TI 2010-062/3
Tinbergen Institute Discussion Paper

Give or Take? Rewards *vs* Charges for a Congested Bottleneck

Jan Rouwendal[a,b]
Erik T. Verhoef[a,b]
Jasper Knockaert[a]

[a] *VU University Amsterdam;*
[b] *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

Give or Take?

Rewards *vs* Charges for a Congested Bottleneck[a]

Jan Rouwendal[b,c], Erik T. Verhoef[c], Jasper Knockaert
Department of Spatial Economics
VU University Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands

This version: September 2, 2010

Key words: Traffic congestion; Road pricing; Subsidies; Rewards; Bottleneck model
JEL codes: R41, R48, D62

Abstract
This paper analyzes the possibilities to relieve congestion using rewards instead of taxes, as well as combinations of rewards and taxes. The model considers a Vickrey-ADL model of bottleneck congestion with endogenous scheduling. With inelastic demand, a fine (time-varying) reward is equivalent to a fine toll, and to a continuum of combinations of time-varying tolls and rewards (including fine feebates). When demand is price sensitive, a reward becomes less attractive from the efficiency viewpoint, because it attracts additional users to the congested bottleneck. As a result, both the second-best optimal rate of participation in the scheme, and the relative efficiency that can be achieved with it, decreases when demand becomes more elastic. Our analytical and simulation results for coarse schemes suggest that a coarse reward is less effective than a coarse feebate, which is itself less effective than a coarse toll. The most efficient coarse system is the step toll, which is also allowed to be positive in the shoulder period.

[a] Financial support from Transumo is gratefully acknowledged
[b] Corresponding author. Tel. 31-20-5986093. Fax 31-20-5986093. E-mail: jrouwendal@feweb.vu.nl
[c] Affiliated to the Tinbergen Institute, Roetersstraat 31, 1018 WB Amsterdam.

1. Introduction

Although the idea of road pricing has been around for a long time, and its popularity seems to be on the rise with successful introductions in cities like London, there still exists resistance against introducing a price for a commodity – access to public roads – that has been freely available for such a long time. As a result, there are many examples of road pricing proposals that do not survive the planning stage, and that are dropped for political reasons before implementation commences. Despite the clear economic case for marginal cost pricing, it may therefore be helpful to explore alternative possibilities to alleviate congestion, that are hopefully more appreciated by the public. An obvious possibility would be to think about the use of 'rewards' instead of tolls or 'penalties'. Even though most economists would hasten to point out that positive marginal external costs would call for a positive Pigouvian tax, that subsidies may induce all sorts of perverse incentives, that taxes needed to raise revenues to finance rewards are likely to cause distortions elsewhere in the economy, and that a net reward is likely to attract additional users whereas a net reduction is desired, there may be reasons to consider rewards anyway. Ultimately, a reward system may be less effective and efficient in combating congestion than a tolling system, but if the latter is unfeasible for political reasons, a more relevant comparison is between a reward system and the absence of any control through financial incentives. Such considerations have been an important motivation for proposals for, for example, combinations of pricing and rationing (Daganzo, 1995), revenue neutral 'credit-based' congestion pricing (Kalmanje and Kockelman, 2005), tradable driving permits (Verhoef, Nijkamp and Rietveld, 1997), exemptions from paying tolls (Daganzo and Garcia, 2000) and so-called Fast and Intertwined Regular (FAIR) lanes (De Corla-Souza, 2000).

It is the purpose of the present paper to explore the possibilities of a reward system. We do so by studying various types of rewards in the context of the well-known bottleneck model (Vickrey, 1969; Arnott, de Palma and Lindsey, 1990, 1993). The idea for this paper originates from a Dutch policy experiment called 'Spitsmijden' (Avoiding the Peak), documented in for example Knockaert, Verhoef and Rouwendal (2009). The purpose of this experiment was to gain insight into the potential of positive financial incentives in the management of peak congestion. To that end, regular users of a given highway (the A12 between Zoetermeer and The Hague) could earn rewards ranging from 3 to 7 Euros for avoiding it during the morning peak (7:30-9:30 am). Given that participation was voluntary, it is no surprise that relatively strong behavioural impacts from rewarding were found, with the number of morning-peak camera registrations of participants' cars roughly halved compared to pre- and post-measurements. Given that departure time adjustments were, by far, the most popular behavioural response to the incentive (roughly accounting for four-fifths of all adjusted trips), it seems very appropriate to use a dynamic model with endogenous departure time choice for the present paper. But because the remaining one-fifth of adjustments concerned alternatives such as public transport, we will be considering a bottleneck with a price-sensitive demand.

Our analysis fits in a much wider literature on second-best road congestion pricing reviewed in, for example, Small and Verhoef (2007). Probably closest to our paper are earlier studies that looked into the efficiency of so-called 'coarse tolls' and 'multi-step tolls' in the bottleneck model, because we too will be considering toll and reward schemes that involve a limited number of discrete levels during the peak. Arnott, de Palma and Lindsey (1990, 1993) were the first to consider this type of problem, and found that the relative efficiency of such measures is above that of 'uniform tolling', which entails a constant toll level throughout the peak, but considerably lower than that of 'fine' tolling, where tolls can be adjusted smoothly over time. The welfare gains are around 50% for the coarse (single-step) toll, a result that was later confirmed by Chu (1999) in a more elaborate model. Quite intuitively, Laih (1994) showed that the efficiency of step-tolls increase with the number of steps. Xiao, Qian and Zhang (2009) explore coarse tolling with heterogeneous users, and find that this improves the relative efficiency of coarse tolling. An important aspect of dynamic equilibria with step tolls in the bottleneck model, already identified by Arnott, de Palma and Lindsey (1990, 1993), is that it entails mass departures in the second part of the peak, since the equilibrium condition of constancy of generalized prices over times requires that a discrete drop in the toll level is matched by a discrete increase in (expected) travel time. The implied mass departures complicate the analytical treatment of step tolls.

The main difference with these earlier studies is our focus on rewards. More specifically, we will be looking at fine and coarse rewards, involving non-negative subsidies for all users; fine and coarse 'feebates', involving budget-neutral combinations of taxes and subsidies producing a zero net revenue for the regulator; and, as important references, fine and coarse tolls, involving non-negative taxes for all users. Given that congestion entails an external cost, one would expect that tolls outperform rewards and feebates in terms of efficiency. It is therefore not primarily the ranking of the different policies that we are most interested in, but rather their relative efficiency. Our analysis should thus give insight into the circumstances under which feebates or rewards may offer a worthwhile alternative to tolling in the management of congestion; and if so, how the policy should be designed to maximize its efficiency.

The paper is organized as follows. In the next section we introduce the basic idea in an informal way. The context is a simple version of the bottleneck model with a homogeneous population of commuters. We will introduce the model and discuss the equivalence of fine tolling, fine feebates and fine rewards when demand is completely inelastic. In Section 3 we study the properties of fine schemes in the context of price sensitive demand. We will require that there be an equilibrium in terms of participation to the reward scheme, in the sense that participants and non-participants face equal generalized equilibrium prices. We derive, for the different policies, the optimal shares of the commuters that should participate in the system to maximize the social benefits, subject to this participation condition. Next, in Section 4, we move another step closer to policy experiments of the type described above, by considering coarse systems, where tolls or rewards change in discrete steps during the peak. Section 5 concludes.

2. The basic bottleneck model and some variants on fine tolling

Our analysis in this section uses the basic bottleneck model, in which during the peak a homogeneous group of users of a given size N has to pass a bottleneck with given capacity s. The free-flow travel time is set equal to zero without loss of generality (given the other model assumptions), but a travel delay of $Q(t_d)/s$ is incurred if at the moment t_d of joining the queue, its length is $Q(t_d)$. As long as there is no queue and the inflow is below capacity s, there are no delays. We consider the following, conventional, cost function:

$$c(t) = \alpha T(t) + \beta \max\{0, t^* - t\} + \gamma \max\{0, t - t^*\} \qquad (1)$$

The Greek letters are positive parameters, with α denoting the 'value of time'; β indicates the unit shadow cost of schedule delay early and γ that of schedule delay late. T is the travel delay incurred in the queue before passing the bottleneck, t the arrival time at the bottleneck's exit, and t^* the preferred arrival time at which schedule delay costs are therefore zero.

