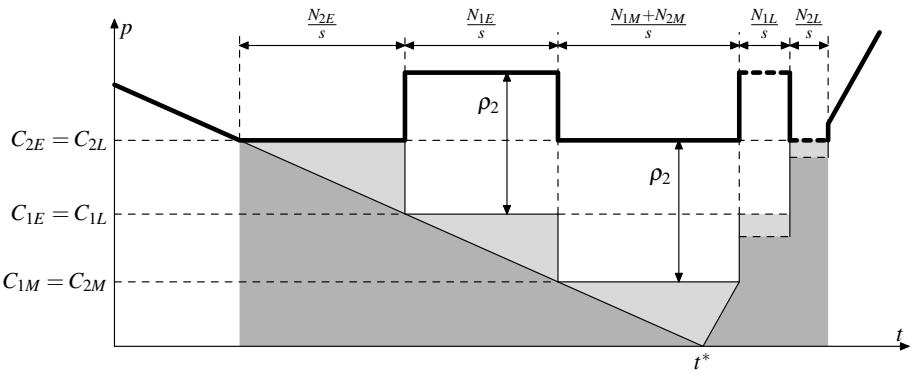

Figure 5: Generalised user cost and charge ρ_2 as a function of arrival time t under the *differentiated coarse* bottleneck charge (the bold line indicates generalised user price $p(t)$ of group N_2 and the shaded area represents total social congestion costs TC)

The dynamic user equilibrium between the subgroups of N_2 requires that $p_{2E} = p_{2M} = p_{2L}$. From this condition and the expressions for generalised travel prices p_{2E} (14) and p_{2L} (16) follows the size of the groups N_{2E} and N_{2L} as a function of ρ_2, N_{1E} and N_{1L}:

$$p_{2E} = p_{2M} \Rightarrow N_{2E} = \frac{\rho_2 s}{\beta} - N_{1E} \tag{17}$$

$$p_{2L} = p_{2M} \Rightarrow N_{2L} = 2\frac{\rho_2 s - \gamma N_{1L}}{\alpha + \gamma} \tag{18}$$

Some obvious conditions apply to the size of the different groups. Provided that banking of travellers is not a real world option, group sizes cannot turn negative. When ρ_2 is smaller than $2\rho_1\gamma/(\alpha+\gamma)$ the subgroup N_{2L} has size zero and hence does not exist in the equilibrium.

The size of $N_{1M} + N_{2M}$ follows from the size of the other groups. There is however a degree of freedom left here, which means that we can arbitrarily choose N_{1M} or N_{2M}. Even could we exempt a group of users of the size $N_{1M} + N_{2M}$ from any charging scheme without affecting the equilibrium.

In order to determine the optimal level of the charges ρ_1 and ρ_2, we optimise social congestion costs. As discussed in the previous section, social congestion costs in the bottleneck model are equal to the sum of total waiting costs and schedule delay. In our analysis we will assume that $\rho_2 > 2\rho_1\gamma/(\alpha+\gamma)$ (or that group N_{2L} has a strictly positive size). We will evaluate this assumption at the end of this section.

Total waiting and schedule delay costs TC can be expressed as:

$$TC = C_{1E}N_{1E} + C_{1M}N_{1M} + C_{1L}N_{1L} + C_{2E}N_{2E} + C_{2M}N_{2M} + C_{2L}N_{2L} \tag{19}$$

Substituting the equilibrium conditions for the group sizes and the generalised costs allows to establish the conditions to be met for optimality of charge levels:

$$\frac{dTC}{d\rho_1} = 2s\varepsilon\rho_1 - s\zeta\rho_2 - (\varepsilon - \zeta)N\delta = 0 \tag{20}$$

and

$$\frac{dTC}{d\rho_2} = s\zeta\rho_1 - 2s\varepsilon\rho_2 - \varepsilon N\delta = 0 \tag{21}$$

with:

- $\varepsilon = (2\alpha + 2\gamma)/(\alpha+\gamma)^2 + 1/\beta$
- $\zeta = 4\gamma/(\alpha+\gamma)^2 + 1/\beta$

Solving (20) and (21) for ρ_1 and ρ_2 yields:

$$\rho_1 = \frac{\varepsilon}{2\varepsilon + \zeta}\delta\frac{N}{s} \tag{22}$$

and

$$\rho_2 = \frac{\varepsilon + \zeta}{2\varepsilon + \zeta}\delta\frac{N}{s} \tag{23}$$

