The Economic Theory of Urban Traffic Congestion: A Microscopic Research Agenda*

Richard Arnott
Department of Economics
Boston College
Chestnut Hill, MA 02467
1-617-552-3674
richard.arnott@bc.edu

July 2001

* For presentation at CESifo Venice Summer Institute 2001, Workshop on Environmental Economics and the Economics of Congestion: Coping with Externalities, Venice International University, San Servolo, July 18-19, 2001. This paper may be found on-line at http://FMWWW.bc.edu/EC-V/Arnott.fac.html

The Economic Theory of Urban Traffic Congestion:
A Microscopic Research Agenda

1. Introduction

The theme of my talk today is that we — the community of urban transport economists — rely excessively on the canonical *macroscopic* model of urban traffic congestion, which has caused us to place excessive emphasis on congestion pricing as a policy tool to improve the efficiency of urban travel. Urban traffic congestion is the outcome of decisions made on many more margins of choice than even the most sophisticated variants of the canonical models capture. To increase the usefulness of our theory in practical policy application, we need to develop a portfolio of models that capture the omitted margins of choice.

Applied microeconomic theory has been so successful largely because of its *method*, which entails working with simple, conceptually consistent models based on maximizing behavior. This method elucidates basic principles and focuses on essentials, abstracting from distracting detail. Recent policy successes of this approach include the development of a market for SO_2, the expanded use of auctions for resource allocation with small numbers of buyers and sellers, and the application of incentive contracting to public utilities (Laffont and Tirola (1993).

Application of this method in the context of urban traffic congestion has led urban transport economists to advocate congestion pricing — ideally perfect congestion pricing but realistically partial or imperfect congestion pricing. But our pleas for congestion pricing have been singularly unsuccessful. With the arguable exception of Singapore, congestion pricing of urban travel has not been implemented anywhere (Small and Gomez-Ibanez (1998)), and many jurisdictions (e.g. Cambridge, Hong Kong (Borins (1986)), Stockholm (Ahlstrand (1981)),and the Randstaad) have backed down from plans to introduce congestion pricing on even an experimental basis. Why congestion pricing of urban auto travel has been received by policy makers with so little enthusiasm has been much discussed in the literature (Gomez-Ibanez (1992), Jones (1998)). My own view is that we urban transport economists have been deceived by the simplicity of our models into greatly underestimating the costs and practical difficulties of implementing congestion pricing *on urban roads*, and that policy makers have wisely resisted doing so. Be that as it may, the social returns to our labors will, I believe, be greater if we devote more of our efforts towards examining *alternatives to congestion pricing*. And this, I shall argue, will require the development of more *microscopic* models of urban travel, which provide a more detailed description of urban traffic flow and urban travel decisions than is done in the canonical macroscopic model; what I have in mind will hopefully be clarified in the talk.

In section 2, I shall provide a brief sketch of the development of urban transport economic theory, as well as its current state, and then provide a critique of it. Then, in section 3, I shall discuss a selection of research topics, which will together illustrate how I think urban transport economic theory should be re-oriented to make it more useful for policy makers.

2. Current Urban Transport Economic Theory and Its Application

The development of urban transport economic theory has entailed the gradual elaboration of a canonical model. The basic model (Beckmann, McGuire, and Winsten (1956)) examines travel on a point-input, point-output road. Individual drivers are identical, and the only economic decision each driver makes is trip frequency. Congestion is captured by a congestion cost function which relates trip cost to traffic volume and capacity.

INSERT FIG. 1 HERE

Figure 1 gives a diagrammatic representation of the basic model (Walters (1961), Mohring (1976)), with capacity fixed. D is the demand curve; AC relates each driver's trip cost to traffic volume, Q, and is variously referred to as trip cost, average cost, user cost, and marginal private cost; and MSC is the marginal social cost of a trip. In the absence of government intervention, the equilibrium occurs where demand intersects average cost. The optimum occurs where demand intersects marginal social cost. The vertical distance between MSC and AC is the congestion externality cost. The minimal government intervention needed to decentralize the social optimum is the imposition of a congestion toll equal to the congestion externality cost, evaluated at the social optimum, τ^*.

The same model may be described algebraically using either social surplus or social welfare analysis (Mayeres and Proost (1997)). Where p is trip price and w capacity, the demand function is $D(p)$, (short-run) average cost $c(Q, w)$, the social benefit function $B(Q)$, and the capacity construction cost function $K(w)$. The direct, long-run social surplus maximization problem may then be written as

$$\max_{Q,w} \qquad B(Q) - Qc(Q, w) - K(w), \tag{1}$$

which gives the optimality conditions:

$$Q: \qquad B^{'}(Q) - \left(c(Q,w) + Q\frac{\partial c(Q,w)}{\partial Q} \right) = 0 \tag{2a}$$

$$w: \qquad -Q\frac{\partial c}{\partial w} - K^{'} = 0. \tag{2b}$$

Eq.(2a) states that optimal traffic volume is such that the marginal social benefit of a trip equals marginal social cost, which equals short-run average cost plus the congestion externality cost. Eq.(2b) states that optimal road width is such that the marginal social benefit from road expansion, the reduction in travel costs holding traffic volume fixed, equals the marginal construction cost. The indirect social surplus maximization problem, where individuals decide on trip frequency based on trip price and the government decides on the congestion toll, with trip price equalling average cost plus the toll, is

$$\begin{aligned} \max_{Q,p,\tau,w} \qquad & (B(Q) - Qp) + (Q\tau - K(w)) \\ \text{s.t.} \quad & \text{i) } Q = D(p) \\ & \text{ii) } p = \tau + c(Q,w), \end{aligned} \tag{3}$$

which reduces to

$$\max_{p,w} \quad B(D(p)) - D(p)c(D(p),w) - K(w). \tag{4}$$

The corresponding first-order conditions are the same as those of the direct maximization problem.

The no-toll equilibrium may be characterized as the solution to $Q = D(p)$ and $p = c(Q,w)$, or as the solution to the following constrained maximization problem

$$\begin{aligned} \max_{Q,p,\tau,n} \quad & B(Q) - Qp + Q\tau - K(w) \\ \text{s.t. } & \text{i) } Q = D(p) \\ & \text{ii) } p = c(Q,w) \\ & \text{iii) } \tau = 0 \end{aligned}$$

The basic model has been enriched to account for other margins of choice. Early on the model was extended to treat *route choice* and *modal choice*. With respect to route choice, an individual chooses his route on a network, from a given origin to a given destination, so as to minimize trip price — the generalized Wardrop principle. With respect to modal choice, when modes are perfect substitutes in demand, the same principle applies, except that there is *congestion interaction* between buses and cars, but no congestion interaction between cars on different links. When modes are not perfect substitutes in demand, the maximization problem is extended to multiple modes, with D, p, and w M-dimensional vectors, where M is the number of modes. The model was also extended early on to treat user heterogeneity (Strotz (1965)). Individuals from different groups have different benefit and average cost functions. And while not analytically necessary, it is almost always assumed that individuals from different groups enter the congestion cost functions symmetrically.

