Congestion Theory and Transport Investment

Author(s): William S. Vickrey

Source: *The American Economic Review*, May, 1969, Vol. 59, No. 2, Papers and Proceedings of the Eighty-first Annual Meeting of the American Economic Association (May, 1969), pp. 251-260

Published by: American Economic Association

Stable URL: https://www.jstor.org/stable/1823678

CONGESTION THEORY AND TRANSPORT INVESTMENT

By William S. Vickrey
Columbia University

"Deepening" and "Widening" of Transport Investment

Investment in transport facilities necessarily begins by being largely investment in the provision of new routes or new services under conditions of substantial indivisibilities and increasing returns to scale. Under these conditions the usual profitability tests for determining the desirability of specific investments lead generally to under- rather than to over-investment in transportation facilities. At this stage, cost-benefit analysis needs to include substantial elements of consumers' surplus on the benefit side in order to arrive at correct evaluations.

As investment proceeds, however, larger and larger proportions of transportation investment are made primarily, or at least in large measure, to relieve congestion on existing routes and to expand overall capacity. In such instances criteria based on apparent profitability may be seriously misleading in the opposite direction, and when notions of consumers' surplus are narrowly applied without regard to the overall situation, the errors may be compounded. This is especially likely to be the case where charges levied for the use of the existing and prospective competing facilities are far wide of the mark of representing marginal cost, as they often tend to be. It is this latter type of investment, designed to relieve congestion, with which this paper is concerned.

Types of Congestion

For purposes of economic analysis it is useful to distinguish at least six types of congested situations, though they are in fact often encountered in various combinations. These can be designated simple interaction, multiple interaction, bottleneck, triggerneck, network and control, and general density.

Single interaction occurs whenever two transportation units approach each other closely enough so that one or the other must be delayed in order to reduce the likelihood of a collision, no other units being sufficiently close to be immediately affected. This is the chief form of congestion encountered in light traffic. Total congestion delay tends to vary as the square of the volume of traffic; thus a motorist deciding on a trip under light traffic conditions will thereby inflict on others an amount of additional delay roughly equal to that which he himself will experience. (To be sure, for some types of vehicles the effect may not be symmetrical: slow-moving vehicles may tend to be relatively little delayed and fast-moving vehicles relatively more. The above relationship holds for an average vehicle.)

Multiple interaction tends to take place at higher levels of traffic density, short of capacity flows, where one can expect the average speed s to be a function of the flow of traffic x: $s=f(x)$. For traffic volumes ranging from about 0.5 to 0.9 of capacity, one can often fit a function of the form

$$(1) \qquad z = t - t_0 = \frac{1}{s} - \frac{1}{s_0} = ax^k$$

where t is the time required to go a unit distance under actual conditions, t_0 is the time required under very light traffic conditions, z is the average delay per vehicle, and a and k are constant parameters. For a relationship of this form, the total increment of delay given by

$$(2)\quad \frac{d(zx)}{dx} = z + x\frac{dz}{dx} = ax^k + xak\,x^{k-1} = (1+k)z$$

that results from a unit increment of traffic thus works out to $k+1$ times the delay experienced by the vehicle itself. That is, for every minute of delay directly experienced by the added vehicle, k minutes of delay are inflicted on the remaining traffic. For situations where considerable congestion exists, k is likely to be in the range of from 3 to 5 or even higher. The previous case is essentially that where $k=1$.

The pure bottleneck situation, which is the one that will chiefly concern us here, is one where a relatively short route segment has a fixed capacity substantially smaller relative to traffic demand than that of preceding or succeeding segments. There is thus relatively little delay as long as traffic remains below the capacity of the bottleneck, though small amounts of delay may occur as a result of stochastic variations in the level of traffic flow when the average flow is just below capacity. The important delays will occur when desired traffic flow continuously exceeds the capacity of the bottleneck for substantial periods. We then find that queues accumulate until either a period is reached when traffic demand is below capacity, or the prospect of waiting in the queue reduces the traffic demand by diverting it to another time or route or by suppressing the trip entirely.