Total waiting time and schedule delay are then:

$$TC = \left(1 - \frac{\varepsilon^2\delta}{2\varepsilon + \zeta}\right)\delta\frac{N^2}{s} \tag{24}$$

As for the size of the groups we find that:

$$N_{1E} + N_{1L} = N_{2E} + N_{2L} = \frac{\varepsilon^2}{2\varepsilon + \zeta} \delta N \tag{25}$$

Throughout our analysis we assumed that $N_{2L} > 0$. This condition can be tested for using the expressions for optimal charges (22) and (23) to calculate the optimal group size N_{2L} using expression (18). Under the condition that both α and β are positive, it can be shown that the optimal group size of N_{2L} is strictly positive as well.

The optimal generic coarse charge presented by Arnott et al. (1990) is a special case of our differentiated scheme where $N_{2E} + N_{2L} = 0$. While we do not claim that the proposed scheme for a differentiated coarse charge (with welfare optimal parameters values) is optimal with respect to alternative schemes, it is obvious from expression (25) which implies $N_{2E} + N_{2L} > 0$ under all circumstances that it does outperform the generic coarse charge.

Comparing figures 4 and 5, it is obvious the the price paid by travellers in group 2 exceeds the price paid by group 1 by an amount of $\rho_2 - \rho_1$. Hence the described equilibrium can only exist if group membership is determined exogenously. In case group choice must be endogenously, an additional uniform charge of $\rho_2 - \rho_1$ needs to be introduced for group 1 to allow for an equilibrium between both groups (but such a charging scheme violates our definition of a coarse charge).

The two groups coarse charge can be converted in an equally well performing two groups coarse reward (under the assumption of inelastic demand). The higher reward then corresponds to a large time slot compared to the lower reward. This seems a very acceptable proposition in a reward setting where travellers could be attributed randomly to one of both schemes. Moreover is there an equilibrium between both groups as they face the same (unrewarded) travel cost for arrivals in the t^* interval, so the coarse reward scheme allows for endogenous group choice without further adaptations.

5 Application

In this section we illustrate the case of the proposed differentiated coarse charge by using generic values of travel time and schedule delay to calculate optimal charge levels.

Assume a bottleneck capacity s of 500 cars per hour and an inelastic peak demand N of 1000 cars. Preferred arrival time t^* is 9 am and value of travel time α is €10 per hour. Schedule delay early β is valued at €5 per hour, and schedule delay late γ at €20 per hour.

The no-charge alternative involves travel time and schedule delay costs of €4000 each. Total social congestion costs TC are €8000. A first-best time varying charge eliminates all waiting and hence overall social costs are equal to schedule delay costs only. Total social cost TC is then equal to €4000.

The coarse toll as presented by Arnott et al. (1990) would amount to $\rho =$ €4. A group of 400 travellers would travel before the charge is turned on and 133 after the charge is turned off. The first traveller departs (and arrives) at 7.27 am and the last traveller arrives at 9.27 am. The charge is turned on at 8.15 am and turned off at 9.11 am. Total social costs TC are €5869 in this scenario, the corresponding welfare gain amounts to 53% of the level realised by the first-best charging scheme.

We use the analysis from the previous section to determine optimal implementation of a two groups differentiated coarse charge. Both groups N_1 and N_2 have a minimal

size of $N_{1E}+N_{1L}=N_{2E}+N_{2L}=346$ travellers. The remaining 308 peak travellers can be attributed at either scheme or be exempt from any charge.

The first group N_1 faces a charge of $\rho_1 = €2{,}59$ between 8.30 am and 9.07 am. The subgroup N_{1E} travelling before the charge is turned on has 260 travellers, of which the first one departs (and arrives) at 7.59 am. After the charge ends another subgroup N_{1L} of 86 travellers departs, the last traveller arriving at 9.17 am.

The second group N_2 faces a charge of $\rho_2 = €5{,}41$ between 7.59 am and 9.17 am. A subgroup N_{2E} of 281 travellers arrives before the charge is levied, with a first traveller activating the bottleneck at 7.25 am. A last group N_{2L} of 65 travellers passes through the bottleneck after the charge ends, with the last traveller arriving at 9.25 am.

