INCENTIVES FOR THE COORDINATION OF DECENTRALIZED TRANSIT CHOICES

Roger Sherman, University of Virginia

In a well-functioning competitive market system, decentralized choices by individuals can serve their interests well; but in a typical contemporary setting, such choices, at least about transit, do not. Flaws in the allocative consequence of decentralized transit choices are illustrated here by reference to an ideal set of price signals and description of some of the ways in which real-world prices differ from the ideal ones. One consequence of this distortion between real and ideal price signals is that bus companies may not be able to survive a profit test in cities even though bus service actually may be desired and could survive under ideal prices. Tax and subsidy incentives that can approximate the ideal price signals are described, and incentives for efficient operation of bus services also are briefly noted.

•IN A well-functioning market economy, each person makes choices about what to consume based on prices that reflect fully the marginal cost to society of providing one more unit of each good or service. Inefficient producers will lose money and automatically will be forced to withdraw from their markets, while profit opportunities will attract more resources to those activities that we consumers wish to have expanded. These properties of a competitive market system are so well known and we are so accustomed to their benefits that we may think decentralized choices always are efficacious. If for any reason, however, prices do not reflect the marginal social costs of goods and services, the signals that individuals act on will be wrong, and resources will not be attracted automatically where they can serve consumers best. We must then appropriately consider alternative incentives that might correct the wrong prices and thereby enable us to use our resources efficiently.

Every day, we make transit choices on a decentralized basis, but are guided in our choices by prices that are faulty. We squander certain of our resources as a result because the faulty prices that we rely on lead us to use too much of some things and not enough of others. Our problem is a very difficult one, however. We fail to solve it efficiently not because we are lazy or dumb but because it is a very difficult problem that still is only partly understood. Our purpose here is, first, to illustrate the kind of problem we face in trying to choose together through decentralized private decisions an efficient solution to our transit problem and, second, to indicate ways that the problem might be solved. Means of controlling the transit agency in case subsidies are given are also considered.

PROBLEMS IN COORDINATING DECENTRALIZED TRANSIT CHOICES

Road-use patterns are commonly accepted and are so similar in U.S. cities that we can easily fail to appreciate their faults. And alternative patterns are hard to imagine. To bring out problems in our present system, let us begin with a scheme for optimal control of highway transit and then move gradually to the system we actually find today in U.S. cities. We consider automobile and bus modes, both using the same road space, and we concern ourselves primarily with the peak hours, when most travel occurs.

One Road, Automobile, and Bus Company

Let us consider an illustrative city. Highways offer the only means of travel, and they accommodate both automobiles and buses. All automobiles, buses, and highways are owned, however, by the city's Grand Road, Automobile, and Bus Company (GRAB). An individual can either hire a car or ride a bus and in each case must pay a specified fee per mile traveled. We shall first assume that GRAB is a truly benevolent welfare-maximizing organization. For such a situation, we ask: What prices should GRAB charge per automobile-mile and per bus-mile? What quantities of road, automobile, and bus facilities should it provide?

We obviously must make some heroic assumptions in posing this problem, pretending for instance that automobile operating and maintenance costs will not be very much affected whether automobiles are privately owned or leased from GRAB and that all the other administrative problems of operating such an unusual organization in a welfare-maximizing way can be handled. But let us first skip such problems to focus on important characteristics of urban road transit, namely, the excessive use of roads by private parties, the consequently high level of congestion, and the fact that automobile passengers contribute more to such congestion than bus passengers do.

To maximize social welfare, GRAB should choose prices that equal the marginal costs of each of the 2 transit modes it can supply. Following this rule may lead to surpluses of profit if average cost is increasing or to losses if average cost is decreasing, and the long-run consequences of these profit or loss results are important. The distribution of profits will affect income distribution and possibly efficiency, and raising funds to cover losses will affect efficiency under any feasible taxation scheme.

Let us set these problems aside for the moment, though, to focus solely on efficiency and the ideal price-equals-marginal-cost rule that will ensure it. The logic of this well-known welfare-maximizing rule rests on the fact that, if one person pays less than the marginal cost to society for a service he consumes, the remaining cost must be borne by others, a result that is not only unfair but that also leads the person to consume a nonoptimal quantity of service. The total cost of serving one person will be higher when the person does not face the marginal cost of his actions, for he will tend to consume more of the service when he pays less for it. His excessive consumption, beyond what he would choose if he paid marginal social cost, is inefficient. Any resources that are used by a person who values them at less than their value in alternative uses ought really to be shifted where they are valued more. In principle, the person now using them could be paid an amount equal to the value he places on them, so that he would remain as well off, while others who value them more could pay that amount and enjoy a net benefit from use of the resources. Using resources where they are valued more will add to welfare.

It is possible in a simple illustrative situation to sort out congestion effects in a reasonably straightforward way and provide GRAB with ideal welfare-maximizing price choices. Let us consider only daytime traffic, which we shall assume to be of uniform density at all times and places so that we can concern ourselves with just one problem and one solution. The ideal price per automobile passenger-mile will reflect the marginal social cost of an additional automobile passenger-mile. That cost not only includes the effects on other automobile travelers but also must take into account the effect that an automobile passenger-mile will have on the marginal and average cost of a bus passenger-mile. This second element captures the increase in congestion due to automobiles that actually affects buses. Similarly, the ideal price per bus passenger-mile will reflect the marginal social cost of another bus passenger-mile, plus an additional element to take into account the increase that a bus passenger-mile will cause in the congestion that affects automobiles.