A triggerneck situation develops from a bottleneck situation whenever the queue backed up from the bottleneck interferes with the flow of traffic not itself intending to use the bottleneck facility. The onset of incremental congestion may be quite sharp, and indeed in extreme circumstances a circular chain of triggerneck situations may bring traffic to a complete standstill, requiring that at least some of the vehicles involved actually back up before a forward movement can be resumed.

Network and control congestion results whenever the levels of traffic during the peak are reached requiring the application of additional control measures, whether in the form of regulations, stop signs, routing limitations, traffic lights, train controls, flight patterns and rules, or otherwise. Aside from the cost of these measures in themselves and even assuming that they are invoked only when the circumstances without them would be demonstrably worse, either from the standpoint of safety or delays, it is generally true that they cannot be applied with complete selectivity as to time and place, so that in most cases control measures required to take care of the most severe conditions will result in more delay under less severe conditions than would occur in the absence of controls, or with less restrictive controls adapted to the less severe conditions. Thus some of the delay experienced by off-peak traffic is in a medium-long-run sense caused by the increase in the peak traffic that made the controls necessary.

Finally there is a sense in which congestion costs in the long run are a function of the overall density of transportation flows in a given area for all modes combined and over all routes, even though some modes may contribute less to the total overall congestion relative to its traffic volume than other modes. Even if route separations are such that traffic on one mode has no immediate impact on traffic on the other modes, the construction of facilities to accommodate additional traffic on the one mode or route will not only encounter increased construction costs by reason of other existing facilities that cross its path, but such construction will at the same time be increasing the cost of constructing any other transportation facilities across its path in the future. In highly congested areas there is a very real sense in which long-run increasing costs may be encountered. It is very rarely that any account is taken, in the estimating of the cost of constructing facilities for a given transportation route and mode, of the increased costs of such future crossings by other links, even though good technical planning may sometimes make provision for such future crossings in the design.

Accidents as a Cost of Congestion

In addition to the cost of delays, the cost of accidents constitute an often overlooked element in the costs of congestion. While the incidence of traffic accidents does not arise with traffic density quite as rapidly as do time delays, one does expect, a priori, that as vehicle interactions per vehicle-mile increase, accidents per vehicle-mile will also increase. There is, indeed, a certain amount of empirical evidence that in a significantly wide range of situations this increase in accident rates with increasing traffic densities does in fact occur: for grade-separated limited access highways in California, it was found that the marginal increment in the number of accidents associated with an increment of traffic on a given type of highway was approximately 1.5 times the average accident incidence per vehicle-mile.[1] Thus whatever may be the effect on accidents of shifting traffic from other highways to grade-separated expressways, there is in addition a favorable effect on accidents of building roads of the same type to more ample dimensions and greater capacity, and an adverse effect on the accident rate per vehicle-mile of increasing the flow of traffic on a given roadway.

Before taking full credit for this benefit, however, it is necessary to examine the net safety effects of increasing total traffic flow overall, and of attracting traffic from other safer modes, such as rail transit. Doing too much in the name of safety considered in a narrow context can actually increase the overall death rate.

Construction to Ease Bottlenecks

Although the pure bottleneck situation is not typical of the general congestion picture, it is an important element in many cases of severe congestion and its relatively simple analysis does provide some valuable insights into the nature of the overall congestion problem.

Assume a situation in which $N = 7200$ commuters want to make a daily trip via a given bottleneck, and that in the absence of congestion their times of passing the bottleneck point would be distributed evenly over a period between $t_a = 8{:}00$ A.M. and $t_b = 9{:}00$ A.M., this permitting each commuter to arrive at his downtown destination at a desired time. If the capacity of the bottleneck were to be enlarged to

$$v_m = \frac{N}{t_b - t_a} = 120 \text{ cars per minute,}$$

then of course the capacity would just meet the requirements and no queue other than that due to stochastic variation would occur.