Under this assumption, in the model extended to treat route choice, modal choice, and user heterogeneity, the full optimum can be decentralized by applying an anonymous toll on each link in the network equal to that link's congestion externality cost. Furthermore, the marginal social benefit of capacity on each link can be computed straightforwardly as the travel cost savings on that link, traffic fixed, without consideration of how travellers switch modes and routes in response to the incremental capacity expansion. These are very important results since they indicate that, in terms of the model, very little information is needed to decentralize the first-best optimum. All that is required is to measure the link congestion externality costs, which requires knowledge of only the link congestion functions and traffic levels. No information is needed on the identity of travellers or on their demand functions. It is therefore easy to understand why economists have pushed so hard for perfect congestion pricing.

The above models have been extended to treat freight traffic. On the assumption that a truck contributes to congestion in the same way as a fixed number of cars, first-best congestion tolling — with trucks paying the car-equivalents toll — continues to decentralize the first-best optimum.

The above models have also been applied to treat a range of second-best problems. Lévy-Lambert (1968), Marchard (1968), Sherman (1971) and Bertrand (1977) examined how other modes should be priced when auto congestion is unpriced or underpriced. Wheaton (1978) and Wilson (1983) considered how optimal road capacity is altered when again auto congestion is unpriced or underpriced. Arnott and Yan (2000) have analyzed simultaneously second-best transit capacity, transit pricing, and road capacity when auto congestion is underpriced. Chia, Tsui, and Whalley (2001) have investigated how much of the efficiency loss from not applying congestion tolls to automobiles can be recovered through a gasoline tax. Verhoef, Emmerink, Nijkamp, and Rietveld (1996) have studied how the value of information to car drivers is modified by not congestion pricing car travel. Several papers (e.g. Braid (1996), Verhoef, Nijkamp, and Rietveld (1996), and Liu and McDonald (1998)) have been written on the proportion of efficiency gains than can be achieved when only a subset of roads can be tolled. Most have come to the pessimistic conclusion that only a small fraction of the gains can be achieved. Small and Yan (1999) and Verhoef and Small (1999), however, argue that this conclusion is too pessimistic, and derives from ignoring user heterogeneity. When heterogeneity and hence self-selection across tolled and untolled links according to the value of time is considered, the efficiency gains from tolling only freeways are considerably magnified.

With two major exclusions — trip timing and land use — such broadly is the current state of the economic theory of urban travel (Lindsey and Verhoef (2000)). The development of the theory has been admirable in many respects. Through elaboration of a canonical model, the theory has moved from a very simple model to models that are increasingly descriptively realistic and incorporate more and more margins of choice. All the model variants meet the standard criteria for good microeconomic modelling. They are thoroughly based on individual maximizing behavior, are conceptually consistent, and are parsimonious. And considerable effort has gone into practical application. There are now large literatures on estimating travel demand functions and on developing efficient algorithms to solve variants of the static network equilibrium problem, including the computation of second-best optimal tolls (e.g. Verhoef (2001) and Hearn ( )), and there is a growing number of city-specific travel simulation models based on the above theory (e.g. Anderson and Mohring (1996)).

These admirable qualities notwithstanding, I have six major criticisms of the current state of the theory.

1. <u>Many relevant margins of choice are ignored</u>

Applied microeconomic has thrived largely by abstracting from the inessential. In many policy contexts, most of the action can be captured by considering only a small number of margins of choice. For example: i) with industrial pollution, most of the action is captured by modeling firms' choices concerning the level of output and the technology, as characterized by the level of emissions of a few pollutants per unit output; ii) with insurance, most of the action on the consumer side can be captured by viewing the consumer as choosing how much insurance to purchase and how much unobservable (observable margins of choice can be written into the contract) effort (which affects the probability of accident or more generally the probability distribution of accident damage) to expend, iii) with housing, most of the action can be captured by viewing the consumer as choosing location, floor area, and quality, and the producer as choosing structural density and quality.

The most sophisticated models of urban travel (e.g. dePalma's METROPOLIS) treat the traveller as choosing trip frequency, route, mode, and departure time. But urban car drivers make many more decisions than this. As I am driving along, I continually decide how rapidly to accelerate or decelerate, which determines my speed and the distance between my car and the car in front. Periodically, I have to choose whether to accept an opportunity to overtake, whether to honk my horn, whether to enter an intersection after the light has turned yellow (or in Boston red) or when it is blocked, and whether to shift to an apparently faster lane. As I am approaching my destination, if I do not have employer-provided parking, I have to decide whether to park on-street or off, and if on-street what parking strategy to adopt, which includes how far from my destination to start cruising for parking and under what circumstances to double park.

One possible reaction to this enumeration of choices is that they are trivial. Each choice by itself may be trivial but cumulatively they are very important. Think how much better traffic would flow and how less stressful urban driving would be if all drivers were to make socially efficient decisions. Another possible reaction is that economists have little useful to say concerning these decisions, even though they are economic, and that regulation of driver behavior should be left to traffic engineers. But traffic engineers decide on traffic regulations with no explicit economic behavioral analysis, and often on the basis of insufficient data and flawed statistical analysis.

If indeed there were first-best congestion pricing, we would not have to worry that our analysis overlooks some microscopic margins of choice. Drivers would face the right prices on every margin, and would therefore make socially efficient decisions on every margin. But in practice congestion pricing cannot be differentiated according to driver behavior. Under anonymous congestion pricing, aggressive and timid drivers impose a larger congestion cost externality than socially responsible drivers but pay the same toll. Less obviously but as importantly, link congestion pricing, as computed, provides no incentive to drive in a socially responsible way. A rational driver who faces the same link toll independently of how he drives will drive selfishingly. Thus, even with homogenous drivers, what we compute as "first-best" link congestion pricing takes as given inefficient driver behavior, which renders computation of optimal tolling that is not based on individual driver behavior an exercise in the theory of the second best. Since, in the world of the second best, all margins of choice should be explicitly accounted for, in even our so-called first-best theory ignoring driver behavior may result in seriously misleading analysis.