These ideal prices can be summarized algebraically. If automobile passenger-miles are M_A, the average cost of an automobile passenger-mile is AC_A, and the total cost of automobile passenger travel is TC_A, then the marginal cost to automobile passengers of an automobile passenger-mile, MC_A, is

$$MC_A = \frac{\partial TC_A}{\partial M_A} = \frac{\partial (AC_A \cdot M_A)}{\partial M_A} = AC_A + M_A \frac{\partial AC_A}{\partial M_A}$$

The same definition of marginal cost to bus travelers alone can hold for bus travel if we replace A subscripts with B subscripts to represent buses:

$$MC_B = AC_B + M_B \frac{\partial AC_B}{\partial M_B}$$

Rather than focus on the cost of automobile or bus travel alone, though, we must consider them together because an additional automobile or bus affects costs in both modes. The effect of a marginal automobile passenger-mile on total cost, $TC_A + TC_B$, is

$$\frac{\partial (TC_A + TC_B)}{\partial M_A} = AC_A + M_A \frac{\partial AC_A}{\partial M_A} + M_B \frac{\partial AC_B}{\partial M_A} = MC_A + M_B \frac{\partial AC_B}{\partial M_A}$$

Because the ideal price per automobile passenger-mile, P_A, will equal marginal social cost, we have

$$P_A = AC_A + M_A \frac{\partial AC_A}{\partial M_A} + M_B \frac{\partial AC_B}{\partial M_A} = MC_A + M_B \frac{\partial AC_B}{\partial M_A} \tag{1}$$

The term, $M_B \partial AC_B / \partial M_A$, represents the addition to marginal cost felt by bus passengers but not experienced by automobile passengers and therefore omitted from the definition of MC_A. Similarly, the marginal social cost and ideal price, P_B, of a bus passenger-mile will be

$$P_B = AC_B + M_B \frac{\partial AC_B}{\partial M_B} + M_A \frac{\partial AC_A}{\partial M_B} = MC_B + M_A \frac{\partial AC_A}{\partial M_B} \tag{2}$$

These expressions reflect the fact that each transit mode ideally will recover its marginal cost and, in addition, the contribution it makes to the marginal cost of the other mode. Let us note that, by taking up more road space, an automobile passenger contributes more to congestion than a bus passenger. More precisely, it is almost certainly true that $\partial AC_A / \partial M_A > \partial AC_A / \partial M_B$ and $\partial AC_B / \partial M_A > \partial AC_B / \partial M_B$. [Support for these conditions is given in another paper (1).] There is an average of only 1.5 passengers per automobile in cities, even during congested hours; and, although a bus may require as much road space as 3 or more passenger cars, it carries many times more passengers.

There is one consequence of such welfare-maximizing pricing policies that might seem embarrassing to the managers of GRAB: Although it pursues wholeheartedly a socially efficient solution, the transit agency will be very profitable. If it merely charged a price equal to the marginal cost of each of the transit modes it controls, the agency would do well because marginal cost exceeds average cost in each case. But the ideal prices are higher still to take account of the contribution each mode makes to marginal cost in the other, and so GRAB will be even more profitable. GRAB achieves economic efficiency with its ideal prices, but in achieving efficiency no attention has been given to income distribution. And yet how GRAB's abundant profit is distributed is an important question.

The presence of congestion means that road space is scarce and that charging a price to ration its use can yield a profit. Let us tackle the question of how to distribute this profit equitably by posing this question: Among all members of society, what sort of bargain might be struck to determine who would use a scarce facility that no one person or group owned? Quite obviously, some would pay others not to use the facility, and those most keen to pay to use it thereby would be accommodated. Now in practice such a bargain may be impossible to arrange. But if we charge a price for the use of roads and distribute any resulting profits to all citizens, road-users and non-road-users alike, we accomplish—crudely—the same sort of bargain. For by charging each traveler the marginal social cost of his travel decision, we discourage from traveling those who do not value travel very highly, and we ration use of a scarce facility to those who value it most. Because the profits of GRAB amount really to the proceeds from such a rationing price, they reasonably could be distributed to all citizens.

Early criticism of road-pricing arguments overlooked this possibility of redistributing the proceeds of a congestion toll or tax. It is true that road users usually will "lose" if taxes are imposed to control congestion and the road user never benefits in any way from the proceeds of the tax. [In his criticism of congestion tolls, for instance, St. Clair (2) remarked, "It seems adding insult to injury first to recognize that that time delay is a cost [to the road user], and then to say he must pay, in order to maximize his own benefits, a tax that will leap from the average to the marginal point." Actually, if the congestion toll proceeds never benefit the road user in any way, a more apt description of the toll is that it "adds injury to efficiency." The toll can achieve efficient road usage. But if it also takes income from road users and transfers it to others, the road user can still lose. In unusual cases it is true that the efficiency gain can make him better off despite the loss he suffers in making toll payments. Moore (3) gives descriptions of such cases. Vickrey (4) reviews evidence on how significant the efficiency losses can be.] But the tax should not work that way if it is to be equitable. The road user pays the tax per unit of travel so that usage of the road can be controlled marginally and thereby can be made efficient, but the proceeds of the tax should be returned to him in a lump-sum form, perhaps as a reduction of his other tax obligations, so the main effect of the tax is to accomplish the marginal adjustment without vastly redistributing income away from automobile travelers at the same time. If we can assume that automobile travelers will share in the proceeds of the congestion toll, or tax, it becomes an equitable as well as an efficient way to deal with traffic congestion.