If the capacity is kept smaller than this, i.e., $v < v_m$, then it becomes impossible for all the commuters to arrive at their destinations just at the desired times. Some, at least, will have to arrive either late or early. In the absence of tolls the steady state that results will involve varying degrees of queuing, with those arriving at their offices closest to their desired times generally having to spend relatively more time in the queue than those who choose to push their arrival time further away from the desired time.

To keep the model simple, let us suppose that all commuters uniformly value time

---

[1] See William Vickrey, "Automobile Accidents, Tort Law, Externalities, and Insurance," *Law and Contemporary Problems*, Summer, 1968, pp. 467–68.

spent at home at $w_h = 2$ cents per minute, and time spent at the office at w_o, which for time prior to the desired starting time we suppose is $w_o = w_p = 1$ cent per minute, and for time after the desired starting time is $w_o = w_j = 4$ cents per minute. Time spent in the queue has a value of $w_q = 0$. It is readily seen that if an individual is to be maximizing the overall value of his time, he must be leaving the bottleneck point, subsequent to any queuing he may have had to endure, at a time such that

$$\frac{d_q}{dt} = \frac{w_h - w_0}{w_h - w_q} = \begin{cases} 0.5 & \text{for } w_0 = w_p = 1 \text{ cent/min} \\ -1.0 & \text{for } w_0 = w_j = 4 \text{ cents/min} \end{cases} \tag{3}$$

where $q(t)$ is the amount of waiting in the queue required in order to leave the bottleneck point at time t. A fraction

$$r = \frac{w_j - w_h}{w_j - w_p} = \frac{2}{3} \tag{4}$$

of the commuters will pass the bottleneck during the period of queue buildup and arrive at work at or before the desired starting time, the remaining fraction $1 - r = 1/3$ of the commuters will leave the bottleneck after 8:40 A.M. during the working off of the queue and arrive at or after the desired time. The total time required for the commuters to pass the bottleneck will be $N/v = 7200/v$, which will also be the length of time that a queue will persist, as long as $v < v_m$.

The length of the queue will build up linearly from zero at time

$$\begin{aligned} t_i &= t_a - r[(N/v) - (t_b - t_a)] \\ &= 8{:}40 - 4800/v \end{aligned} \tag{5}$$

to a maximum wait in the queue of

$$q_p = \frac{N}{v}(r)\frac{w_h - w_p}{w_h - w_q} = \frac{2400}{v} \tag{6}$$

for cars leaving the bottleneck at

$$t_p = t_a + r(t_b - t_a) = 8{:}40 \text{ AM} \tag{7}$$

after which it will again decline linearly to zero at

$$\begin{aligned} t_j &= t_b + (1 - r)[(N/v) - (t_b - t_a)] \\ &= 8{:}40 + \frac{2400}{v}\,. \end{aligned} \tag{8}$$

There will be a sharp discontinuity in the amount of delay experienced as the capacity of the facility is expanded past the point where $v = v_m = 120$ cars/minute. Below v_m, delay is inversely proportional to the capacity v, while at and above v_m the delay from queuing is zero. To be sure, matters will not usually work out as sharply as this, for there will usually be some variation in the desired rates of traffic flow near the peak rather than a peak that has an absolutely flat top, as has been assumed here for the sake of simplicity. Moreover, there will usually be some elasticity of traffic demand with respect to queuing time such as to suppress some traffic entirely rather than merely to shift its timing. Nevertheless, sharp discontinuities such as this emphasize the need for careful analysis of practical situations.

In practice, too, we are usually dealing with a dynamic situation in which traffic levels are generally growing at a substantial rate, while construction of additional facilities takes time and usually involves substantial lumps of additional capacity. In the face of such substantial penalties for either over- or underinvestment, substantial waste is likely unless some form of control over the use of the facilities is applied, such as is available through pricing, and which does not involve the wastes of queuing. Unlike the construction of additional capacity, prices can be adjusted upward or downward, as proves to be desirable, on relatively short notice and by relatively small increments.