In the above discussion, I have focused on *individual* margins of choice that the conventional analysis ignores. Policies associated with these margins of choice, most notably traffic regulations, are correspondingly ignored. Also ignored are car manufacturers' margins of choice. Automobile characteristics affect the congestion caused by a car. Since these characteristics are at least partially observable, congestion pricing can be based on them, but if it is not consumers have no incentive to purchase "congestion-efficient" cars or car makers to manufacture them.

2. <u>The congestion function captures not only technology but also behavior</u>

The congestion function is treated as being a technological datum, but in fact incorporates many margins of choice. This point has been demonstrated formally with respect to users' trip-timing decisions, but applies to many other margins of choice. A major development in urban transport economic theory, which was pioneered by Vickrey (1969), has been the extension of the conventional static theory to treat the dynamics or evolution of congestion over the rush hour. This has been done by combining a particularly simple state variable characterization of nonstationary traffic flow — queuing behind a bottleneck — with a behavioral equilibrium condition analogous to the Wardrop condition, that each driver chooses his departure time to minimize his trip price. Equilibrium aggregate travel costs (which include travel time costs and schedule delay costs — the costs of being early or late relative to desired arrival time) may then be computed as a function of the number of drivers of various types and of the tolling "régime". This aggregate travel cost function may then be treated as the total congestion cost function in a static, reduced-form social surplus or social welfare maximization problem. Thus, the conventional static model may be interpreted as incorporating traffic dynamics and users' trip-timing decisions implicitly via the total congestion cost function. According to this interpretation, the total congestion cost function is not completely technological but incorporates users trip timing decisions and is influenced by the tolling régime in effect.

The same point applies with respect to other margins of choice not explicitly treated in the conventional static model. The total congestion function can be interpreted as treating implicitly these other margins of choice. Hence, the conventional model is more generally applicable than one might have suspected. At the same time, treating margins of choice implicitly carries with it twin dangers: first, policy instruments that affect only the implicit margins of choice — such as traffic regulations — may be overlooked; and second, when a policy instrument affects both explicit and implicit margins, it is easy to forget the impact of the policy instrument on the implicit margins — for example, raising the toll may not only lower demand but also cause motorists to drive faster thereby altering the form of the "congestion cost function".

The congestion cost function is sometimes interpreted as incorporating the costs of traffic noise, pollution, and accidents. That is conceptually acceptable, but carries with it the danger that the policy analyst will neglect policy instruments affecting corresponding margins of choice.

A related point is that treating the value of time — which enters the congestion cost function — as a datum makes it easy to overlook policy instruments that affect the value of time. Traffic regulations which discourage aggressive driving make driving more pleasant. This reduces the value of time and, holding traffic volume fixed, reduces congestions costs, but also causes demand to rise, increasing the level of congestion.

3. Capacity is too aggregated a policy variable

Transport planners do not choose capacity *per se.* Instead, they choose road width, pavement quality, gradient, banking, ramp metering, speed limits, and so on, which together determine capacity. Since transportation planners tend to use engineering rules of thumb, without reference to economic variables, their choice of how to provide a given level of capacity may differ significantly from the design which minimize social costs. Economists have a role in advising transportation planners how to provide a given level of capacity efficiently in different economic environments. To do this, a set of transport economic models is needed that provide a richer treatment of traffic engineering. Two examples of excellent work along these lines are Newbery (1988) and Small, Winston and Evans (1989). Newbery examines the economics of pavement resurfacing, Small,Winston and Evans looks at the economics of road damage, considering not only how vehicles should be charged for the road damage they cause but also how pavement durability should be chosen.

Economists can contribute to the traffic engineering/transportation science literature as well by developing microscopic models of traffic flow with behavioral foundations. Some work has already been done along these lines. Rotemberg (1985) and Verhoef, Rouwendal, and Rietveld (1999) provide models in which drivers decide on speed, spacing, or acceleration so as to maximize utility, trading off travel time against the probability of accident. And Mohring ( ) derives utility-maximizing speed as a tradeoff between travel time and gasoline consumption. Further work along these lines should incorporate recent developments in microscopic traffic flow theory (Transportation Research Board), including car following theory which derives aggregate traffic flow from a difference-differential equation describing the acceleration of individual vehicles.

While Vickrey is best known among transport economists as a crusader for congestion pricing and as the developer of the bottleneck model, he also pioneered in the engineering economics of congestion. He regularly attended the Transportation Research Board annual meetings, as well as international conferences in traffic engineering; he developed several models of congestion other than the link flow and bottleneck models; and he also considered efficient subway scheduling, seating, braking, fare collection, and platform length, and devised schemes to mitigate bus bunching.

4. Link flow congestion is not the only form of congestion

The conventional model treats only one form of congestion, link flow congestion whereby a driver's costs on a link are positively related to traffic volume or flow on the link. As has already been noted, the conventional static link congestion cost function can also be interpreted as providing a reduced form representation of bottleneck congestion. But there are many traffic congestion phenomena that are not consistent with link flow congestion. Link flow congestion excludes transient, non-steady-state flow phenomena such as shock waves and hypercongestion. It also ignores congestion at nodes. In telephone traffic, congestion at nodes (in switching circuits) is more important than congestion on links (Syski (1986)). Examples of nodal congestion in the context of urban travel are intersection[1] , freeway entrance and exit, and parking congestion. Link congestion dominates nodal congestion in freeway travel, but not on city streets. Other forms of congestion include pedestrian-car interaction, entry into and exit from parking, merging, and phenomena deriving from the physical length of cars such as gridlock.

Of these, perhaps the most important is parking congestion. Remarkably little is known empirically about parking, but there is some evidence that in car travel with a downtown destination, the average time devoted to finding a parking space may be as large as the average time lost to congestion en route. Cars cruising for parking also contribute significantly to traffic congestion. One encounters the assertion in the literature that in congested downtown areas, half the cars driving are cruising for parking. This seems too high, but cruising for parking no doubt contributes significantly to congestion. On-street parking also reduces capacity, and double parking and entry into and exit from on-street

[1] In applications of traffic network equilibrium theory, which incorporate only link flow congestion, intersections are treated as *virtual links*. For example, a northbound car which makes a right-hand turn is viewed as traveling along a link joining the N-S road to the E-W road. This treatment is better than nothing, but not entirely satisfactory since it ignores the interaction between traffic travelling in different directions.

parking can seriously impede traffic flow.