The long-run decision about the quantity of road, automobile, and bus facilities can be a very difficult one. If the short-run marginal costs of automobile and bus travel can be identified for any existing level of road supply and trip demand, then optimal prices can be determined that will result in using existing facilities efficiently. Such short-run marginal cost prices also can indicate the value consumers place on existing facilities. By their size, those prices therefore can tell whether the cost of expanding road space and other facilities ought to be incurred. Here there are especially serious difficulties, however. First there is a discreteness in road expansion that requires large rather than small changes. Expansion calls for adding new lanes and possibly new roads rather than making small marginal adjustments within the range of existing usage, where evidence of marginal values can be more reliably estimated. Second, these changes can affect the demand for trips as well as their supply. Knocking out stores to make room for an additional road lane may remove (or relocate) the causes of some trips (to go to those stores) and thus can cause a reduction in trip demand at the same time it makes an increase in supply possible.

The main point is that reasonably efficient short-run prices (or tolls or taxes) are needed first, for they will give some indication of the value road users place on increased system capacity. Congestion itself can never justify road expansion; the congestion may be due to faulty pricing of existing roads rather than inadequate provision of roads. And if road capacity expansion is characterized by increasing cost, as is probably the case in real cities, efficient short-run tolls or taxes should allow net proceeds in excess of existing road costs. A break-even rule is not enough. Road operations ideally will be "profitable" instead, and the net proceeds should be returned to citizens through some sort of payment that will not affect their marginal travel choices. Then transit choices can be efficiently and also equitably coordinated.

Private Ownership of Automobiles

Now let us weaken GRAB's control of all metropolitan transit and permit private ownership of automobiles. Doing so can interfere seriously with the efficiency of urban transit. First, private automobile ownership allows, in effect, a 2-part price for automobile travel, while GRAB presumably will continue to collect a single fee for each bus trip. Automobile owners pay fixed outlays to become owners and then can make smaller outlays per mile traveled because they pay some costs (e.g., insurance, license fees, part of depreciation) independent of the amount they travel. The consequent 2-part price for automobile travel enables those who travel more to achieve a lower average cost because they can spread the fixed costs of automobile travel over more miles. Those

who travel more are therefore more apt to commit themselves to automobile ownership to obtain the benefits of a 2-part price (5, 6). In what follows we shall ignore this slight bias toward automobile ownership brought about by lower average cost through a 2-part price, although it actually would tend to bias our private choices more in the direction of the automobile.

Apart from this ownership bias, in the absence of special efforts to coordinate the perceived automobile and bus travel costs, automobile travel will no longer be decided based on marginal social cost. An automobile owner can join a traffic stream and share in the average delays that are experienced, so his or her decision will be based on average social cost. There will be an excessive, inefficient use of roads by automobile passengers as a result because, when it is equal only to average cost, the marginal benefit for automobile passengers will lie below marginal cost. Indeed, the well-known traffic congestion problem that we see daily in real-world city streets is caused by our failure to charge a proper price to those automobile travelers who cause highway congestion. Each of us reckons with average costs and delays when we decide to join in road traffic, but we do not consider the delays that we cause others on the road; those delays to others nonetheless are part of marginal social cost. In terms of Eq. 1, we pay only AC_A per automobile passenger-mile and ignore $M_A \partial AC_A/\partial M_A + M_B \partial AC_B/M_A$, so our effective automobile travel price is below the ideal P_A. Because average cost rises with an increase in traffic, marginal cost must be higher than average cost; and the equilibrium we reach is not optimal because it occurs where the benefit from a marginal trip equals its average social cost and thus lies below marginal social cost. That is why the marginal social cost of trips on congested roads exceeds their social benefit.

Now what should GRAB do? If automobile travel decisions are based on average rather than marginal social cost, what is the optimal price for bus trips? If GRAB tries to maximize welfare by dealing as efficiently as possible with congestion, it now may set a bus price below the average cost of providing bus service. Such a choice can be appropriate because without it too many resources will go to automobile travel, and automobile passengers pay only average rather than marginal social cost and yet contribute more to congestion. Drawing passengers away from automobiles with a lower price for bus transit can offset this misallocation. If a marginal bus passenger would eliminate a marginal automobile traveler, for instance, the effect of their attendant contributions to congestion could be to reduce overall congestion because the automobile traveler's contribution is greater. In that case of induced substitution from automobile to bus travel, the net social cost of a marginal bus passenger would be low because by choosing to travel by bus he actually would reduce congestion. His addition to congestion as a bus passenger would be lower than the reduction in congestion caused by his withdrawal as an automobile traveler. Such an effect depends again, of course, on some substitutability of bus travel for automobile travel.

Thus, in the presence of nonoptimal pricing of road use for automobiles, the price that GRAB should set for its bus passenger-miles will depend on the tendency of travelers to respond to price changes by substituting one mode of travel for another. If they do not respond to prices at all, there is little we can do through bus pricing alone about excessive automobile use and its consequences for the level of congestion. But evidence that individuals respond to relative prices in their other decisions is so abundant that we cannot fail to expect it here. So let us describe the conditions that determine whether the optimal price per bus passenger-mile is above or below average cost.