Indeed, in the above situation one can readily compute the price structure that would just eliminate the queue and lead to efficient use, at least in the short run, of whatever facilities are actually in place. This will consist of a toll rising linearly from 0 at t_i to

$$p_p = \frac{Nr}{v}(w_h - w_p) = \frac{4800}{v}, \tag{9}$$

at t_p and then declining linearly to zero at t_j. With this pattern of tolls, each commuter will find that he can do no better for himself than to set out in time to pass the bottleneck at the same time that he would have left it after waiting in the queue in the zero (or constant) price situation. If traffic should fail to adjust its movement in such a fashion as to eliminate the queue, those finding themselves in the queue would have a motive to shift their travel time in such a way that the queue would be eliminated.

In the short run, the commuters are just as well off paying the variable toll and having no queue as they were before with no toll but with an equivalent queue; moreover there is no change in the pattern of arrivals at the city center. The revenue derived from the charges thus represents clear gain. We thus have an example of tax revenue that not only has no excess burden, it has no burden at all! To the extent that any of the revenue from the variable toll is used to reduce a preexisting flat toll, the motorists will be better off.

Obviously this does not mean that expansion of the capacity of the facility is never justified, but it does mean that the justification for such investment must be considered in an entirely different light if congestion charges are a possibility than if they are not. In the absence of congestion charges, a decision to expand facilities may have to be taken on an all or nothing basis. Expansion inadequate to take care of the entire traffic demand may result in a relatively slight improvement in conditions and may turn out to be hardly worthwhile, while a just slightly larger expansion might clear conditions up rather dramatically. Unfortunately, in a dynamically changing situation it is difficult to predict just what size of an improvement will in fact get over this threshold and for how long. It is easy to think of cases where an expansion of capacity felt to be quite ample when planned has turned out to serve merely to attract additional traffic until conditions are almost as bad as they were originally.

If, under conditions similar to the above, the levying of congestion charges is either an actuality or an alternative under consideration, benefits from the expansion of capacity are likely to be both smaller and less capricious in their behavior than if no pricing is contemplated. The net gain from the expansion of the bottleneck, assuming the adjustment of charges both before and after so as to just eliminate queuing, consists not of any reduction in queuing time (since there isn't any queuing) nor still less of the reduction in tolls, since this is merely a transfer from the government or operating agency to the users (and may entail substantial costs involved in securing an equivalent revenue from other sources), but simply in the fact that users will be traveling at times closer to the preferred times. The value of this shift in time may be measured by the difference in the value they place on their time at the two ends of the journey, i.e., $w_h - w_p = 1$ cent/minute for reductions in the amount by which commuters travel in advance of the preferred time, and $w_j - w_h = 2$ cents/minute for reductions in lateness. The total value of this delay, under optimum charging and no queue, is given by

$$\begin{aligned} & Nr(t_a - t_\alpha)(w_h - w_p)/2 \\ & \quad + N(1 - r)(t_\beta - t_b)(w_j - w_h)/2 \\ (10) \quad & = N^2 r \frac{1 - r}{2}(w_j - w_p)\left[\frac{1}{v} - \frac{1}{v_w}\right] \\ & = \$1440 \frac{120 - v}{v} \cdot \end{aligned}$$

Where queuing results from the absence of tolls, the average queuing time is half the maximum, or $q_p(\frac{1}{2}) = 1200/v$, which when evaluated for 7,200 cars per day at w_h 2 cents/minute amounts to $172,800/$v$.

The results, for various values of v, are summarized in Table 1.

Imposition of the optimal variable toll in each case eliminates queuing and results in toll revenues equal to the cost of the eliminated queuing. The displaced arrival cost will be the same whether or not the optimal toll is imposed, as this depends only on the capacity of the facility.