To derive efficient urban transport systems we shall need to develop richer and more microscopic models of congestion.

5. Interaction between urban travel distortions and other distortions in the economy may be important

The conventional modeling of urban travel ignores the interactions between urban travel distortions and other distortions in the economy. Two of these are probably particularly important. The first, the analog of which has been discussed at length in the environmental economics literature (Bovenberg and Goulder (1996, 1998), and Parry and Bento (2000)), concerns the interaction between the labor-leisure distortion caused by the taxation of wage income and urban travel. The second, which to my knowledge has not been mentioned before, concerns the connection between *interaction externalities* and urban travel. In recent years, there has been considerable research on the *economics of agglomeration* (see Fujita and Thisse (2000) for an excellent survey). Most experts consider that the non-market exchange of information through face-to-face interaction is a primary, and perhaps the primary, force encouraging firms to cluster. If this is correct, the associated positive interaction externality must be of the same order of magnitude as the externality associated with unpriced urban auto congestion. To mitigate the deadweight loss associated the interaction externality, interaction should be encouraged, and perhaps *subsidizing* urban travel is a way to deal with this. We know little empirically about the interaction between the labor-leisure distortion and that arising from unpriced auto congestion, and even less empirically about the interaction between the interaction externality and the urban auto congestion externality.

What implications these interactions between distortions have for urban travel policy is not at all clear. One possibility, however, is that congestion pricing will become less attractive and microscopic policies aimed at improving the efficiency with which a given pattern of traffic is accommodated more attractive.

6. The demand for travel is predominantly a derived demand

This is a familiar criticism. The conventional model treats individuals as deriving utility from travel *per se.* But individuals derive utility from activities arrayed over time and space, as well as from goods and services. These activities require transportation, as well as other goods and services as inputs. Thus, the demand for travel is predominantly a derived demand. While almost everyone acknowledges the correctness of this criticism, little progress has been made in developing activity-based models of derived travel demand. The associated scheduling problems are very difficult, and no progress has been made in the solution of scheduling problems which require schedule coordination between individuals. Thus, it seems that we are stuck with treating travel as a final good. This is unfortunate since it muddies application of production efficiency arguments from optimal tax theory in the context of urban travel.

In conclusion: Urban transport economic theory has developed primarily through the elaboration and refinement of a single, canonical model. In many respects, the process has been admirable and fruitful. Since, however, we are so familiar with the model, we are apt to forget that it is only one of many possible simplified representations of very complex and varied traffic congestion phenomena. By drawing so heavily on this single model as the conceptual basis for our policy analysis, we have probably placed excessive emphasis on congestion pricing while ignoring other potentially valuable policy tools. A promising avenue towards redressing this imbalance is to develop more microscopic economic models of urban travel.

3. A Selection of Research Topics

In the remainder of the talk, I shall discuss several research topics which will I hope serve to illustrate the points I have made.

1. Regulation of freight deliveries — time of day and truck size

In downtown Boston at least, freight delivery contributes considerably to traffic congestion. Large interstate (designed for freeway travel) trucks have trouble manoeuvering round corners or narrow streets; they completely block traffic on streets when entering and exiting from loading docks; and their double-parking for deliveries where loading docks are absent severely reduces capacity.

I know of no modern economic study of urban freight delivery in the context of urban travel congestion. We seem simply to treat trucks as so many car-equivalents, and assume that whatever policy is best for cars is best for trucks. By employing such a crude treatment of urban freight transport, we overlook many policies that might significantly reduce the unpriced congestion externality imposed by trucks.

One such policy is imposing restrictions on the times of day at which downtown freight deliveries can be made. Such a policy is, I understand, on the books in Paris, though I have no idea how strictly it is enforced. Evaluating such a policy requires estimating the costs, which include inconvenience costs of restricted delivery hours to shippers and receivers, as well as determining more precisely the "technology" of the congestion interaction between cars and trucks. As has been noted, it is standard to treat trucks as car-equivalents, which assumes that cars and trucks enter the congestion function additively. Marvin Kraus, however, argues that trucks enter the congestion function as reductions in capacity, which accords better with my intuition. Let A denote the number of cars, T the number of trucks, and k the amount by which a truck reduces capacity. Then the link travel-time function may be written as $t\left(\frac{A}{w-kT}\right)$. Let e denote the car equivalents of a truck in terms of congestion

$$e(A,T,w) = \frac{\partial t(\cdot)/\partial T}{\partial t(\cdot)/\partial A} = \frac{Ak}{w-kT}, \tag{5a}$$

so that

$$\frac{\partial e}{\partial T} = \frac{Ak^2}{(w-kT)^2} > 0; \tag{5b}$$

that is, the car equivalents of a truck is increasing in the number of trucks. Thus, in the absence of congestion pricing, correcectness of the Kraus conjecture would strengthen the argument for smoothing truck traffic over the course of the day, and probably too for restricting delivery hours.

Another freight-delivery-related policy is regulating truck size in areas of severe congestion. Evaluating this policy requires knowledge of warehousing technology and practice. Suppose, at one extreme, that current practice is for all goods transported by inter-city truck to be unloaded at suburban warehouses and reloaded onto smaller trucks for delivery within the metropolitan area. In this case, relatively little additional cost would be imposed by restricting truck size for urban deliveries. Suppose, at the other extreme, that current practice is for all goods to be transported from supplier to receiver door-to-door. Regulating truck size for urban delivery would then be considerably costlier. If door-to-door deliveries were to continue, the smaller truck size would substantially increase inter-city shipping costs. Otherwise, new warehousing districts would have to be constructed where goods shipped inter-city in large trucks would be unloaded and loaded onto smaller trucks for urban delivery.

Geho (2000) is currently writing her thesis on urban freight consolidation, and Sivitanidou before her untimely death was studying the spatial economics of urban warehousing. More research along these lines is badly needed, but will be hampered by deficiencies in data.

More generally, study of the contribution to congestion caused by urban freight delivery should be high on the urban transport economic research agenda.