Equilibrium for this model is discussed in detail in for example Arnott, de Palma and Lindsey (ADL) (1990). They show that in the no-toll equilibrium, denoted with superscript 0, the equilibrium cost of the commuting trip is for all drivers equal to:

$$c^0 = \delta \cdot \frac{N}{s} \equiv \frac{\beta \cdot \gamma}{\beta + \gamma} \cdot \frac{N}{s} \qquad (2)$$

where the composite schedule delay cost parameter δ is introduced for notational convenience.[1] The total social cost is therefore equal to $C^0 = \delta \cdot N^2 / s$. The peak hour starts at a moment that we will refer to as $B = t^* - (\delta / \beta) \cdot (N / s)$ and ends at $E = t^* + (\delta / \gamma) \cdot (N / s)$. At these moments, the queue has a zero length, and it is easily verified that the schedule delay costs at these two moments are both equal to $\delta . (N / s)$. The driver who passes at the desired arrival time t^*, in contrast, incurs no schedule delay costs, but only travel delay costs because of time spent in the queue. For all other drivers, the cost is a combination of positive schedule delay costs and positive travel time costs. The dynamic equilibrium in the model is thus caused by a queue that first grows and then shrinks linearly, so as to make the generalized cost in (1) constant over time.

ADL (1990) also show that the total cost can be reduced by 50% if a fine toll, that is a toll for which the value varies continuously over time, is introduced. The fine toll replicates the no-toll equilibrium pattern of travel delay costs and thus completely eliminates the queue and therefore all travel delay costs. It does so by substituting a monetary cost for the time cost. The crucial difference between both is that the time spent in the queue is a social loss, whereas the money paid as toll is merely a transfer. This creates the welfare gain from optimal pricing.

From the perspective of the present paper, it is interesting to observe that ADL (1990) present the optimal fine toll $\tau^f(t)$ as:

$$\tau^f(t) = \begin{cases} a - (t^* - t) \cdot \beta & \text{if} \quad t \le t^* \\ a - (t - t^*) \cdot \gamma & \text{if} \quad t > t^* \end{cases} \tag{3}$$

for all *t* during the peak. Here *a* is a constant that should satisfy $a \le \delta \cdot (N / s)$ if the toll is not to be positive outside the peak. When the inequality applies strictly, this formula implies that the toll is negative, and hence a subsidy, for some *t* in the peak period. Indeed, if we choose *a*=0, the toll will be non-positive for all drivers. In other words: the fine toll may in fact be a reward for some drivers, or even for all. Braid (1996) observes that it is in fact second-best optimal for a time-varying toll to start and end as a net subsidy at a bottleneck with an

[1] That is, $\delta \equiv \dfrac{\beta \cdot \gamma}{\beta + \gamma}$.

unpriced substitute.[2] This is needed to achieve equalization of marginal costs at the two bottlenecks in that second-best optimum.

In the basic bottleneck model, the fine toll that is always non-negative, with $a = \delta \cdot (N/s)$, is equivalent in its effects to a fine 'toll' that is always non-positive, with a = 0. The latter will be referred to in this paper as a fine reward. Other values of a may also be chosen. One potentially interesting possibility is a fine 'feebate', which we define as a budget-neutral fine toll that takes on positive as well as negative values and generates no net revenues. It requires $a = \frac{1}{2} \cdot \delta \cdot (N/s)$.

The insensitivity of welfare with respect to a stems from the assumed perfect inelasticity of demand in this basic bottleneck model. It can be shown that the price paid by each driver is equal to the marginal social cost under fine tolling if $a = \delta \cdot (N/s)$,and this is the optimal choice if demand is price-sensitive.[3] One would therefore also expect that the relative benefits of a reward system are smaller when demand is sensitive to the price. This motivates our choice for considering price-sensitive demand in the remainder of this paper. We will sometimes call this 'elastic' demand. Note that this term is sometimes used to denote an absolute elasticity above unity; in our paper it is used as a short-hand for price-sensitive demand that is not perfectly inelastic.

3. Elastic demand, fine tolling and rewarding, and endogenous participation

To analyze the situation in which demand is sensitive to the price, we postulate a downward sloping inverse demand function $D(N)$ and define social surplus S as the difference between aggregate consumer benefit and the cost $C(N)$ associated with using the bottleneck:

$$S = \int_0^N D(x)\,\mathrm{d}x - C(N) \tag{4}$$

The number of users N is now endogenously determined by the generalized price p associated with passing the bottleneck, because in equilibrium:

$$D(N) = p \tag{5}$$

Now assume that fine tolls/rewards are set using (3). Such a toll scheme eliminates all queuing at the bottleneck, and implies that the average generalized cost per user (that is, net of

[2] The substitute is a second bottleneck that suffers congestion too.

[3] Because optimal tolling eliminates queuing, the social cost becomes half the level in the no-toll equilibrium, and thus equals $\frac{1}{2} \cdot \delta \cdot N^2 / s$. The marginal cost is then $\delta \cdot N / s$, which is equal to the toll at t^* and hence the generalized price for the person arriving at t^* when $a = \delta \cdot N / s$. Because the generalized price is equalized in the decentralized optimum, this optimal price level applies for all drivers.

tolls) equals $\frac{1}{2}\cdot\delta\cdot(N/s)$. However, the policy maker is still free to set the generalized price by choosing the parameter *a*. Taking this into account, and realizing that the generalized price will be equal to *a* since the person arriving at t^* will have neither travel delay nor schedule delay cost, and only faces a toll *a*, and the generalized price will be equal for all users in equilibrium, we write the Lagrangian:

$$\Lambda = \int_0^N D(x)\,\mathrm{d}x - \tfrac{1}{2}\cdot\delta\cdot\frac{N^2}{s} + \lambda\cdot\left(a - D(N)\right) \tag{6}$$

and derive the first-order conditions:

$$\frac{\partial\Lambda}{\partial N} = D(N) - \delta\cdot\frac{N}{s} - \lambda\cdot D'(N) = 0 \tag{7a}$$

$$\frac{\partial\Lambda}{\partial a} = \lambda = 0 \tag{7b}$$

$$\frac{\partial\Lambda}{\partial\lambda} = a - D(N) = p - D(N) = 0 \tag{7c}$$

It follows from these conditions that *a* and hence *p* should equal $\delta\cdot(N/s)$, which implies a fine tolling system as defined by $a = \delta\cdot(N/s)$. Other values of *p* are incompatible with the first-best situation. The fact that the first-best price is uniquely determined implies that the optimal toll is also uniquely determined. Clearly, a reward system is incompatible with a first-best optimum when demand is price-sensitive.

This does, of course, not imply that a reward system can only play a useful role when demand is completely inelastic. In many actual policy situations first-best situations are – for various reasons – not implementable and second-best measures are the only relevant policy instruments, but may still achieve substantial welfare gains. In our example, a reward system may be used to remove part of the users from the peak, and thus alleviate the congestion problem. To see how this can be done, consider Figure 1, which refers to a no-toll equilibrium. The duration of the rush hour in equilibrium is indicated by the line segment *B*–*E*, which equals *N*/*s*. To see the potential usefulness of rewarding drivers, consider what happens if we could persuade a given number of drivers – to be denoted as N_P, the number of 'participants' to the rewarding scheme – not to use the bottleneck during the peak. The result is that the extent of the rush hour decreases, as the total number of commuters that wants to pass the bottleneck around time t^* is now lower. The base of the triangle in Figure 1 would shrink, but the slopes of the two sides would remain unchanged. Graphically, the original triangle moves downward. What remains could be the dashed triangle indicated in Figure 1.