In a simple model it can be shown that, when automobile users make their private travel decisions at average rather than marginal social cost, we can use our transit resources more efficiently by pricing bus travel below its average cost if the following condition is met:

$$\frac{\partial AC_B/\partial M_B}{\partial AC_B/\partial M_A} < \frac{\partial M_A/\partial P_B}{-\partial M_B/\partial P_B}$$

[This condition is a slightly modified version of the requirement derived in other papers (1, 7). The derivation actually assumed that $(\partial AC_A/\partial M_B)/(\partial AC_A/\partial M_A) = (\partial AC_B/\partial M_B)/$

($\partial AC_B/\partial M_A$). Also the $\partial M_A/\partial P_B$ and $\partial M_B/\partial P_B$ responses are "income-compensated" responses that reflect effects of substitution but not income change. Because income effects resulting from transit outlays probably are small, this proviso should not be crucial.] We have assumed that a marginal automobile traveler causes more congestion for both automobile and bus modes, and when that is true $\partial AC_B/\partial M_B < \partial AC_B/\partial M_A$, so the left side of this inequality will be less than 1. Because its own price usually has more effect than prices of other services on the quantity demanded of any one service, we can expect the denominator of the right side to be larger than its numerator, too. But, although each side is less than 1, we cannot be sure whether this inequality condition will be satisfied. It certainly can be satisfied though (7), which means it can be more efficient to subsidize a bus service in order that it can set its price below its average cost.

We might refer here to the much-heralded experiment in Rome, where free public transit failed to solve a problem of traffic congestion that was due primarily to automobiles. Observers now concede that, perhaps because of the design of existing transit service in Rome, reducing its price brought former pedestrians to use it rather than automobile travelers. In terms of the inequality above, this would mean that $-\partial M_B/\partial P_B$ was large but $\partial M_A/\partial P_B$ was not large, so the right side probably was too small to satisfy the condition that would warrant subsidizing the transit service. A different kind of service, however, that would attract automobile travelers instead of pedestrians still might have warranted a subsidy.

There actually is a way to approach much more closely the original optimal solution chosen by GRAB when it controlled prices of both automobile and bus passenger-miles. Suppose that, in addition to setting P_B, GRAB is given authority to impose a tax, t, on all the inputs that account for operating costs AC_A and AC_B (such as gas, oil, tires, vehicles, and repairs). We assume that any tax must affect both modes because it is not feasible to tax gasoline at different rates for buses and automobiles without incurring prohibitively high enforcement costs. Thus, the input tax necessarily would raise costs in both modes to $AC_A(1 + t)$ and $AC_B(1 + t)$. By adjusting the input tax rate and the level of P_B, the transit firm can achieve prices exactly equivalent to marginal social cost for each mode, just as if it chose P_A and P_B. The only remaining difference is that individuals (rather than GRAB) now decide the quantity of cars, and because of the 2-part price advantage of automobile ownership more automobiles may be chosen by individuals.

It is worth emphasizing, however, that with the input tax the optimal level of P_B is even more apt to be below the average cost of bus travel, which now includes the cost of the input tax. So a subsidy to GRAB will be needed. The ideal P_B will lie below $AC_B(1 + t)$ as long as the ratio of bus to automobile contribution to congestion is lower than the ratio of bus to automobile average cost (7). That is, the ideal level of P_B will be below this average cost, $AC_B(1 + t)$, when

$$\frac{\partial AC_B/\partial M_B}{\partial AC_B/\partial M_A} < \frac{AC_B}{AC_A}$$

This requirement appears very likely to be fulfilled in a city; so on its bus operations alone, GRAB would appear to lose money. It is possible to show, however, that proceeds from the input tax will always exceed these apparent losses on bus operations. Some of the input tax proceeds thus can go to provide the subsidy without causing new tax burdens elsewhere. Because bus passengers contribute proportionately less to congestion but the input tax cannot be adjusted for that difference, a rebate for them on the input tax is appropriate on both equity and efficiency grounds.

Private Ownership of Automobiles With GRAB as a Private Monopolist

Suppose now that we permit private ownership of automobiles, turn operation of the bus service over to GRAB, suspend its welfare-maximizing purpose to let it maximize profit, have local government decide on input taxes and the quantity of roads, and leave the road to the use of both automobiles and buses. The result then will approximate more the sort of situation we have now in many cities. As long as the automobile can

serve as a good substitute for bus travel, the monopoly position of GRAB is not really very lucrative, for its attempts to raise price will be met by substitution of the automobile mode for bus travel. (As a recent case in point, the Louisville Transit Company recently gave notice that it will give up its bus franchise in 1974, although it has a right to keep it until 1981.) Thus, GRAB has a monopoly of bus service, but not a monopoly in the classical sense of producing a service for which no good substitutes are available. And as the automobile, its main substitute, becomes more popular, the service a bus can offer is apt to become more costly and less attractive because of greater congestion delays.