If we now suppose that initially we have a 2-lane bottleneck with $v = 60$ cars per minute, then without the control provided by the variation in the toll rate congestion costs will total $4,320 per day. Without help from toll adjustments, opening up a third lane at a cost of $2,000 per day would reduce congestion costs by only $1,920, and so would not be worthwhile, although opening up two new lanes at a cost of $4,000 would eliminate the entire $4,320 worth of congestion and yield a net gain of $320. On the other hand, the institution of variable tolls according to an appropriate pattern would cut congestion costs of the 2-lane bottleneck by $\frac{2}{3}$ to $1,440 and result in a budget inflow of $2,880 instead of a budget outflow of $4,000 if the additional two lanes were built. If the overhead costs of obtaining public funds from other sources were as much as 10 percent—a not unreasonable figure if all of the unfavorable results of increased tax rates are allowed for—this budgetary shift would constitute a further gain of $688, for a total gain of $2,880+$688 =$3,568, a substantially better result than any that can be obtained without toll variation, and much better, in particular, than the $320 gain from the 2-lane expansion.

If, on the other hand, the cost of expanding the bottleneck is relatively low, say $20 per day for each vehicle per minute of increase in capacity, so that some addition to capacity will be worthwhile in any event, variable tolls can still play a

TABLE 1

Capacity (cars per minute)	Equivalent Number of Lanes	Duration of Queue or Toll (minutes)	Maximum Wait in Queue (minutes)	Average Toll Rate (cents)	Congestion Cost ($/day)		
					Displaced Arrival	Waiting in Queue (=toll rev.)	Total without Pricing
50	1.67	144.00	48.00	48.00	2016	3456	5472
60	2.00	120.00	40.00	40.0	1440	2880	4320
70	2.33	102.9	34.29	34.3	1029	2469	3498
80	2.67	90.0	30.00	30.0	720	2160	2880
90	3.00	80.0	26.67	26.7	480	1920	2400
100	3.33	72.0	24.00	24.0	288	1728	2016
110	3.67	65.6	21.91	21.9	131	1571	1708
115	3.83	62.6	20.87	20.9	63	1503	1566
118	3.93	61.0	20.33	20.3	24	1464	1488
119	3.97	60.5	20.17	20.2	12	1452	1464
119.999	4.00−	60.0	20.00	20.0	0.12	1440	1440
120.001	4.00+	0	0	0	0	0	0

role in reaching the optimal result. In the absence of any tolls the best available alternative would again be to expand capacity to slightly above $v=120$ cars per minute, so as to eliminate congestion entirely. If toll controls are available, however, it would not pay to carry the expansion much beyond $v=90$, since expanding from $v=90$ to $v=100$ would reduce losses from displaced arrival by only $192, as compared with the cost of $200 for this expansion. In addition, such an increase might entail an increase in budgetary problems by the difference in revenues of $392.

The use of congestion tolls as an element in developing an efficient transportation system is thus not only a means of providing optimal adjustment in the short run but is likely to remain an important element even in the long run. It is only if the increment to capacity is provided in a manner that incidentally provides a substantially new and different route or a shorter origin to destination time for a substantial amount of traffic, or possibly where incremental costs of adding facilities for the collection of congestion charges *de novo* bulk large, that it would be possible to omit such charges from an optimally efficient scheme.

In practice, of course, most bottleneck situations are not as simple or as clear cut as the above case. The desired times of passing the bottleneck are not usually uniformly distributed, individuals vary in the values they asign to time spent in various places at various times, and to some extent the total number of trips made through the bottleneck would be affected by the tolls or the congestion conditions, probably in different ways for different users. These complications tend to lessen the sharpness with which critical capacity is determined, but in other ways they may enhance the effectiveness of appropriately graduated tolls and charges improving the efficiency of whatever facilities are constructed.