Since the height of this new triangle is smaller than that of the original one, the equilibrium cost of all remaining unrewarded drivers has gone down. If their demand is completely inelastic, the number of drivers passing the bottleneck in the peak is equal to $N–N_P$. If demand is elastic, there will be induced demand: additional drivers start using the bottleneck in the peak after the average cost decreases because some original participants have been removed. The downward shift of the triangle will of course be smaller when there is induced demand.

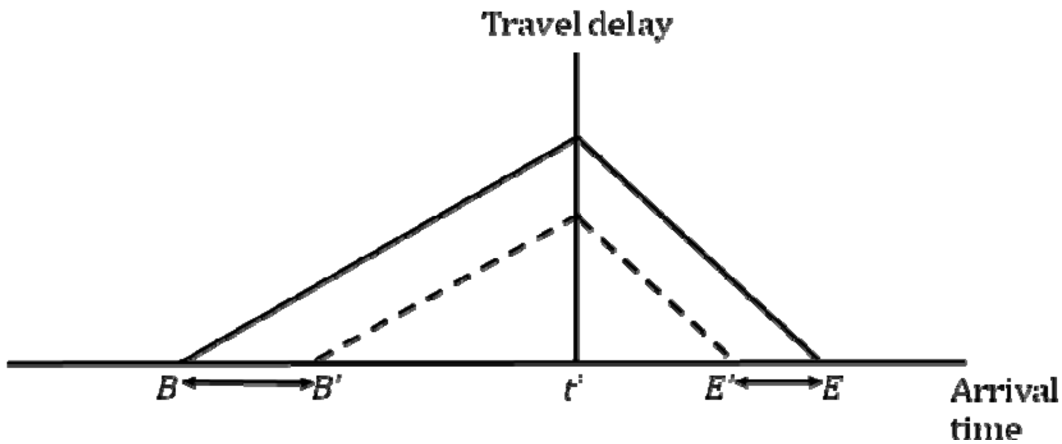

Figure 1. Effect of removing some commuters from the rush hour

Besides options such as changing mode or suppressing the trip altogether, rewarded drivers may pass the bottleneck outside the new peak period B'–E'. In fact, because the peak period has become shorter, there cannot be an equilibrium if none of the rewarding drivers would use that option: the generalized price for travelling just before B' or after E' would then be lower than the original price. If they do use the bottleneck, to minimize their schedule delay cost, they should be allowed to pass it right before B' and after E'. Because the capacity of the bottleneck does not change, all rewarded drivers could then pass the bottleneck during the two time intervals $[B,B']$ and $[E',E]$ when demand is perfectly inelastic. If demand is price-sensitive and additional unrewarded drivers are attracted, some rewarded drivers have to pass the bottleneck before B or after E. This is illustrated in Figure 2. The rewarded 'participants' pass the bottleneck during the time intervals $[B'',B']$ and $[E',E'']$. The absence of queuing by participants requires the reward to vary over time during these intervals in a way similar to the pattern of the toll for first-best pricing (*i.e.,* the reward should fall at a rate β for early arrivals and increase at a rate γ for late arrivals). Time-invariant rewards will be considered later. There are no rewards, but time-varying travel delays, in the remaining peak period <B',E'>.

When induced demand increases the duration of this remaining peak period [B',E'], average cost rises both for the unrewarded and for the rewarded drivers. This suggests that induced demand makes the reward system less efficient. To investigate this more formally, we derive the relevant cost functions. First observe that the average cost for the unrewarded drivers is identical to what it would be in a conventional bottleneck equilibrium with $N_N \equiv N - N_P$ users: $c_N = \delta \cdot (N_N / s)$. If we use fine rewarding for the participants, they only experience schedule delay costs, which are on average equal to $c_P = \frac{1}{2} \cdot \delta \cdot (N_N + N) / s$. For both cost levels, the levels of use of course refer to traffic volumes in the new equilibrium.

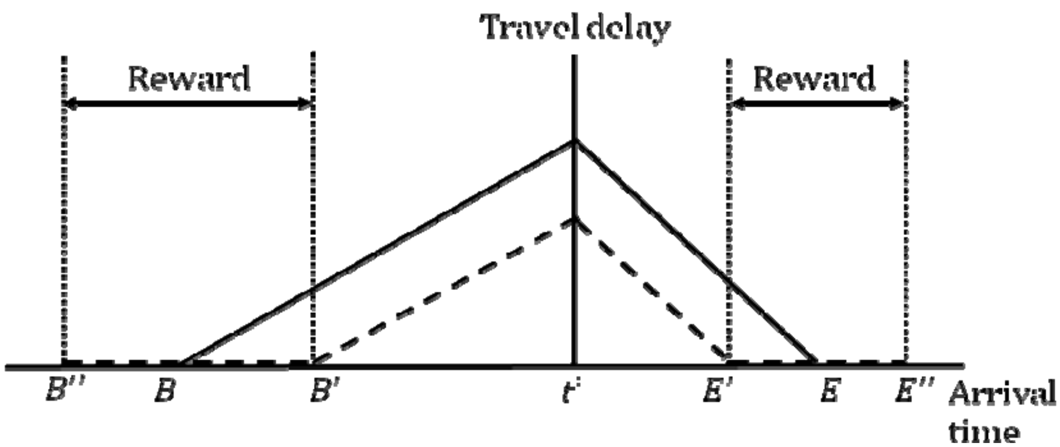

Figure 2. Equilibrium with time-varying rewards for passing the bottleneck early or late

If participation in the reward system is voluntary and open to all, as we will assume, the equilibrium generalized prices must be equal for all users of the bottleneck. For non-participants this price is equal to their average cost. For participants, the fine reward has to be equal, at each instant, to the difference between their schedule delay cost and the generalized price for non-participants, c_N. Because the reward itself is a transfer and we ignore the cost of public funds, it is not part of the social cost $C(N_N, N_P)$:

$$C(N_N, N_P) = N_N \cdot c_N + N_P \cdot c_P$$
$$= \delta \cdot \frac{N_N^2}{s} + \delta \cdot \frac{(N - N_N) \cdot \frac{1}{2} \cdot (N + N_N)}{s} = \frac{1}{2} \cdot \delta \cdot \frac{N^2 + N_N^2}{s} \qquad (8)$$

If demand is completely inelastic, maximization of social surplus is equivalent to minimizing social costs, and it is easily verified in (8) that in this case it is optimal to let all drivers participate in the reward so that $N_N = 0$. The time-varying reward then reproduces the first-best outcome. Its time variation eliminates all queuing, and the associated below-optimal generalized price does not distort overall demand because of the perfect inelasticity. When demand is price-sensitive, in contrast, there is the induced-demand effect of rewarding to take into account, because the first-best optimality conditions of (7a) and (7b) can no longer be fulfilled. To see how this affects the second-best reward, we maximize social surplus by solving the Lagrangian:

$$\Lambda = \int_0^N D(x)\mathrm{d}x - \tfrac{1}{2}\cdot\delta\cdot\frac{N^2 + N_N^2}{s} + \lambda\cdot\left(\delta\cdot\frac{N_N}{s} - D(N)\right) \tag{9}$$

The constraint reflects that the generalized price will be equal for all drivers, and will amount to c_N as derived above. The first-order conditions are as follows:

$$\frac{\partial\Lambda}{\partial N} = D(N) - \delta\cdot\frac{N}{s} - \lambda\cdot D'(N) = 0 \tag{10a}$$

$$\frac{\partial\Lambda}{\partial N_N} = -\delta\cdot\frac{N_N}{s} + \lambda\cdot\frac{\delta}{s} = 0 \tag{10b}$$

$$\frac{\partial\Lambda}{\partial\lambda} = \delta\cdot\frac{N_N}{s} - D(N) = 0 \tag{10c}$$

where $D' \equiv \mathrm{d}D/\mathrm{d}N$. These first-order conditions imply that the relative numbers of non-rewarded and rewarded drivers should be set such that:

$$N_N = \frac{\delta/s}{\delta/s - D'(N)}\cdot N \quad \Leftrightarrow \quad N_P = \frac{-D'(N)}{\delta/s - D'(N)}\cdot N \tag{10d}$$