Any effect of the 2-part price opportunity available through automobile ownership now will be harmful to GRAB, for a lower average cost for automobile travel via the 2-part price will tend to increase M_A and reduce M_B. Moreover, because the upper limit that the private automobile ownership option places on the bus travel price may be below the average cost of bus service, it is very possible that GRAB will lose money. Imposing an input tax to reduce congestion probably will be slightly prejudicial against buses, too; automobile and bus costs will tend to go up proportionately because of the tax, but the contribution of automobile passenger-miles to congestion exceeds that of bus passenger-miles. As the input tax is set higher, the ideal price for bus travel will fall relative to the automobile travel price that is perceived by travelers because the input tax alone tends to affect bus passengers more, relative to their marginal contribution to congestion. A lower fare for bus travel can offset that input tax bias. In principle, when combined with input taxes, a subsidized bus service can permit final effective automobile and bus travel prices that equal marginal social costs, equivalent to the ideal P_A and P_B values considered earlier and given in Eqs. 1 and 2. Thus, the money-losing bus service we see in the real world may not be a sign of faulty pricing because the ideal price per bus-mile probably should be below its average cost. A subsidy to make up this loss properly should be paid, however, out of proceeds of the input tax.

To the extent that a profit-maximizing bus service could survive profitably in a city, probably those services that did not cover their average costs would be suspended. Yet such services conceivably could add to overall transit efficiency by easing road congestion and might be the most valuable to a community. The profit test simply is inappropriate in the urban transit setting and can yield perverse results. The final outcome also would handicap a city government in choosing what quantity of roads to provide because the prices implicit in road use no longer would reflect faithfully the value of roads to consumers. The wrong price for automobile travel—at average rather than marginal social cost—has induced too many of us to choose the automobile as a mode of travel. We have driven profit-seeking, private-enterprise bus services out of business at the same time we have brought ourselves a large inefficiency burden in the form of excessive traffic congestion.

INCENTIVES TO COORDINATE DECENTRALIZED TRANSIT CHOICES

Urban transit choices are made by individuals on a decentralized basis, just as choices in our market economy generally are made on a decentralized basis. Complicated interdependencies occur in the urban transit setting, however, and they affect both the demand and the supply of transit in complicated ways. These interdependencies alone would make price-setting for decentralized coordination decisions (about what transit modes to use and, thus, about how to allocate resources) very difficult because they can prevent some persons from facing the full cost consequences of their actions. But, in addition, prices cannot serve their familiar purpose in the transit setting because the relatively high cost of actually carrying out the necessary transactions prohibits the use of prices to ration road space. As a result much road use is not properly priced and road use is inefficient; we usually have more congestion than is optimal. Moreover, because of the interdependence of automobiles with other transit modes, this failure to coordinate one aspect of road usage—if uncorrected—can have serious perverse effects on other transit modes.

Pricing of Road Use

Various forms of user taxation—primarily fuel taxes, excise taxes, license fees, parking fees, and tolls (road, bridge, and tunnel)—combine along with other costs of vehicle operation and the time cost of traffic congestion delays to form a crude price for road usage. Where there is little or no congestion most of the time on rural roads and late at night or early in the morning on urban roads, taxes and tolls are not appropriate and they actually will reduce welfare. But they can reduce congestion and invite more efficient road use where there is congestion during peak usage periods in urban areas. Of course such taxes cannot be effective at one time of day without being effective at another. The failure of such taxes to distinguish peak from off-peak time periods may not be very serious, however (8). A greater problem seems to be due to the currently low level of user taxes where congestion is most serious. Fuel taxes, for example, may be at less than half the level needed to ensure efficient use of roads in large urban centers in the United States (9). As a result, excessively high congestion can prevent the most efficient use of the road. A crude way to improve coordination of transit choices is to raise substantially the tax on inputs necessary for road use.

As input taxes go up to ration the use of existing roads and thereby improve the efficiency of road use, a problem arises with respect to the relation between revenues from bus operations and the total cost of those operations. We have already noted that, if we apply a percentage tax on inputs, the average cost of bus operations will tend to go up at the same rate as average cost on automobile operations. The contributions these 2 modes make to traffic congestion are not necessarily in those proportions, however, for automobiles contribute more per passenger-mile to congestion than buses do. Relative to the effect of automobile usage on marginal social cost, the increase in marginal social cost due to greater bus usage will be less than the initial ratio of average bus cost per passenger-mile to average automobile cost per passenger-mile. That means that relative to the ideal price for automobile travel, the ideal price for bus travel will go up by less than the ratio of bus average cost to automobile average cost, and yet the tax will raise both costs by the same percentage. So if we rely on input taxes for achieving an effective automobile travel price, we shall tend to have a higher effective price for bus transit than would be ideal.

It is possible to correct this probable bias of an input tax against buses in several different ways. One way is to subsidize bus operations because, if there is any profit from bus operations, the price for bus transit probably is too high relative to an ideal effective price for automobile travel. For, if the bus travel price is near its optimum level, losses are likely, and a subsidy will be needed to cover the total cost of providing the bus service. An alternative arrangement would reduce the cost interdependency brought on by automobiles and buses contributing to each other's congestion, perhaps by providing exclusive lanes for bus travel. Then a marginal automobile passenger-mile would be less apt to add to the congestion delays felt by bus passengers and vice versa. But even without congestion interdependence—automobile passengers slowing bus passengers and vice versa—there may be reason to price bus transit below its average cost when input taxes are in effect to control congestion. For the automobile still can contribute more to its own congestion and, thus, for efficiency can warrant a higher congestion toll than buses. Yet because the tax on inputs cannot easily be adjusted based on how the inputs are finally used, a tax level that will control automobile usage ideally can make the effective tax on bus travel too high. A subsidy can lower the levy per bus-mile then, however, and bring the effective price for bus travel more in line with its marginal social cost.