Expansion of Routes in the Presence of Alternative Routes

One situation that makes appropriately graduated charges even more essential is where a significant part of the traffic has closely competing alternative routes available to it. The classical paradigm of this situation is one where the alternative to the bottleneck route is, for a substantial portion of the traffic, a more circuitous or slower route of ample capacity. In the absence of any charges, the traffic will divide between the two routes so as to equalize total travel costs per vehicle, including travel time and also the queuing time on the bottleneck route. An enlargement of the bottleneck under these conditions will, if it falls short of being able to accommodate all of the traffic, simply result in enough traffic being diverted from the circuitous route to the enlarged bottleneck route to maintain the queue at the former level. The enlargement may thus produce no improvement in travel times at all, at least during periods of peak traffic. In a sense, such a costly enlargement proves worthless precisely because it is free.

In a more general vein, traffic often behaves like population. It has been said that if nothing stops the growth of population but misery and starvation, then the population will grow until it is miserable and starves. Similarly, if the use of private automobiles for access to the cores of large metropolitan areas is so attractive, under uncongested toll-free conditions, relative to other modes, that in effect nothing stops the growth of such traffic but congestion and delay, then such traffic will grow until sufficient congestion and delay are generated to constitute a deterrent, or until the core begins to suffer from gangrene, at which point a cumulative decline may set in that may be difficult to reverse, even with a belated introduction of appropriate toll graduation.

In practice a situation not too far from the classical paradigm often presents itself where attempts are being made to improve access to the core of a metropolitan area. In the absence of pricing, the alternatives may consist of (1) the *status quo*, (2) building an access facility sufficient for the traffic bound for the center, but which will immediately become so clogged with through traffic as to provide only moderate improvement in the speed and convenience of access to the center for a substantial part of the day, and (3) building a huge artery sufficient to take care not only of all local traffic but of any through traffic for which circumferential routes cannot be made sufficiently attractive to divert traffic from the central artery under uncongested conditions, and at the same time trying to bring the circumferentials close enough in and of such high grade as to divert as much of this traffic as possible. This third alternative is likely to prove astronomically expensive as well as disruptive of community amenities; the second alternative may in the end yield a very low return on the investment cost in terms of improvement of traffic conditions. Thus on balance the prospects of a net gain over the *status quo* may be rather dim whatever is done.

The availability of pricing opens up a new alternative of constructing central access facilities scaled to the requirements of traffic actually requiring to go to or from the center, with through traffic during the peak period being fairly thoroughly diverted to the circumferential routes by the charges imposed for the use of the central route. In some instances circumferentials that might not be worth their cost in the face of the difficulty of locating them so as to attract a sufficient volume of traffic in the absence of toll controls will become more worthwhile if pricing is available to help in the optimum distribution of traffic. With pricing such circumferentials may also be more readily located where construction is cheaper and less disruptive of amenities, and possibly also better suited for that part of their traffic that is not in any case tempted to use the central route, without fear that impairing their competitive relation to the central route would lead to undue congestion on the latter.

Variations in the Value of Time

An important but not essential element in the strategic importance of pricing as a factor influencing investment decisions is the existence of variations in the value of time, not only for different persons at the same time, but for the same individual at different times. In the absence of pricing, expansion of capacity must provide indifferently for individuals for whom the improvement will be worth relatively little as well as for those for whom it may be worth a good deal more. Pricing makes it possible to exclude the low-value uses and base the magnitude of the improvement primarily on the uses that are valued sufficiently highly so that they warrant the marginal cost of the final increment to the magnitude of the improvement. The selective effect of pricing on the costs of congestion would be to reduce still further the figures near the bottom of the "Displaced Arrival" cost column of Table 1, enhancing the gains from the earlier increments to capacity and reducing the potential gains from the final increments. Potential improvements in efficiency and savings in construction costs are thus increased significantly over the amounts calculated on the basis of a uniform value for time.

There is, to be sure, likely to be an outcry at this point that pricing discriminates against the poor by forcing them off the congested highways. Actually the number of really poor individuals who are under any strong compulsion to drive cars with any regularity on the congested highways at peak hours appears to be quite negligible. The poorest among those significantly affected by a program of congestion charges are likely to be still somewhat above the poverty line. To the extent that this level of incomes is considered to be in need of a subsidy, there are surely better ways of determining needs than the amount of congested driving done.