This confirms our earlier conclusion that with perfectly inelastic demand ($D'(N) \to -\infty$) all drivers should participate to maximize social surplus: it implies $N_N = 0; N_P = N$. In contrast, with perfectly inelastic demand ($D'(N) = 0$) no-one should participate: $N_N = N; N_P = 0$. The reason is that latent demand is then so strong that any attempt to decrease the number of drivers in the remaining peak hour is fully ineffective, as the number of newly attracted drivers is exactly equal to the number of removed 'rewarded' drivers. Hence there are no benefits from reduced congestion associated with removing some drivers from the peak, while there will be a social loss from subsidizing travel outside the peak period because it attracts drivers with a marginal benefit below the marginal social cost. The fine reward is then a completely ineffective instrument, and cannot produce any welfare gain.

For intermediate elasticities, we expect the relative efficiency of fine rewards to decrease with demand sensitivity, along with the second-best optimal share of rewarded drivers as implied by (10d). Figure 3 confirms this by depicting for a numerical example the relative efficiency of fine rewarding for various levels of demand elasticity (in the no-toll equilibrium) and for the full range of possible participation levels. The latter is indicated as $\pi = N_P/N$ along the horizontal axis. On the vertical axis, ω denotes the relative efficiency: the gain in social surplus with the reward system as a fraction of the surplus in the first best optimum. The results are obtained for a numerical specification with the following parameters: α=7.5; β=3.75; γ=15; N=9000; and s=3600. The implied ratios of utility parameters reflect the usual values (*e.g.,* Small, 1982; Arnott, de Palma and Lindsey, 1993); the value for α when expressed in Euros is close (rounded) to the current "official" value in The Netherlands; and the implied peak duration of 2.5 (hours) seems reasonable for a morning peak bottleneck. The two parameters in the assumed linear demand function are chosen such that both the target level of N and the target demand elasticity are achieved in the base (unpriced) equilibrium.

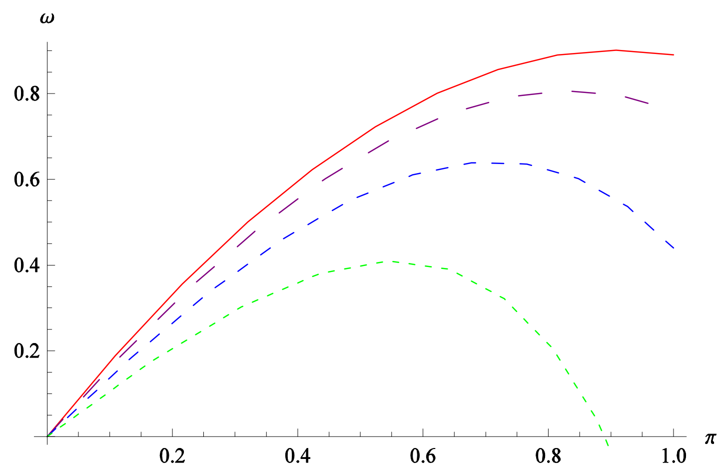

Figure 3. Relative efficiency (ω) of fine rewarding by participation rate (π) for various elasticities of demand in the no-toll equilibrium: –0.1 (red, solid), –0.2 (purple, large dashing), –0.4 (blue, medium dashing), and –0.8 (green, small dashing)

The upper solid line in the Figure 3 shows that with a small absolute value of the elasticity, the efficiency of the reward system approaches that of the first-best situation: ω approaches 1. It is also optimal to let almost all drivers participate in the reward system: the top of the curve is not far from π=1. When demand is more elastic, represented by the lower lines, both the relative welfare gains and the optimal share of participants – the π for which a maximum ω is reached – decrease. This confirms the conjecture that strong latent demand effects make a reward system less effective. Note that for the highest (absolute value of the) demand elasticity, a reward system where more than 90% of the drivers participate in fact lowers welfare compared to the no-toll equilibrium. The adverse induced demand effect then outweighs the benefit of eliminating queuing for the rewarded drivers.

4. Coarse rewards, feebates and tolls

We now turn to consider a simpler but probably more realistic version of the reward system. This concerns a 'coarse' reward, that has the same value independent of the moment of driving – as long, of course, as it is outside some specified central peak period. Such a coarse reward is some sort of mirror image of the coarse toll considered by Arnott, de Palma and Lindsey (1990, 1993), which is a uniform toll that is levied during the central part of the peak period only, for instance between B′ and *E*′ in Figures 1 and 2, and zero outside that period – including the shoulders of the peak. The coarse reward, in contrast, is zero in the central part of the peak and positive in the shoulders. We will discuss the similarities and differences between the two systems below. In addition, we will consider what we will call the coarse 'feebate', defined as a budget neutral combination of a coarse toll (in the central part of the peak) and reward (in the shoulders). Not surprisingly, we will find that the performance of this measure is between that of the coarse toll and the coarse reward. For completeness, we will also consider an unrestricted 'step toll', for which it is assumed that the peak is divided in three periods during which differentiated shoulder and central toll or reward levels can apply, but no constraints on the signs or values of these tolls or their net revenues have to be met. In other words, this is a benchmark regime, that gives the maximum achievable welfare when the peak is to be subdivided in three periods with constant toll or reward levels.

Because tolls are piecewise constant in the regimes considered in this section, queuing will remain existent in equilibrium, for the drivers in the shoulders ('participants' in the previous section) as well as in the central period (the non-participants).[4] In all equilibria that we will consider, all drivers experience the same generalized prices, which is consistent with endogenous participation for a reward scheme. The costs of the two groups differ, and the difference in toll or reward levels is used to compensate for this difference, and to establish a market equilibrium. The only difference between the four regimes will be the way in which the cost difference is compensated for: through a reward given to the high-cost group, a coarse toll levied on the low-cost group, a toll difference between both groups (for the step toll), or a combination of a reward and a toll (for the feebate).

[4] For notational convenience, we will continue to use the subscript P for users in the shoulders (think of pre- and post periods) and N for those in the central period.

We start the discussion by emphasizing that, as explained by Arnott, De Palma and Lindsey (1990), equilibrium requires a mass departure at instant E', which leads to the sudden emergence of a queue at that moment. Only then is it possible that the generalized price just before E', when a higher toll or lower reward applies, is equal to the (expected) generalized price right after E'. All travellers in the mass departure have the same probability of obtaining a particular place in the queue, so that all users who pass the bottleneck between E' and E'' experience the same expected cost. The first one can pass the bottleneck at E', and experiences only a schedule delay cost equal to $\gamma \cdot (E' - t^*)$. The final one passes at E'', and therefore spends time $E'' - E'$ in the queue, implying a generalized cost $\alpha \cdot (E'' - E') + \gamma \cdot (E'' - t^*)$. Because the realization of generalized cost rises linearly between these two instants, the expected travel cost of the users in the mass departure is the average of these extremes, and therefore equals $\frac{1}{2} \cdot \gamma \cdot (E' - t^*) + \frac{1}{2} \cdot \left[\alpha \cdot (E'' - E') + \gamma \cdot (E'' - t^*)\right]$.

Throughout our analysis of coarse instruments we will assume $\gamma > \alpha$. This is generally believed to be the more relevant case empirically for car traffic in the morning peak (e.g. Small, 1982; Arnott, de Palma and Lindsey, 1990, 1993). As pointed out by Arnott, de Palma and Lindsey (1990) the relative size of α and γ is important in the analysis of coarse pricing because it affects the qualitative properties of the equilibrium. In particular, when $\gamma < \alpha$, there will be users departing after the mass departure, joining the queue before it has fully dissipated and the departure time interval will not shift as it does with $\gamma > \alpha$.