2. <u>The engineering economics of urban auto congestion</u>

I have already mentioned the exemplary work by Newbery (1988) and Small,Winston, and Evans (1989) related to the economics of freeway/highway design, construction, and maintenance. Their work corresponds to sound cost-benefit practice. If sound cost-benefit analysis were practiced in all aspects of urban road engineering, very considerable cost savings could be achieved, especially if second-best considerations were properly accounted for (Kanemoto (1999)). Though the work might be rather unglamorous and conceptually prosaic, we can make valuable contributions by demonstrating in practical applications how cost-benefit analysis should be done and by pressing hard for the adoption of sound cost-benefit procedures by state and local governments. We can also contribute by applying economics to the nuts and bolts of road design. For example, even the most sophisticated cost-benefit procedures currently employed provide very crude treatments of uncertainty, taking no account of the literature on irreversible investment and real options. Rules for road resurfacing should take into account that future traffic volumes are generated by a stochastic process, that our understanding of pavement damage will improve, and that technological advances in pavement design will occur, and should accordingly be more adaptive and flexible.

Ezra Hauer ( ) has argued that the bulk of transport engineering standards and rules are based on scant empirical testing and on-site data collection, and often faulty statistical analysis. To this I would add that engineering standards tend to be applied without reference to economic variables — the discount rate, the value of time, etc. Economic studies which devise guidelines for such standards would be valuable.

To do transport engineering economics well requires some expertise in transport engineering. At present there is little professional communication between traffic engineers/transportation scientists and transport economists. The situation is improving, however. Economics is now taught in many graduate transportation engineering programs, and the transportation science literature is drawing increasingly on economics, particularly with regards to pricing solutions. And at least a subset of urban transport economists keep abreast of the relevant engineering literature (e.g. *Transportation Science* and *Transportation Research).* But we can do better. We should not only encourage transportation engineering programs to teach more and better economics, perhaps volunteering to teach a course ourselves, participating in thesis supervision of transportation engineering students, and attending transportation engineering seminars, but we should also include more transportation engineering in our teaching and in our research.

3. Automobile noise, traffic accidents, and urban auto pollution

The automobile congestion cost function is often interpreted as incorporating the costs associated with the noise, accident, and pollution caused by cars. Treating the congestion cost function as technologically determined therefore ignores all the behavioral and engineering decisions that influence how traffic congestion affects the levels of noise, accidents, and pollution.

There has been considerable research into automobile pollution and alternative policies for reducing automobile emissions (e.g., Small and Kazimi (1995), Brownstone ( )), and policy in this area has been rather successful. We need a comparable body of work on the noise pollution generated by cars and on the economics of traffic accidents.

A German transport economist recently informed me that Germans are more concerned by the noise generated by traffic than by the time delays due to traffic congestion. While I doubt this, it is nonetheless true that traffic noise in European cities tends to be significantly higher than in U.S. cities (where population is less dense, streets wider, and building setbacks higher) and can be very irritating. Urban transport economists have paid little attention to traffic noise. It is, however, an area of policy where effective policy remedies would be popular and relatively easy to implement. What needs study is the cost-benefit calculus of alternative policies. It would not be difficult to design quieter cars, but at present car makers have little incentive to do so because each driver incurs only a fraction of the noise cost generated by her car; the same point applies to trucks and buses, which contribute disproportionately to traffic noise. Horn-honking, a curse of living in Boston, could be dealt with by making it a traffic violation except when done to avoid an accident, as I gather is done in many European cities. Road work and garbage collection could be made quieter.

Traffic accidents are costly not only for the direct damage they cause but also for the non-recurrent congestion (including "curiosity congestion") they induce. Economists have paid some attention to traffic accidents (e.g., Vickrey (1968)), especially to the effect of insurance on the incentive to drive safely (e.g., Boyer and Dionne (1987)) but have left other aspects such as the regulation of unsafe driving, the design of roads for safety, and accident follow-up procedures, for traffic engineers who, in this context too, tend to choose policy with little or no explicit attention to economics. Virtually no attention has been paid by either economists or engineers to the link between traffic accidents and the "technology" of congestion. The primary tradeoff determining the many small decisions drivers make is between reducing travel time and increasing accident risk. Thus, policies which affect the private costs of accidents may have a significant effect on how traffic flows.

4. Uncivil driving behavior and the value of time

The standard model of auto congestion treats an individual's value of time as exogenous (e.g. Calfree and Winston (1998)), but it isn't. It depends on the scheduling constraints she confronts and on how pleasant or unpleasant she finds driving. A simple way to reduce the cost of traffic congestion and the congestion externality cost is to make driving more pleasant, thereby reducing the value of time (though doing so also lowers trip price which stimulates demand).

The market takes care of automobile comfort, but not the equilibrium stress level associated with driving. I conjecture that driving stress is strongly related to the incidence of uncivil and dangerous driving: tailgating, honking at the slightest provocation, running yellow and red lights, making dangerous and excessive lane changes. It is unclear how effective public policy can be in discouraging anti-social driving[2], but the topic is worthy

[2]Only half in jest, I would like to see bad drivers ostracized from the road. The difficulty lies in designing a technology which allows a driver to report another without interfering with his driving concentration.

an enquiry by economists.

5. Economics of mass transit and pedestrian traffic

The bulk of the work done on the economics of urban traffic congestion has concerned cars. Relatively little has been done on the economics of mass transit, presumably because until recently most of the innovative research in the field was done in the United States where mass transit is relatively unimportant. Mohring (1972) explored some of the basic economic principles of mass transit, in particular economies of service frequency and service density; and there is a substantial literature which estimates mass transit cost functions (e.g. Berechman (1993)) and a smaller one which examines capital-intensity bias (e.g. Frankena (1987)). But urban transport economists have devoted little attention to the microscopics of urban mass transit. The major exception is Vickrey. As noted earlier, he did considerable work in the area but most remains unpublished and much is particular to New York City. We would do well to follow his lead. Sample topics include second-best train/bus size and service frequency and density, procedures to mitigate bus bunching (Vickrey (1979)), to expedite passenger entry and exit, and to reduce the congestion imposed by buses.

The economics of pedestrian congestion is unexplored. The proposal to impose a minimum walking speed on one side of Oxford Street was met with considerable amusement, but merits serious analysis[3]. Pedestrian-car congestion interaction is potentially important. Should jaywalking be discouraged? How wide should sidewalks be?

6. Hypercongestion

Hypercongestion is the phenomenon whereby a given flow occurs at a significantly lower speed than is possible. Hypercongestion has puzzled and intrigued transport engineers and economists. Reducing its incidence would provide substantial efficiency gains. How can this be achieved? The current wisdom (Hall (), Small and Chu (), Verhoef ()) is that

[3] Apparently pedestrian traffic too is characterized by hypercongestion.

hypercongestion is a transient phenomenon generated by events (such as a slow car in the passing lane or a double-parked car) which trigger backward shockwaves. Accordingly, such events should be penalized. How specifically this should be achieved is an exercise which blends traffic flow theory and economics at the microscopic level.