New Jersey, Pennsylvania, and on a less complete scale several other states now provide significant operating subsidies to urban transit. There also is systematic relief for publicly owned transit operators from the federal tax on gasoline and from the federal excise taxes on buses and parts. And privately owned transit operators have been granted relief from fuel tax increases that were approved in 1956 under the federal highway program (10). All of these efforts to ease the impact of fuel taxes on urban transit operators probably are consistent with efforts to maximize welfare. Their effectiveness is limited, of course, where fuel taxes tend to be too low to control urban traffic congestion adequately.

Several cities are now turning to transit subsidies (11). Atlanta's reduction in fare from 40 to 15 cents already has increased bus patronage by more than 20 percent, and the adjustment to the new fare may not yet be complete. The service also is being improved by a rapid transit system that is being financed in part by a 1 percent city sales tax. Denver voters passed a $4 million bond issue in order to reduce all bus fares, including a lower off-peak bus fare. And Akron, Dayton, and Toledo have recently approved property tax increases in order to fund transit improvements.

By lowering the price of bus usage more in line with that of automobile usage, these subsidy efforts will be able to reduce the bias that otherwise would favor automobile travel. And because automobile travelers cause more congestion, that will be somewhat relieved. All road travelers typically still will pay less than the marginal social cost of their travel, however, so neither their payments nor the level of congestion they experience will serve as a reliable indication of the value they actually place on road capacity. Only an increase in input taxes, or some other means of collecting a fee for road use, can offer such an indication reliably.

Quantity of Roads

Proceeds from user fees now seem adequate on average to meet the costs of highway construction, and that fact may suggest to some that road-user fees already are high enough. But equality in these magnitudes is not necessarily in order. A user fee, which could be an input tax, serves first the purpose of rationing existing roads so they may be used efficiently. The level of fees or taxes that accomplishes that result also can indicate whether the existing stock of roads should be increased. For instance, in paying their fees, users may reveal a willingness to pay more than the cost involved in expending roads, and then roads should be expanded. This expansion question is a difficult one, though, as we have noted. In urban areas especially, expansion of the highway system is apt to encounter diminishing returns, and an optimal price for road usage then should provide a surplus of revenue over past road construction cost to go as a rent to the advantageously located but scarce land already devoted to road use. Because this land is publicly owned, all citizens reasonably can claim a share in such rents. And so some of the payments by users may go appropriately as nonmarginal transfers to other users and nonusers alike, rather than as investments in new highways. Fees and taxes that make the effective price paid for highway usage equal to the marginal social cost of highway usage will not necessarily also make the total proceeds from fees and taxes equal to construction costs because the long-run average and marginal road construction costs will not necessarily be equal. A surplus of user fees over highway construction expenditures can be perfectly appropriate, especially in cities where land is valuable.

Suppose, as is true in U.S. metropolitan areas, that user fees are not great enough to make the usage of existing roads efficient during daytime hours. As a consequence there will be excessive congestion on roads. And the presence of serious congestion could lead to political pressure to build roads, even though the main reason for congestion is a faulty low effective price for road use (12). In connection with the Interstate highway program, we have an arrangement called the Highway Trust Fund, which urges that fees from users be used for construction of roads. Indeed, it offers to local communities the return of user tax proceeds only if the communities use them for roads. And each community need pay as little as 10 percent of its roads' costs; the balance comes from the Highway Trust Fund. All users already have contributed to this fund through federal gasoline taxes, but can benefit from it only by building highways in their communities. The trust fund arrangement thus systematically allocates enormous resources to highways by distorting extraordinarily the relative costs faced by local communities as they choose between roads and other modes of transit. Although perhaps justified temporarily during a one-time expansion of the national network of highways, the extreme distortion introduced in our long-run transit choices by such a scheme certainly should not be allowed to persist. As a result of such a distortion in relative prices, communities will be persuaded to build more roads when other uses of the resources actually would benefit the society more.

If input taxes are raised to a level that will reduce congestion in urban areas to an efficient level, there will be a surplus of current tax revenue over amortized road construction cost. All tax proceeds thus should not be spent on roads. On equity grounds the surplus of tax proceeds over road construction cost ought to go to highway users in a way that will not influence their marginal road use (as a reduction in input tax would do), and some of it reasonably could go to nonusers as well. After all, a rent for scarce public land can be shared by all citizens. In particular, those who have been discouraged from using the road—the tolled-off—can claim a share in these excess payments. And there is a perfectly sound equity argument for supporting hospitals with the surplus, for instance, because motor vehicle air pollution has been shown quite convincingly to aggravate bronchitis, emphysema, and lung cancer (13). Because input taxes affect bus travelers more than automobile travelers, relative to their respective contributions to congestion, some additional return also should go to bus travelers in some form of subsidy. Because many citizens benefit from local government expenditures, it seems perfectly appropriate to approximate nonmarginal transfers by using proceeds of these travel input taxes as local revenues. Doing so would even alleviate the inefficiencies associated with other taxes. If they do not trust their political institutions, automobile travelers may treat suspiciously any claim that the use of motor vehicle input taxes will permit a reduction in other local taxes. But apart from such distrust it is a perfectly reasonable solution.