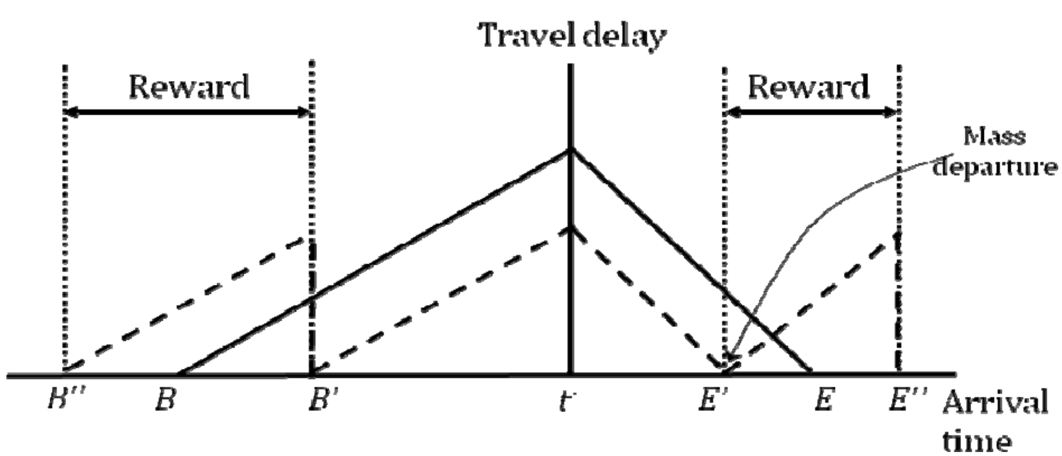

Figure 4. Travel delay in the coarse systems

In Figure 4, travel delay is pictured as a function of the time at which the bottleneck is passed. The solid line gives the original equilibrium, and the dashed line that with a coarse reward.

Since demand is price-sensitive and the generalized price decreases with a reward, the total duration of the peak increases. The dynamic user equilibrium condition implies that a queue will start growing at B''. This queue will be completely resolved at B', because it is not optimal to let central users already depart when there is still a queue of early shoulder users. The queue for central users starts growing at B', and shrinking after t^*, until it has resolved at E'. At E', there next is the mass departure of participants, and depending on their place in the queue, there is again a positive waiting time. The user who passes at E'' has both the longest travel time and the largest schedule delay late.

We can now establish the generalized costs for the participants, c_P, and non-participants, c_N, as functions of the numbers of drivers. These relations are:

$$c_N = \delta \cdot \frac{N_N}{s} \quad \text{with} \quad \delta \equiv \frac{\beta \cdot \gamma}{\beta + \gamma} \tag{11a}$$

$$c_P = c_N + \delta^* \cdot \frac{N_P}{s} \quad \text{with} \quad \delta^* \equiv \frac{\beta \cdot \frac{1}{2} \cdot (\alpha + \gamma)}{\beta + \frac{1}{2} \cdot (\alpha + \gamma)} \tag{11b}$$

Equation (11a) follows immediately from noting that the cost for the non-participants will be equal to the level that would apply in a conventional no-toll equilibrium with N_N drivers; compare equation (2). Equation (11b) follows from solving the set of two equations that, first, defines the two periods of shoulder arrivals to be together sufficiently long to host these drivers, and secondly, that equates the (expected) generalized cost to be equal for early and late shoulder drivers:

$$\frac{N_P}{s} = (B' - B'') + (E'' - E') \tag{12a}$$

$$\beta \cdot (B' - B'') + c_N = \tfrac{1}{2} \cdot (\alpha + \gamma) \cdot (E'' - E') + c_N \tag{12b}$$

Equations (12ab) can be solved to produce:

$$(B' - B'') = \frac{\frac{1}{2} \cdot (\alpha + \gamma)}{\beta + \frac{1}{2} \cdot (\alpha + \gamma)} \cdot \frac{N_P}{s} \tag{12c}$$

Next, observe that c_P is equal to the sum of c_N and the additional schedule delay cost at B'' compared to B':

$$c_P = c_N + \beta \cdot (B' - B'') \tag{12d}$$

Substitution of (12c) into (12d) then produces the desired result of (11b). Note that (11b) applies whenever the toll or reward levels in the shoulders, for the participants, are equalized for the early and late shoulders, as we will assume. It is therefore not an optimality condition that would only apply for optimized tolls or rewards.

It is clear from (11b) that the shoulder users have a higher generalized cost than the central users. User equilibrium requires that the generalized prices of both groups are equal; the difference in toll or reward levels should bridge the gap. That is, the net difference Δ between toll and/or reward levels in the central versus the peak period should for all schemes be equal to:

$$\Delta = \delta^* \cdot \frac{N_P}{s} \tag{13}$$

In what follows, we apply these results to find the optimal numbers of users in the various systems. We do so by maximizing social surplus under the side-condition that the generalized price faced by all users will be the same. The exact formulation of this condition differs over the systems.

Coarse reward

A coarse reward attracts some users to the shoulder periods; the equilibrium generalized price will be equal to the generalized cost for central drivers. The Lagrangian is:[5]

$$\Lambda = \int_0^N D(x)\mathrm{d}x - \delta \cdot \frac{N \cdot N_N}{s} - \delta^* \cdot \frac{(N - N_N)^2}{s} + \lambda \cdot \left(\delta \cdot \frac{N_N}{s} - D(N) \right) \tag{14}$$

Note that the constraint is the same as in problem (9) with fine rewards. The first-order conditions are as follows:

$$\frac{\partial \Lambda}{\partial N} = D(N) - \delta \cdot \frac{N_N}{s} - 2 \cdot \delta^* \cdot \frac{N - N_N}{s} - \lambda \cdot D'(N) = 0 \tag{15a}$$

$$\frac{\partial \Lambda}{\partial N_N} = -\delta \cdot \frac{N}{s} + 2 \cdot \delta^* \cdot \frac{N - N_N}{s} + \lambda \cdot \frac{\delta}{s} = 0 \tag{15b}$$

$$\frac{\partial \Lambda}{\partial \lambda} = \delta \cdot \frac{N_N}{s} - D(N) = 0 \tag{15c}$$

Equation (15c) can be substituted in (15a) to derive the second-best equilibrium expression for λ:

$$\lambda = \frac{2 \cdot \delta^* \cdot \frac{N - N_N}{s}}{-D'(N)} \tag{15d}$$

Recall that this shadow price should give the marginal impact on second-best optimized social surplus from a relaxation of the constraint; *i.e.*, from an increase in the generalized price. With

[5] The formulation for total cost in (14) follows after reorganization of $N_N \cdot c_N + (N - N_N) \cdot c_P$; note that we have substituted out $N_P = N - N_N$.

rewarding, this generalized price is below marginal social cost, so the shadow price is positive. The denominator shows that the shadow price decreases and approaches zero as demand becomes less elastic, and induced demand is consequently less distortive. The numerator shows that the shadow price decreases as fewer users are rewarded, capacity is higher, and the composite cost coefficient δ^* is smaller.

Substitution of (15d) in (15b) reveals, after some manipulations, that the ratio of rewarded shoulder travelers to total use should amount to:

$$\frac{N_P}{N} = \tfrac{1}{2} \cdot \frac{\delta}{\delta^*} \cdot \frac{-D'}{\delta / s - D'} \tag{15e}$$

The third term gives the ratio that was found for a fine reward, in (10d), and the interpretation is similar. In particular, the ratio approaches zero as demand approaches perfect elasticity, for the same reason as with a fine reward. The upper limit, which applies with perfectly inelastic demand, however, is now not a share of unity, as for the fine reward, but a share of $\tfrac{1}{2} \cdot \delta / \delta^*$. This is the same share that Arnott, De Palma and Lindsey (1990) find for a coarse toll.[6] This reflects that for perfectly inelastic demand, the coarse reward is as efficient as the coarse toll. The two instruments are then equally effective in affecting departure time choice, the only relevant margin of behaviour if there is no induced demand. The instruments should therefore be set in an similar way, the only difference being that the shoulder users now receive a subsidy that is equal to the tax that central users would pay in the coarse toll schedule.