7. Flextime and staggered working hours

A generation ago urban transport economists (with the exception of Vickrey) essentially ignored the dynamics of rush hour traffic congestion. Because of excessive focus on the conventional models, they overlooked the trip timing decision. That oversight has now largely been rectified in the now substantial literature on bottleneck congestion, initiated by Vickrey (1969). That literature has strengthened the case for congestion pricing by demonstrating that time-varying congestion pricing can substantially reduce congestion, holding fixed the time pattern of arrivals (Arnott, dePalma, and Lindsey (1993)).

The bottleneck literature takes the distribution of work start times as exogenous. The endogeneity of the distribution of work start times has been studied by Henderson (1981). In his model, each employer decides when to have his employees start work, trading off the benefit from having his employees interact with more employees from other firms against the higher wage that he must pay his workers for commuting in congested conditions. There is then an interaction between interaction and traffic congestion externalities. Is it worthwhile for the government to attempt to modify private firms' work start times? Whether it is or not, can the government, as the dominant employer in many jurisdictions, significantly improve efficiency by modifying the work start times of its employees (Bonsall (1978))?

8. Non-commuting trips

It has been argued that late twentieth-century urban economics was preoccupied with refining a model of the nineteenth-century city. Urban transport economics can be subjected to a similar criticism, that it is refining models of commuting traffic at a time when an increasing proportion of rush-hour travel has a non-commuting purpose. Fifteen years ago, the figure was bandied about that less than fifty percent of rush-hour trips are for commuting, and recently the figure for Chicago has fallen to thirty percent. One must take such figures with a grain of salt, since they are sensitive to the purpose assigned to chained trips. But, however measured, a steadily increasing proportion of urban travel has a non-commuting purpose.

The difficulty with treating non-commuting trips within a dynamic model of congestion is that desired arrival time and correspondingly schedule delay costs become fuzzy. Rather, an individual schedules his activities taking traffic congestion into account. Increased congestion will cause him to cancel some activities and reschedule others. Unfortunately, it has proved very difficult to operationalize this conceptualization. Until we do, the demand side of dynamic urban travel models will remain disconcertingly weak.

9. Parking

It was remarked earlier that the focus of our theory on link flow traffic congestion has distracted us from other forms of congestion and associated policy tools. Particularly important is parking-related congestion. At least for auto travel with a downtown destination, the average time lost in searching for a parking spot may be as large as the average time lost due to congested traffic, and cruising for on-street parking probably contributes significantly to downtown traffic congestion.

To date, most of the work on parking by urban transport economists (Vickrey (1954), Roth (1965), Gillen (1977,1978), Shoup (1982, 1987),Shoup and Willson (1992), Glazer and Niskanen (1992), Verhoef, Nijkamp, and Rietveld (1995), Calthrop, Proost, and van Dender(2000)) regard parking as a price, either fixed or per unit time, payable at the destination for a trip by car. In modeling modal choice, and the effects of cashing out employer-provided parking, this is a convenient simplification. However, it also misses a lot. It ignores the contribution of parking to traffic congestion — cruising for parking, capacity reduction from on-street parking, and congestion due to entry into and exit from on-street parking, double parking, and queues at entries to off-street parking. It also ignores the congestion cost parkers impose on one another, in terms of the time to find a parking space and to walk from a parking spot to the destination, as well as the attendant uncertainty. These phenomena can be crudely captured by reduced-form models; for example, one might model a driver's parking time costs as increasing in the ratio of the number of auto commuters already parked to capacity. But this approach ignores the microscopics of parking-related traffic congestion, the stochasticity inherent in the parking search process, the spatial aspects of searching for parking, and the allocation of land to parking. A more satisfactory approach would be both spatial and structural, modelling drivers' spatial search for parking.[4] This would permit the analysis of parking policies at an appropriately micro level: the allocation of land to on- and off-street parking, the spacing between parking garages, private vs public ownership of parking garages, second-best on- and off-street parking fee structures, the allocation of on-street parking between metered, resident, and unrestricted parking, time limits for on-street parking, and so on.

Traffic engineers have devoted surprisingly little attention to parking-related congestion. The dearth of stylized facts will impede modeling, and may require that we undertake the data collection and analysis ourselves.

---

[4] Axhausen (1990), Axhausen and Polak (1995), and Arnott and Rowse (1999) provides first steps in this direction.

4. Conclusion

We — the community of urban transport economists — have been less effective than we could have been in advising policy makers, by focusing excessively on congestion pricing. I am certainly not opposed to congestion pricing, but think that we have considerably underestimated the costs of implementation, and may have overestimated the benefits as well. Be that as it may, there seems to be almost overwhelming opposition to congestion pricing. Its time may yet come, but for the moment we should be pragmatic and devote more of our efforts to considering more immediate, practical, and mundane policy issues.

In my talk today, I have argued that our preoccupation with congestion pricing has stemmed from excessive reliance on a single model framework — which I have termed the *canonical macroscopic link flow model* of urban traffic congestion. Elaboration of this model has given our field an impressive, consistent, and coherent body of the theory. But looking at urban transport congestion through a single lens has distorted our perception, causing us to neglect many real-world aspects of urban traffic congestion and consequently to overlook many promising avenues of policy-related research. Most of my criticisms concerned the *link congestion cost function*. By treating congestion at such an aggregate level, it has caused us to look at policy at too aggregate a level, leaving microscopic policy to engineers whose decisions are typically ill-informed by economics. By treating congestion as purely technological, it has caused us to overlook many individual margins of choice than can be influenced by policy. And by treating only link flow congestion, it has caused us to overlook other forms of traffic congestion and the associated policy variables.

I also put forward a research agenda. The topics chosen were not intended to be exhaustive, but rather to illustrate the type of research that would address my criticisms of our existing body of theory. By and large, the research agenda is *microscopic* in nature, and complements rather than competes with the macroscopic theory which dominates our field. The topic headings give a flavor of the type of research I have in mind:

- regulation of freight delivery — time of day and truck size
- the engineering economics of urban auto congestion
- automobile noise, traffic accidents, and urban auto pollution
- uncivil driving behavior and the value of time
- economics of mass transit and pedestrian traffic
- hypercongestion
- flextime and staggered work hours
- non-commuting trips
- parking

The research agenda I put forward was frustratingly — even annoyingly — non-specific. But I am normally not someone who criticizes without offering constructive suggestion, and I hope than in the years ahead that my own research will go some way to meeting the challenges I have posed.