INCENTIVES FOR TRANSIT AGENCY OPERATING EFFICIENCY

Setting the price for a transit service deliberately below its average cost by providing a subsidy might destroy any incentive for operating efficiency. So to set prices below average cost in order to coordinate better our decentralized transit choices, we may have to sacrifice the efficiency incentives that exist when an agency must seek a profit or even break even. On the other hand, the profit test may never have provided good incentives. To make a profit a firm might have cut costs not by reducing waste and operating inefficiencies but by terminating services that were socially the most desirable. The profit test serves in a well-functioning competitive market to distinguish efficient firms that satisfy consumers from inefficient ones that do not. The inefficient ones will be driven from the field, and this harsh penalty serves as an incentive to encourage efficient operation. When the full competitive market profit test is inappropriate, however, as in the urban transit setting, its spur to operating efficiency can be lost.

Government action to remedy an excess profit or a loss position in a transit agency cannot then accept as an entirely reliable indication of what is possible and efficient the cost that the transit agency reports. For example, if any loss, no matter how large, is to be made up automatically out of government funds, no incentive will remain for the transit agency to control its costs; and the agency's reported loss may be larger as a consequence. Other bases for determining a subsidy payment, such as an allowable return to private owners based on capital employed, suffer also from biases they may introduce in the choice of inputs, which in turn will increase costs. But different bases for subsidy payment are worth examining for their effect on efficiency incentives because alternative incentives for efficiency are needed when the profit test is not available.

Let us consider 4 main ways that a government may provide subsidy payments to a public or private transit agency: (a) make up any deficit, (b) share fractionally in any deficit, (c) allow a subsidy payment to be "earned" based on an input measure, and (d) allow a subsidy payment to be "earned" based on an output measure. Any one of these methods may provide subsidy funds to a transit agency that is deemed to warrant them, but the methods can have very different effects because of the opportunities and constraints they place on the managers of the transit agency.

Although it may offer some sort of relief in unusual circumstances, the first type, the make-up-any-deficit scheme, removes all incentive for operating efficiently. It can be complicated in practice by the organization of public agencies, for the deficit will depend on the scope of the operating agency and the costs that are included in its operations. Some agencies do not consider capital costs, for instance, and an additional

deficit would no doubt be incurred, at least implicitly, to cover the cost of such capital. Where capital is privately owned and a payment must be targeted for its use, we would have the third type of subsidy payment scheme, which we shall discuss in a moment.

Under the second basis for subsidy payments, deficits are shared fractionally by the transit agency, and the government subsidy source can preserve an incentive for efficiency; the transit agency will then enjoy a larger residual sum for it is more efficient (14). Such incentives are common in government contracting and seem useful, although they cannot be claimed as unqualified successes. In typical applications, a price and cost plan is established in advance for a particular project, and departures from the profit implicit in such a plan are shared by customer and vendor according to a preset formula. In a continuing service operation like transit, determining such profit targets for a privately owned transit agency would be difficult without moving to some extent into the third or fourth methods of subsidizing, where some reference to an input or an output is made. But the idea of sharing alone is a useful one and can salvage some efficiency incentive by making the agency suffer when its costs are higher.

The third type of subsidy is keyed to the usage of an input. Public utilities are regulated in the United States by rate-of-return constraint, which relies on the amount of one input, capital, to determine a profit figure that the firm is allowed to earn. In principle, this allows a subsidy to be earned per unit of the capital input, as the third subsidy method requires. In practice as it is applied to public utilities, this method is much like the first method, however, because, although the targeted profit depends on the capital input, that profit is then almost assured the utility, and cost increases lead merely to higher prices; very little incentive for efficiency remains. Indeed, the public utility firm can enhance its allowed profit by distorting its mixture of inputs away from the most efficient blend in favor of more capital on which its allowed profit depends (15). Because of long delays that occur in the administration of this regulatory scheme, its more perverse aspects may actually be muted, for a firm cannot win price increases promptly enough to make relief from financial penalties for inefficient operations immediately available. If no relief were granted and the firm were merely credited with payments based on the amount of capital (or other input) it employed, an incentive for keeping costs low would remain. Earning a subsidy payment based on usage of a particular input is almost certain to cause distortion in the mixture of inputs, however, and so it is undesirable for that reason. The urban transit capital grants have been shown convincingly to introduce a bias toward the use of capital, and thus to distort input mixtures away from the most efficient ones (16).

The earning of a subsidy payment per unit of output avoids any distortion away from efficient input mixtures and also can invite the expanded provision that typically is desired of public services (17, 18) A transit subsidy per passenger, or per passenger-mile, is therefore to be preferred over a subsidy based on any one input or on the size of the agency's loss. In the case of a privately owned transit agency, a subsidy per unit of passenger service can be effective not only because, with fixed price, it leaves cost reduction incentives in the firm but also because it urges expansion of service by the firm in order to win greater subsidy payments.