Finally, the equilibrium value for the coarse reward σ can be found by substitution of (15e) into (13). This produces:

$$\sigma = \tfrac{1}{2} \cdot \frac{\delta}{s} \cdot \frac{-D'}{\delta / s - D'} \cdot N \tag{15f}$$

For perfectly inelastic demand, this again simplifies to an expression that applies for the coarse toll of Arnott, De Palma and Lindsey (1990).[7] Note that, given N, this level is equal to the average level of the fine toll. For perfectly elastic demand, the second-best optimal reward becomes zero.

Coarse toll

[6] Arnott, De Palma and Lindsey do not mention this share explicitly. But using their expressions for the coarse toll (their equation 14a), and the moments at which it is switched on (their 14c) and off (their 14d), this share can be computed, and it indeed turns out to be equal to the one in our equation (15e).
[7] See their equation (14a).

As mentioned, the coarse reward resembles the coarse toll, the difference being that in the former case a subsidy is given, whereas in the latter a tax applies. The coarse toll is therefore a natural reference for judging the performance of the coarse reward. Arnott, De Palma and Lindsey (1990) derive the second-best coarse toll for perfectly inelastic demand. In their later paper, Arnott, De Palma and Lindsey (1993) consider coarse tolling with price-sensitive demand, but they do not derive the corresponding second-best toll level analytically. To derive it, we solve a Lagrangian that is rather similar to the one in (14), the difference being that the generalized price will now be the generalized cost level incurred by the shoulder drivers, given in (11b), rather than the central drivers, given in (11a):

$$\Lambda = \int_0^N D(x)\mathrm{d}x - \delta \cdot \frac{N \cdot N_N}{s} - \delta^* \cdot \frac{(N-N_N)^2}{s} + \lambda \cdot \left(\delta \cdot \frac{N_N}{s} + \delta^* \cdot \frac{N-N_N}{s} - D(N) \right) \tag{16}$$

The first-order conditions are now:

$$\frac{\partial \Lambda}{\partial N} = D(N) - \delta \cdot \frac{N_N}{s} - 2 \cdot \delta^* \cdot \frac{N-N_N}{s} + \lambda \cdot \left(\frac{\delta^*}{s} - D'(N) \right) = 0 \tag{17a}$$

$$\frac{\partial \Lambda}{\partial N_N} = -\delta \cdot \frac{N}{s} + 2 \cdot \delta^* \cdot \frac{N-N_N}{s} + \lambda \cdot \frac{\delta - \delta^*}{s} = 0 \tag{17b}$$

$$\frac{\partial \Lambda}{\partial \lambda} = \delta \cdot \frac{N_N}{s} + \delta^* \cdot \frac{N-N_N}{s} - D(N) = 0 \tag{17c}$$

Equations (17c) and (17a) imply the following second-best equilibrium expression for λ:

$$\lambda = \frac{\delta^* \cdot \frac{N-N_N}{s}}{\frac{\delta^*}{s} - D'(N)} \tag{17d}$$

For given levels of the right-hand side variables, this implies a lower shadow price than for the coarse reward in (15d): the numerator is now twice as small, whereas the denominator is larger. This lower shadow price reflects that the problem of induced demand is, for the coarse toll, smaller than for the coarse reward. But the shadow price is still positive, reflecting that an increase in the generalized price by raising the shoulder toll level (now zero) would raise welfare – see also the step toll below.

Substitution of (17d) in (17b) shows that the ratio of shoulder travelers to total use should now amount to:

$$\frac{N_P}{N} = \tfrac{1}{2} \cdot \frac{\delta}{\delta^*} \cdot \frac{\delta^*/s - D'}{(\delta + \delta^*)/2s - D'} \tag{17e}$$

For perfectly inelastic demand, we again find a share of $\frac{1}{2}\cdot\delta/\delta^*$, which is again equal to the ratio implied by the results of Arnott, De Palma and Lindsey (1990) for a coarse toll and perfectly inelastic demand. This ratio is therefore the same as for the coarse reward with perfectly inelastic demand. However, when approaching perfect elasticity of demand, (17e) implies a positive ratio of $\delta/(\delta+\delta^*)$, which for example equals ½ if $\alpha=\gamma$. This is in sharp contrast with the results for the coarse reward, where the share would approach zero. It reflects that the toll has a positive effect of discouraging users of the congested bottleneck, whereas the reward tends to attract users. Note that the ratio of $\delta/(\delta+\delta^*)$ for perfectly elastic demand is equal to that of $\frac{1}{2}\cdot\delta/\delta^*$ for perfectly inelastic demand when $\alpha=\gamma$ (so that $\delta=\delta^*$).[8]

Finally, the equilibrium value for the coarse toll implied by (13) and (17e) is:

$$\tau=\tfrac{1}{2}\cdot\frac{\delta}{s}\cdot\frac{\delta^*/s-D'}{\left(\delta+\delta^*\right)/2s-D'}\cdot N \tag{17f}$$

Consistent with what was just said, this toll remains positive for a perfectly elastic demand, while the reward in (15f) becomes zero. Also note that for given right-hand side variables, the toll will certainly be larger than the subsidy as long as $\gamma>\alpha\Rightarrow\delta>\delta^*$, a condition usually found and assumed to apply in reality (*e.g.*, Small, 1982). Under those conditions, the share in (17e) will also certainly exceed that in (15e). Note that this restriction $\gamma>\alpha$ is more than sufficient. This confirms the intuitive expectation that the coarse toll is, in absolute terms, bigger than the coarse reward, because it does not attract additional users to an already congested facility. As anticipated, with perfectly inelastic demand, the toll is again equal to the level found by Arnott, De Palma and Lindsey (1990).

Step toll

The step toll is quite similar to the coarse toll just discussed, but allows a positive toll to be charged also in the shoulders. This shoulder toll may – and will – be different from the central toll. One may expect the step toll to be at least as efficient as the coarse toll, because the shoulder toll may be set at zero, which would make the two policies identical. The Lagrangian now becomes:

[8] When $\alpha<\gamma$, $\delta>\delta^*$, and the ratio of shoulder travellers with perfectly elastic demand is smaller than that with perfectly inelastic demand. There does not seem to be a simple intuition behind these results.

$$\Lambda = \int_0^N D(x)\mathrm{d}x - \delta \cdot \frac{N \cdot N_N}{s} - \delta^* \cdot \frac{(N-N_N)^2}{s} + \lambda \cdot \left(\delta \cdot \frac{N_N}{s} + \delta^* \cdot \frac{N-N_N}{s} + \tau_L - D(N) \right) \quad (18)$$

where τ_L denotes the ("low") shoulder toll. The first-order conditions are now:

$$\frac{\partial \Lambda}{\partial N} = D(N) - \delta \cdot \frac{N_N}{s} - 2 \cdot \delta^* \cdot \frac{N-N_N}{s} + \lambda \cdot \left(\frac{\delta^*}{s} - D'(N) \right) = 0 \quad (19a)$$

$$\frac{\partial \Lambda}{\partial N_N} = -\delta \cdot \frac{N}{s} + 2 \cdot \delta^* \cdot \frac{N-N_N}{s} + \lambda \cdot \frac{\delta - \delta^*}{s} = 0 \quad (19b)$$

$$\frac{\partial \Lambda}{\partial \lambda} = \delta \cdot \frac{N_N}{s} + \delta^* \cdot \frac{N-N_N}{s} + \tau_L - D(N) = 0 \quad (19c)$$

$$\frac{\partial \Lambda}{\partial \tau_L} = \lambda = 0 \quad (19d)$$

The shadow price λ is now zero, reflecting that the generalized price and hence overall demand can be controlled perfectly using the shoulder toll level τ_L. The zero shadow price reflects that, among the coarse schedules considered, this step toll is the most efficient: there is no binding constraint on the two toll levels (given the symmetric two-levels structure).