References

[1] Ahlstrand, I. (1998). "The rise and fall of the heroic transport plan for Stockholm", Transport Policy 5:205-211

[2] Anderson, D. and H.Mohring (1996). Congestion Costs and Congestion Pricing for the Twin Cities, Report No. MN/RC-96/32 to Minnesota Department of Transportation (Department of Economics, University of Minnesota, Minneapolis/St. Paul)

[3] Arnott, R.(1994)."Alleviating traffic congestion: alternatives to road pricing", working paper No.282, Department of Economics, Boston College

[4] Arnott, R., A. de Palma and R.Lindsey (1993). "A structural model of peak-period congestion: a traffic bottleneck with elastic demand", American Economic Review 83: 161-79

[5] Arnott R., and J. Rowse (1999). "Modeling Parking", Journal of Urban Economics 45: 97-124.

[6] Arnott, R. and A. Yan (2000). "The two-mode problem: second-best pricing and capacity", Review of Urban and Regional Development Studies 12:170-199

[7] Axhausen, K.W. (1990). "A simultaneous simulation of activity chains", in P.M.Jones (Ed), New Approaches in Dynamic and Activity-based Approaches to Travel Analysis, 206-225, Avebury, Aldershot.

[8] Axhausen, K.W. and J.W.Polak (1995)."A disaggreagte model of the effects of parking guidance systems", 7th WCTR Proceedings

[9] Beckmann, M., C.B. McGuire and C.B. Winsten (1956). Studies in the Economics of Transportation, Yale University Press, New Haven.

[10] Berechman, J. (1993). Public Transit Economics and Deregulation Policy (North-Holland, Amsterdam)

[11] Bertrand, T.(1977). "Second-best congestion taxes in transportation systems", Econometrica 45: 1703-15

[12] Bonsall, T.(1978). "Flexible work hours and public transit in Ottawa", presented at the annual conference of the Roads and Transportation Association of Canada, Toronto, Ontario, September 23-4

[13] Borins, S.F.(1986). "The political economy of road pricing: the case of Hong Kong", Proceedings of the World Conference on Transport Research, Vancouver, B.C., Vol.2, 1367-78

[14] Bovenberg, A.L. and Goulder, L.H.(1996). "Optimal environmental taxation in the presence of other taxes: general equilibrium analysis", American Economic Review 86: 985-1000

[15] Bovenberg, A.L. and Goulder, L.H.(1998). "Environmental Taxation", in A.Auerbach and M.Feldstein, eds, Handbook of Public Economic, 2nd edition, North-Holland, New York

[16] Boyer, M. and G.Dionne (1987). "The economics of road safety", Transportation Research 21B:413-431

[17] Braid, R.(1996). "Peak load pricing of a transportation route with an unpriced substitute", Journal of Urban Economics 40: 179-97

[18] Brownstone

[19] Calfree, J.E. and C.Winston (1993). "The value of automobile travel time: implications for congestion policy", Journal of Public Economics 69: 83-102.

[20] Calthrop, E., S. Proost, and K. van Dender (2000). “Parking policies and road pricing”, Urban Studies 37: 63-76.

[21] Chia, N.-C., A.Tsui, and J.Whalley (2001). “Ownership and use taxes as congestion correcting instruments”, NBER working paper 8278, Cambridge, MA

[22] Frankena, M. (1987).“Capital-biased subsidies, bureaucratic monitoring, and bus scrapping”, Journal of Urban Economics 21: 180-93

[23] Fujita, M. and J.-F.Thisse (2000). “The formation of economic agglomerations:old problems and new perspectives”, in J.-M.Huriot and J.-F.Thisse, eds, Economics of Cities:Theoretical Perspectives, Cambridge University Press,Cambridge UK

[24] Geho (2000). Ph.D. thesis proposal, University of California, Irvine.

[25] Gillen, D.W.(1977). “Estimation and specification of the effect of parking costs on urban transport mode choice”, Journal of Urban Economics 4: 186-99

[26] Gillen, D.W.(1978). “Parking location decisions and the distribution of congestion”, Transportation 7: 69-85

[27] Glazer,A. and Niskanen, E.(1992). “Parking fees and congestion”, Regional Science and Urban Economics 22: 123-32

[28] Gómez-Ibáñez, J.(1992).“The political economy of highway tolls and congestion pricing”, Transportation Quarterly 46:343-360

[29] Gómez-Ibáñez, J. and K.Small (1994). Road Pricing for Congestion Management: A Survey of International Practice, National Cooperative Highway Research Program, Synthesis of Highway Practice 210 (National Academy Press, Washington, DC).

[30] Hall, F.

[31] Hauer, E.

[32] Henderson, J.V.(1981)."The economics of staggered work hours", Journal of Urban Economics 9: 349-64

[33] Institute of Transportation Engineers (1982). Transportation and Traffic Engineering Handbook, Prentice-Hall, London

[34] Jones, P.(1998). "Urban road pricing: public acceptability and barriers to implementation", in K.J. Button and E.T. Verhoef, eds, Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility , 263-84, Edward Elgar: Cheltenham, UK

[35] Kanemoto, Y.(1999)."Benefit -cost evaluation of transportation investment with underpriced congestion", University of Tokyo, mimeo.

[36] Kraus, M. and Y.Yoshida (2000)."The consumer's time-of-use decision and optimal pricing and service in urban mass transit", Boston College, mimeo, forthcoming in Journal of Urban Economics

[37] Laffront, J.-J. and J.Tirole.(1993). A Theory of Incentives in Procurement and Regulation, MIT Press, Cambridge, MA

[38] Lévy-Lambert, H.(1968),"Tarification des services à qualité variable– application aux péages de circulation", Econometrica 36: 564-74

[39] Lindsey, R. and E.T. Verhoef (2000). "Congestion modelling ", forthcoming, in Handbook of Transport Modelling, Vo l.1, D.A. Hensher and K.J. Button, eds, Elsevier Sciences, Oxford

[40] Lindsey, R. and E. Verhoef (2001). "Traffic congestion and congestion pricing", mimeo

[41] Liu, L, and J.F.McDonald.(1998).“Efficient congestion tolls in the presence of unpriced congestion: a peak and off-peak simulation model”, Journal of Urban Economics 44:352-66

[42] Marchand, M. (1968).“A note on optimal tolls in an imperfect environment”, Econometrica 36: 575-81

[43] Mayeres, I. and S.Proost (1997).“Optimal tax and public investment rules for congestion type of externalities”, Scandinavian Journal of Economics 99: 261-279

[44] Mohring, H. (1972).“Optimization and scale economies in urban bus transportation”, American Economic Review 62: 591-604

[45] Mohring, H.