Total travel time and schedule delay costs in this scenario amount to €5232. The two groups differentiated coarse charge realises 69% of the first-best welfare gain.

We observe that the generic coarse charge realises about $1/2$ of the first best welfare improvement, which goes up to roughly $2/3$ for the two groups coarse charge. This observation compares to the finding by Laih (1994) that the ratio of efficiency from the optimal n-step toll to that of the first best toll is $n/(n+1)$ for the case where $\gamma < \alpha$.

Looking for real world opportunities to apply a differentiated coarse charge, one could think of a bottleneck where two highways merge. In such case the inflow rate of traffic from both motorways in the bottleneck is pretty much stable. An application would then be to apply a different charging scheme on the two constituting flows of the bottleneck. A prerequisite is however that no substitution is possible for the individual traveller.

Another potential application for a differentiated coarse charge is where a congestion charge is levied at motorway access points. Charging schemes could in such a case be differentiated over access points.

6 Conclusions

There has been much debate on the optimal degree of differentiation to apply in battling road congestion. Normative economic theory has favoured infinite degrees of differentiation arguing that they correspond to welfare optimal settings. But technological and cognitive limitations have put a practical limit on the amount of differentiation that can be handled in real world cases.

In this paper we added to the debate by introducing a way to overcome the cognitive barrier towards more differentiation: while keeping the cognitive effort for the individual traveller constant, the welfare efficiency of a coarse charging scheme increases by differentiating its parameters across groups of travellers.

We derived analytically the optimal parameter values for a differentiated coarse charging scheme in the case where we limit to two groups. We illustrated our scheme using generic shadow cost values and find that differentiating the coarse charge over two groups increases the relative efficiency to 69% up from 53% for the undifferentiated coarse charge. This increase is of a magnitude which is comparable to the welfare gain that is obtained by moving from the coarse charge to a generic two step charge.

The presented charging regime has a considerable degree of flexibility with respect to implementation. Not only allows the optimal setting for a significant tolerance with respect to the relative size of the groups attributed to each scheme, also is it possible to exempt a considerable share of the travellers without facing a negative welfare impact. Moreover, our analysis makes clear that similar welfare gains can be obtained by non-price measures, such as number plate policies, that simply ban certain sub-groups from

travelling during certain intervals. With alternating groups, all users would benefit from such policies; at least when demand is perfectly inelastic. This robustness leaves plenty of space to accommodate for practical and political considerations which invariably come with real world implementations.

Acknowledgement

This study is financially supported by the TRANSUMO foundation.

References

Arnott, R., Palma, A. de and Lindsey, R. (1990). Economics of a Bottleneck. *Journal of Urban Economics*, *28*, 111–130.

Arnott, R., Palma, A. de and Lindsey, R. (1993). A Structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand. *American Economic Review*, *83*(1), 161–179.

Arnott, R., Palma, A. de and Lindsey, R. (1998). Recent developments in the bottleneck model. In K. Button and E. Verhoef (Eds.), *Road pricing, traffic congestion and the environment: Issues of efficiency and social feasibility.* Cheltenham, England: Edward Elgar.

Fosgerau, M. (2010, May). *How a fast lane may replace a congestion toll* [Working Paper]. Technical University of Denmark.

Laih, C.-H. (1994). Queueing at a bottleneck with single- and multi-step tolls. *Transportation Research Part A–Policy and Practice*, *28A*, 197–208.

Lindsey, R. and Verhoef, E. (2001). Traffic congestion and congestion pricing. In K. J. Button and D. A. Henscher (Eds.), *Handbook of Transport Systems and Traffic Control.* Amsterdam: Pergamon.

Norwood, F. B. (2006). Less Choice is Better, Sometimes. *Journal of Agricultural & Food Industrial Organization*, *4*(3), 1–21.

Shen, W. and Zhang, H. (in press). Pareto-improving ramp metering strategies for reducing congestion in the morning commute. *Transportation Research Part A–Policy and Practice*.

Small, K. A. (1982). The Scheduling of Consumer Activities: Work Trips. *American Economic Review*, *72*(3), 467–479.

Vickrey, W. S. (1969). Congestion Theory and Transport Investment. *American Economic Review*, *59*(2), 251–260.