The ratio of shoulder users can be found after substituting (19d) into (19b):

$$\frac{N_P}{N} = \tfrac{1}{2} \cdot \frac{\delta}{\delta^*} \quad (19e)$$

Given that overall demand can now be controlled perfectly using τ_L, it is no surprise that this ratio is the same as what was found for the coarse toll and the coarse reward above for perfectly inelastic demand. There is now no reason to deviate from this ratio in an attempt to restrict demand.

Substitution of (19d) and (19c) into (19a) gives the level of the low toll:

$$\tau_L = \delta^* \cdot \frac{N_P}{s} = \tfrac{1}{2} \cdot \delta \cdot \frac{N}{s} \quad (19f)$$

Adding Δ then gives the ("high") central toll τ_H:

$$\tau_H = \delta \cdot \frac{N}{s} \quad (19g)$$

The central toll is therefore twice as large as the shoulder toll. The shoulder toll is now equal to the average level of what the fine toll would be with N users, and the central toll is equal to its maximum. The total revenues from step tolling are therefore, for a given N, higher than those from fine tolling.

Feebate

Finally, we consider the condition of a budget neutral combination of a toll τ that has to be paid by the central drivers, and a reward σ that is received by the shoulder drivers. The revenues from the toll have to be equal to the expenditure on the reward, and (13) still applies, so we can find an expression for τ as a function of N and N_P:

$$\begin{cases} \tau + \sigma = \delta^* \cdot \dfrac{N_P}{s} \\ \sigma = \tau \cdot \dfrac{N_N}{N_P} \end{cases} \Rightarrow \tau = \delta^* \cdot \frac{N_P^2}{s \cdot N} \tag{20a}$$

Adding c_N gives the generalized price in equilibrium:

$$p = \delta \cdot \frac{N_N}{s} + \delta^* \cdot \frac{(N - N_N)^2}{s \cdot N} = \delta \cdot \frac{N_N}{s} + \delta^* \cdot \left(\frac{N}{s} - 2 \cdot \frac{N_N}{s} + \frac{N_N^2}{s \cdot N} \right) \tag{20b}$$

This means that the Lagrangian becomes:

$$\begin{aligned} \Lambda = \int_0^N D(x)\mathrm{d}x - \delta \cdot \frac{N \cdot N_N}{s} - \delta^* \cdot \frac{(N - N_N)^2}{s} \\ + \lambda \cdot \left(\delta \cdot \frac{N_N}{s} + \delta^* \cdot \left(\frac{N}{s} - 2 \cdot \frac{N_N}{s} + \frac{N_N^2}{s \cdot N} \right) - D(N) \right) \end{aligned} \tag{21}$$

The first-order conditions are now:

$$\frac{\partial \Lambda}{\partial N} = D(N) - \delta \cdot \frac{N_N}{s} - 2 \cdot \delta^* \cdot \frac{N - N_N}{s} + \lambda \cdot \left(\frac{\delta^*}{s} \cdot \left(1 - \frac{N_N^2}{N^2} \right) - D'(N) \right) = 0 \tag{22a}$$

$$\frac{\partial \Lambda}{\partial N_N} = -\delta \cdot \frac{N}{s} + 2 \cdot \delta^* \cdot \frac{N - N_N}{s} + \lambda \cdot \left(\frac{\delta}{s} - 2 \cdot \frac{\delta^*}{s} + 2 \cdot \frac{\delta^* \cdot N_N}{s \cdot N} \right) = 0 \tag{22b}$$

$$\frac{\partial \Lambda}{\partial \lambda} = \delta \cdot \frac{N_N}{s} + \delta^* \cdot \left(\frac{N}{s} - 2 \cdot \frac{N_N}{s} + \frac{N_N^2}{s \cdot N} \right) - D(N) = 0 \tag{22c}$$

Somewhat surprisingly, because the term between large brackets in (22b) is equal to $-1/N$ times the term in front of the λ-term, the ratio of shoulder users can be found directly from (22b), independent of λ:

$$\frac{N_P}{N} = \tfrac{1}{2} \cdot \frac{\delta}{\delta^*} \tag{22d}$$

It is the familiar expression that we already found for perfectly inelastic demand, and for the step toll, and this reflects that the feebate is ineffective at addressing the problem of induced demand. That gives a somewhat different reason than for the step toll to aim for the cost-minimizing ratios of shoulder and central users without considering induced demand effects. The equilibrium expression for λ is tedious for this regime, but again reflects that it will approach zero as demand approaches perfect inelasticity:

$$\lambda = \frac{\delta^* \cdot N \cdot \left(N^2 - N_N^2\right)}{\delta^* \cdot \left(N^2 - N_N^2\right) - D' \cdot s \cdot N^2} \qquad (22e)$$

Using (22d) and (20a) we can compute the following levels for τ and σ:

$$\tau = \tfrac{1}{2} \cdot \delta \cdot \frac{N_P}{s} = \tfrac{1}{4} \cdot \frac{\delta}{s} \cdot \frac{\delta}{\delta^*} \cdot N \qquad (22f)$$

$$\sigma = \tfrac{1}{2} \cdot \delta \cdot \frac{N_N}{s} = \tfrac{1}{4} \cdot \frac{\delta}{s} \cdot \left(\frac{2 \cdot \delta^* - \delta}{\delta^*}\right) \cdot N \qquad (22g)$$

Note that (22f) and (22g) imply that the difference in monetary prices between the shoulder and central periods is consistent with (22d) in the sense that it is again equal to the coarse toll level for inelastic demand:

$$\tau + \sigma = \tfrac{1}{2} \cdot \frac{\delta}{s} \cdot N \qquad (22f)$$

The reader may also verify that substitution of (22d) into (22f) confirms that (13) is satisfied.

Numerical results

It is clear from the above derivations that the main factor determining the relative performance of the four coarse schedules will be the elasticity of demand. We illustrate this in the context of the same numerical model as we used in the previous section. The results are summarized in Figures 5, 6 and 7. The horizontal axes of these figures give the absolute value of the elasticity of demand in the no-toll equilibrium.

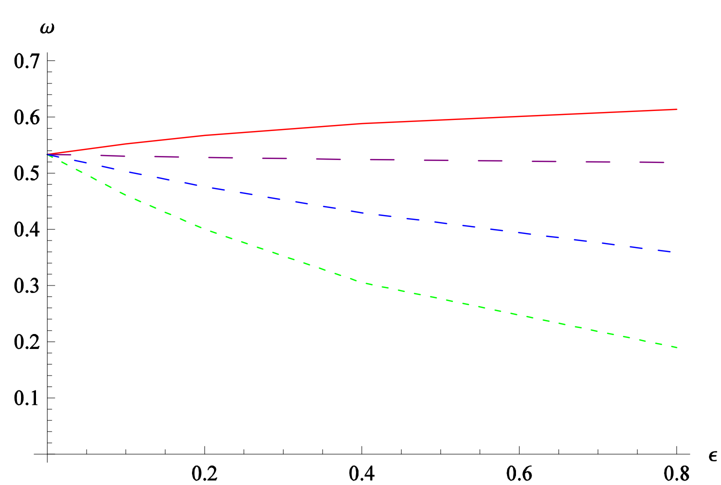

Figure 5. Relative efficiency (ω) of step tolls (red, solid), coarse tolls (purple, large dashing), coarse feebates (blue, medium dashing), and coarse rewards (green, small dashing) by elasticity of demand (ε) in the no-toll equilibrium

Figure 5 shows the relative efficiency, ω. Consistent with our analytical findings, the four coarse regimes are equally efficient when the price elasticity of demand in the original equilibrium is equal to 0. When demand becomes more elastic, the step toll becomes relatively more efficient, the relative efficiency of the coarse toll remains more or less unchanged, but the other two systems – which rely partly or completely on rewarding and therefore suffer from induced demand – become gradually less efficient.