If ideal input taxes are in force, an ideal subsidy per bus passenger-mile can be estimated. The subsidy can be derived from an admittedly simple model of urban transit, and it takes the following form (7, p. 29):

$$S = M_A\left(\frac{AC_B}{AC_A} \cdot \frac{\partial AC_A}{\partial M_A} - \frac{\partial AC_A}{\partial M_B}\right) + M_B\left(\frac{AC_B}{AC_A} \cdot \frac{\partial AC_B}{\partial M_A} - \frac{\partial AC_B}{\partial M_B}\right)$$

To illustrate application of this formula, consider \$0.0002 a rough estimate of $\partial AC_A/\partial M_A$ or $\partial AC_B/\partial M_A$, where M_A and M_B represent trips that take roughly 1 hour [this crude estimate is based on relations given in the Highway Capacity Manual (19) and a time value of \$1/hour], and let $\partial AC_A/\partial M_B$ and $\partial AC_B/\partial M_B$ be about ⅓ as great (7). Then, if $AC_B = AC_A$ and $M_A + M_B = 800$ trips/hour, S = 10 cents/passenger trip. The appropriate value of S could vary for each bus route depending on traffic volume, the levels of average costs, and the effects of each mode on congestion costs.

Thus, when a profit test is foregone, as we have argued it should be with respect to

urban transit agencies, a subsidy must be given to the agency providing transit service, and efficiency incentives can be lost or distorted as a consequence. The most direct way to determine an appropriate amount of subsidy is to estimate effects per passenger or per passenger-mile based on differences between average and marginal costs due to bus and automobile passengers. The ideal price, including effects due to any input tax, can be determined that way, and implicit in an ideal price is a subsidy per unit of service. Because a subsidy per unit of service also can offer an effective incentive for efficiency, it is to be preferred over alternative methods of providing a subsidy.

CONCLUSION

It is doubtful that present transit choice opportunities in cities reflect marginal social costs well enough to enable us to coordinate our actions efficiently. In choosing to travel by automobile, we pay only the average rather than the marginal social cost of our trip, with the result that together we create more than an efficient amount of traffic congestion. Ideally we would employ devices, which only now are technically feasible, to record the presence of vehicles and charge optimal tolls by time and place in order to ration road space. But we have been unable politically to implement such congestion controls. Because bus passengers contribute less marginal road congestion than automobile passengers, we can relieve this congestion problem somewhat if we can lower the price of bus travel and at the same time offer a service that will persuade some travelers to use buses rather than automobiles.

But higher taxes should be imposed on the inputs needed for road use, too, in order to make the perceived private cost of all road use approach more closely its marginal social cost. And the proceeds from such road use taxes reasonably can subsidize a substitute bus service and still should yield net proceeds beyond the cost of constructing roads. There is no compelling logic to support break-even rules in the urban transit setting. Of course, an attempt to achieve efficient use of town roads is apt to require a subsidy to the provider of a bus or transit service, and yet it is difficult to provide a subsidy and still encourage efficient operation of the agency. An award per unit of service probably will preserve an internal efficiency incentive better than alternative bases for the subsidy payment, and it can provide a better incentive for welfare-maximizing actions than the profit goal can offer.

ACKNOWLEDGMENT

Helpful comments by Charles Hedges and Irving Hoch and financial support from the National Science Foundation to the Thomas Jefferson Center at the University of Virginia are gratefully acknowledged.

REFERENCES

1. Sherman, R. Congestion Interdependence and Urban Transit Fares. Econometrica, May 1971.
2. St. Clair, G. P. Congestion Tolls—An Engineer's Viewpoint. Highway Research Record 47, 1964, p. 75.
3. Moore, J. H. Congestion and Welfare—Comment. Economic Jour., March 1968.
4. Vickrey, W. Pricing as a Tool in Coordination of Local Transit. In Transportation Economics, Columbia Univ., New York, 1965.
5. Sherman, R. A Private Ownership Bias in Transit Choice. American Economic Review, Dec. 1967.
6. Sherman, R. Club Subscriptions for Public Transport. Jour. of Transport Economics and Policy, Sept. 1967.
7. Sherman, R. Subsidies to Relieve Urban Transit Congestion. Jour. of Transport Economics and Policy, Jan. 1972.
8. Mohring, H. Relation Between Optimum Congestion Tolls and Present Highway User Charges. Highway Research Record 47, 1964, pp. 1-14.
9. Walters, A. A. The Theory and Measurement of Private and Social Cost of Highway Congestion. Econometrica, Oct. 1961.

10. Feasibility of Federal Assistance for Urban Mass Transit Operating Costs. U.S. Department of Transportation, Nov. 1971.
11. The Christian Science Monitor. Dec. 28, 1972.
12. Meiburg, C. O. An Economic Analysis of Highway Services. Quarterly Jour. of Economics, Nov. 1963.
13. Lave, L. B., and Seskin, E. P. Air Pollution and Human Health. Science, Aug. 21, 1970.
14. Cross, J. G. Incentive Pricing and Utility Regulation. Quarterly Jour. of Economics, May 1970.
15. Averch, H., and Johnson, L. L. Behavior of the Firm Under Regulatory Constraint. American Economic Review, Dec. 1962.
16. Tye, W. B. The Economics of Urban Public Transit Capital Grants. Published in this Record.
17. Roth, G. The Regulation of Buses in Cities. Published in this Record.
18. Miller, D. R The Urban Transit Industry: Deficits and Subsidy Mechanisms. Paper presented at Engineering Foundation Conference on Public Transportation, Henniker, New Hampshire, July 14, 1972, and to be published by the Highway Research Board.
19. Highway Capacity Manual. U.S. Government Printing Office, Washington, D.C., 1950.