A somewhat parallel outcry against the use of appropriately graduated landing fees as a means of controlling congestion at busy airports is even more difficult to justify. It is bad enough when a facility used primarily by the well-to-do is subsidized from tax revenues derived at the margin in large part from taxpayers of lower incomes; at uncongested airports this has at least the virtue of promoting better utilization. But when landing fees geared to congestion costs would substantially improve utilization while costing relatively little to assess and collect, even this excuse is lacking. In the airport case, moreover, those who would be charged the highest tolls, on a per capita basis, would be primarily general aviation planes operated in many cases for company executives and the like or private planes used for recreational and other purposes. Such users would in general be far better able to find acceptable alternatives, such as use of some of the smaller airports, if they are unwilling to pay the appropriate charges, than the patrons of the scheduled airlines. Landing fees reflecting congestion costs at various times could bring about a coordination of use that might well defer for a considerable time the need for resort to costly additional construction, often at less convenient locations. This can come about in part through the diversion of general aviation flights to other airports, through adjustment of airline schedules to reduce the concentration at the peak hours, by the use of larger planes and by scheduling for increased load factors, even without adjustment in the fare structure. If in addition some of the congestion charges can be shifted forward to passengers through the fare structure, some diversion of travel to less congested times and via less congested interchange airports may also aid in alleviating congestion. Such fare adjustments may also be essential if the best allocation of traffic between short-haul air travel and ground transportation, high-speed or otherwise, is to be achieved. A rush to construct additional airports to take care of threatened congestion may prove particularly costly at the present juncture in that improved navigational and flight control methods seem to be on the verge of substantially increasing the capacity of present airports.

Evaluation of Investment in Congestion Relief

Finally, the information provided by a system of congestion control through pricing has an essential role to play in the evaluation of investments designed to afford relief from congestion. In the absence of congestion pricing, very little solid data exists on the value of varying degrees of congestion alleviation, and much of what exists is subject to considerable bias in the direction of overestimating the value of such improvements. For example, a recent study of changes in street use and traffic patterns in central London over the period from 1960 to 1966 came to the conclusion that what was widely touted as a significant improvement in traffic flow was actually no net improvement and possibly a deterioration from the standpoint of origin-to-destination volumes and times.[2] Many of the "improvements" during this period consisted of conversion to one-way traffic and the prohibition of certain turns involving

---

[2] J. M. Thomson, "The Value of Traffic Management," *J. of Transp. Econ. and Policy*, Jan., 1968, pp. 3–32.

crossing other traffic flows. Although average speeds and flow volumes of vehicles passing given points may have increased, these increases appear to have been used up in traversing more circuitous routes between given origins and destinations. In a similar way, higher speeds and volumes of traffic recorded on turnpikes and expressways often significantly overstate the increase in the transportation service accomplished as a result of their construction in terms of delivery from a specific origin to a specific destination. This is because distances, especially for shorter trips, are often longer via the new routes than via the old, though perhaps not as much as in the London example just cited.

If, indeed, all routes were subject to appropriate congestion tolls, the level of these tolls would then be a good initial approximation to the value of the congestion relief afforded by investment in increased capacity, at least for small increments. For larger increments one could also then rely, to some extent, on estimates of consumers' surplus under each demand curve separately. But where charges for the use of alternative routes fail to reflect congestion costs at the margin, the problem becomes much more complicated. Not only must allowance be made for the indirect effects on competing routes but consideration must be given to the possibilities for improving efficiency through introduction of appropriate patterns of user charges. In the absence of the information that would be provided by the charging of appropriate tolls, planning of investment in expanded transportation facilities is half blind, and resort is sometimes had to arbitrary rules of thumb, such as that of providing capacity adequate to handle the traffic during the thirtieth heaviest hour of traffic out of the year. The capriciousness of such a rule should be fairly obvious.

Appropriate patterns of congestion tolls are thus essential, not only to the efficient utilization of existing facilities, but to the planning of future facilities.