Figure 6 shows the optimal share of shoulder drivers in the four systems, N_P/N. There is a large discrepancy between the systems that rely partly or completely on tolling, and the 'purest' reward system. The optimal share of participants is clearly decreasing in the elasticity of demand for the latter system, whereas it remains rather constant in the other regimes.

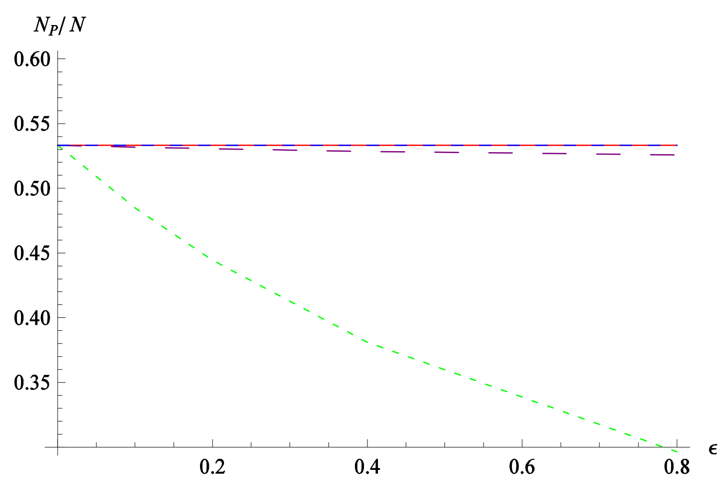

Figure 6. Second-best optimal share of "participants" (shoulder-period users) with step tolls (red, solid), coarse tolls (purple, large dashing), coarse feebates (blue, medium dashing), and coarse rewards (green, small dashing) by elasticity of demand (ε) in the no-toll equilibrium

Finally, Figure 7 shows the optimal values of the tolls and rewards in the four regimes. Note that negative values in the Figure denote positive rewards. In the step toll, the optimal value of both tolls is slightly decreasing in the elasticity of demand. This is consistent with total use declining more strongly due to step tolling as demand becomes more elastic. The optimal reward decreases in magnitude when demand becomes more elastic, confirming our analytical results. The coarse toll, and the toll and reward in the feebate regime, turn out to be rather insensitive to the elasticity of demand.

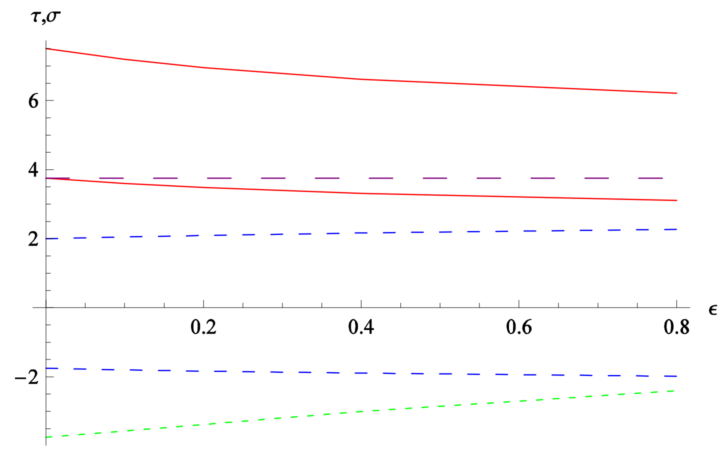

Figure 7. Second-best optimal tolls (τ) and subsidies (σ, in negative quadrant) with step tolls (red, solid; upper curve is central peak and lower curve is shoulder), coarse tolls (purple, large dashing), coarse feebates (blue, medium dashing; upper curve is central peak and lower curve is shoulder), and coarse rewards (green, small dashing) by elasticity of demand (ε) in the no-toll equilibrium

5. Conclusion

This paper analyzed the possibilities to relieve bottleneck congestion by using rewards instead of – or in combination with – taxes. We have shown that with inelastic demand a fine (time-varying) reward is equivalent to a fine toll, and to a continuum of combinations of time-varying tolls and rewards (including fine feebates). When demand is price sensitive, a reward becomes less attractive from the efficiency viewpoint, because it attracts additional users to the congested bottleneck. As a result, both the second-best optimal rate of participation in the scheme, and the relative efficiency that can be achieved with it, was found to decrease when demand becomes more elastic.

We also studied the properties of coarse schemes. Both our analytical and simulation results suggest that a coarse reward is less effective than a coarse feebate, which is itself less effective than a coarse toll. The most efficient coarse system is the step toll, which is also allowed to be positive in the shoulder period. These conclusions reflect that congestion entails an external *cost*, which should ideally be internalized.

Our analysis nevertheless suggests that a reward system can relieve congestion problems by persuading some drivers to pass the bottleneck earlier or later than others, who do not receive a reward. Especially in situations in which tolls are unusually unacceptable –

one may think of road works, bad weather forecasts or large scale sport events (such as the Olympic Games) a reward to stimulate people to avoid the peak may be more attractive to combat congestion than a toll. When demand is expected to be rather elastic, it may be preferable to look for possibilities to impose a feebate system, rather than a reward system.

Many potentially important aspects of a reward system have been left undiscussed in this paper. Heterogeneity among users – possibly in their preferred arrival time, but also in their shadow costs of travel delays and schedule delays – may be particularly important, as may be uncertainty about demand and the capacity of the bottleneck. These issues are left for future work.

References

Arnott, R., A. de Palma en R. Lindsey (1990) "Economics of a bottleneck" *Journal of Urban Economics* **27** 111-130.

Arnott, R., A. de Palma en R. Lindsey (1993) "A structural model of peak-period congestion: A traffic bottleneck with elastic demand" *American Economic Review* **83** 161-179.

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

Chu, X. (1999) "Alternative congestion pricing schedules" *Regional Science and Urban Economics* **29** 697-722.

Daganzo, C.F. (1995) "A Pareto optimum congestion reduction scheme" *Transportation Research B*, **29** 139-154.

Daganzo, C.F. and R.C. Garcia (2000) "A Pareto improving strategy for the time dependent morning commute problem" *Transportation Science*, **34**, 303-311.

De Corla-Souza, P. (2000) "Making pricing of currently free highway lanes acceptable to the public" Transportation Quarterly, **54**, 1-20.

Kalmanje, S. and K. Kockelman (2005) "Credit-based congestion pricing: A proposed policy and the public's response" *Transportation Research* **39A** 671-690.

Knockaert, J., E.T. Verhoef and J. Rouwendal (2009) "The Spitsmijden experiment: reward to battle congestion" TRB 88th Annual Meeting Compendium of Papers DVD.

Laih, C-H. (1994) "Queuing at a bottleneck with single- and multi-step tolls" *Transportation Research* **28A**(3) 197-208.

Small, K.A. (1982) "The scheduling of consumer activities: work trips" *American Economic Review* **72** 467-479.

Small, K.A. and E.T. Verhoef (2007) *The Economics of Urban Transportation* Routledge, London.

Xiao, F., Z. Qian and H.M. Zhang (2009) "The morning commute problem with coarse toll and nonidentical commuters" University of California Davis, working paper.

Verhoef, E.T., P. Nijkamp and P. Rietveld (1997b) "Tradeable permits: their potential in the regulation of road transport externalities" *Environment and Planning B: Planning and Design* **24B** 527-548.

Vickrey, W.S. (1969) "Congestion theory and transport investment" *American Economic Review* **59** 251-261.