[46] Mohring, H. (1976). Transportation Economics, Ballinger, Cambridge MA

[47] Newbery,D. (1988). “Road damage externalities and road user charges”, Econometrica 56: 295-316.

[48] Parry, I. and A.Bento (2000). “Tax deductions, environmental policy, and the‘double dividend’ hypothesis”, forthcoming, Scandinavian Journal of Economics

[49] Rotemberg, J. (1985). “The efficiency of equilibrium traffic flows”, Journal of Public Economics 26:191-205

[50] Roth, G.J.(1965).Paying For Parking, The Institute of Economic Affairs, London

[51] Sherman, R. (1971). “Congestion interdependence and urban transit fares ”, Econometrica 39: 565-76

[52] Shoup, D. (1982). “Cashing out free parking”, Transportation Quarterly 35: 351-364

[53] Shoup, D. (1997).“Evaluating the effects of cashing out employer-paid parking: eight case studies”, Transport Policy 4: 201-16

[54] Shoup, D. and Willson,R.(1992).“Employer-paid parking : the problem and proposed solutions”, Transportation Quarterly 46: 169-92

[55] Small, K., and J.Gómez-Ibáñez (1998). “Road pricing for congestion management: the transition from theory to policy, ” in K.Button and E.Verhoef, eds., Road Pricing, Traffic Congestion, and the Environment: Issues of Efficiency and Social Feasibility, 263-84, Edward Elgar: Cheltenham, UK

[56] Small, K.A. and C.Kazimi (1995). “On the costs of air pollution from motor vehicles”, Journal of Transport Economics and Policy 29: 7-32.

[57] Small, K., and J.Yan (1999) “The value of ‘value pricing’ of roads: second-best pricing and production differentiation”, Journal of Urban Economics, forthcoming

[58] Small, K., C.Winston, and C.Evans (1989) Road Work:A New Highway Pricing and Investment Policy, Brookings Institution, Washington, D.C.

[59] Strotz, R. (1965). “Urban transportation parables”, in J.Margolis, ed., The Public Economy of Urban Communities (Resources for the Future:Washington, DC)

[60] Syski, R.(1986).Congestion Theory, North-Holland, Amsterdam.

[61] Transportation Research Board (). Traffic Flow Theory

[62] Verhoef, E.

[63] Verhoef, E, Nijkamp, P. and Rietveld, P.(1995).“The economics of regulatory parking policies: the (im)possibilities of parking policies in traffic regulation”, Transportation Research, 29A: 141-56

[64] Verhoef, E, Nijkamp, P. and Rietveld, P.(1996).“Second-best congestion pricing: the case of an unpriced alternative”, Journal of Urban Economics 40: 279-302

[65] Verhoef, E, J.Rouwendal and Rietveld, P.(1999).“Congestion caused by speed differences”, Journal of Urban Economics 45: 533-56

[66] Verhoef, E and K.A. Small (1999).“Product differentiation on roads: second-best congestion pricing with heterogeneity under public and private ownership”, Discussion Paper TI99-066/3. Tinbergen Institute, Amsterdam-Rotterdam

[67] Vickrey, W.S.(1954). “The economizing of curb parking space”, Traffic Engineering Magazine, Nov. Reprinted in Journal of Urban Economics 36: 56-65(1994)

[68] Vickrey, W.S.(1959). “Statement on the pricing of urban street use”, Hearings, U.S.Congress Joint Committee on Metropolitan Washington Problems, November 11.[Reprinted in Journal of Urban Economics 36: 42-65 (1994).]

[69] Vickrey, W.S.(1968). “Automobile accidents, tort law, externalities, and insurance: an economist’s critique”, Law and Contemportary Problems 33:464-487

[70] Vickrey, W.S.(1969) “Congestion theory and transportation investment”, American Economic Review 59: 251-61

[71] Vickrey, W.S.(1979). “Spacing out those gregarious buses”, New York Times, June

[72] Walters, A. (1961). “The theory and measurement of private and social cost of highway congestion”, Econometrica 19:679-699

[73] Wheaton, W.C.(1978) “Price-induced distortions in urban highway investment”, Bell Journal of Economics 9: 622-32

[74] Wilson, J.D.(1983) “Optimal road capacity in the presence of unpriced congestion”, Journal of Urban Economics 13: 337-57

Hearn

dePalma, METROPOLIS

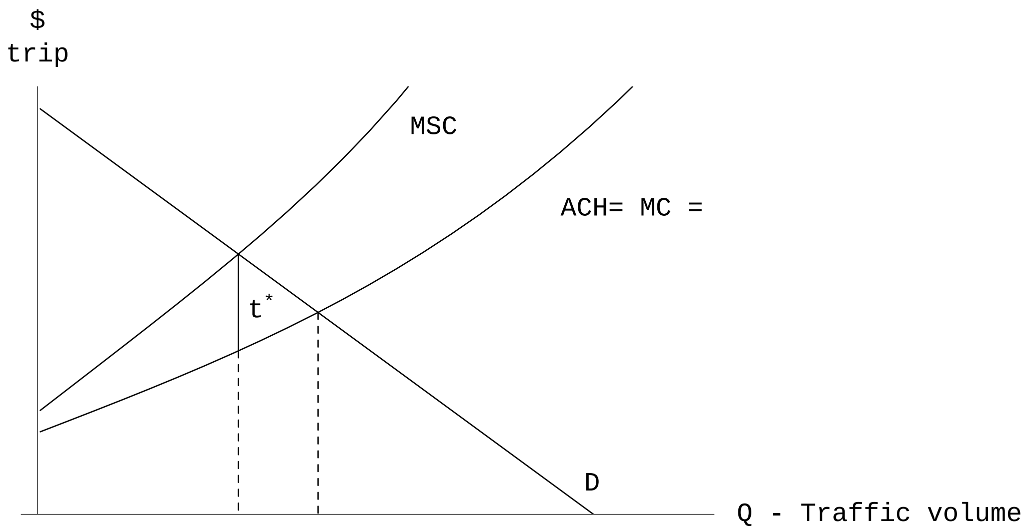

Figure 1. Diagramatic representation of the basic model of trafic congestion