An Economic Analysis of Construction Bottlenecks

William P. O'Dea
Department of Economics and Business
SUNY-Oneonta
Oneonta, NY 13820
(607) 436-2127
ODEAWP@ONEONTA.EDU

ABSTRACT

Highway repaving and repairs often require that two (or more) lanes of traffic be condensed into one lane around construction sites. As a rule, the merging process is unmanaged which in peak periods results in a traffic queue at the bottleneck. Traffic moves through the queue in a stop-and-go manner which increases travel time. This paper computes the amount of time spent in the traffic queues which result when the behavior of drivers at bottlenecks is unmanaged. It then explores a strategy for imposing order on driver behavior and determines the reduction in queuing time that would result. An example shows that the value of the reduction could be dramatic.

INTRODUCTION

As drivers are well aware, it is frequently necessary to close highway lanes for resurfacing or repair. These closures can be a source of considerable annoyance to drivers. A number of papers have been written about the congestion problems created when one lane on a two-lane road is closed and vehicles traveling in both directions are forced to alternate their use of the open lane. See, for example, Son (1999). The focus of this paper is on the congestion that results when one lane (or perhaps more than one lane) on a limited access highway with multiple lanes in each direction must be closed. There is a substantial literature dealing with the economics of bottlenecks. For example, Mun (1994) considers the case of a roadway which contains a bottleneck whose capacity (measured in vehicles per hour) is less than the volume of traffic on the roadway which feeds into it. A queue then forms at the bottleneck entrance. Arnott, de Palma and Lindsey (1990) analyze the evolution of a traffic queue at a bottleneck in a situation where drivers incur penalties for arriving at their destinations either earlier or later than they would desire. In their model, drivers can reduce time spent in the queue at the cost of an increase in scheduling delay. Neither of these models considers the behavior of drivers within the queue in any detail. For example, Mun does not address the problems associated with the process of merging two lanes of traffic (in one direction) into one lane. His model would be applicable in a situation where two lanes of traffic feed into a road which also contains two lanes but which has narrower shoulders which reduce capacity. In this case, there would be no merging problem. In the Arnott *et al.* model, Verhoef (2003) points out that the queue of vehicles at the bottleneck takes up no physical space and that "vehicles leave the queue on a first-in-first-out basis (p. 534)." In effect, each vehicle joining the queue is added to a vertical stack. Since drivers are not in competition for access to the bottleneck, the capacity of the bottleneck is independent of the size of the queue of vehicles waiting to enter. At construction sites on multi-lane highways, the need to condense two (or more) lanes of traffic into one lane is at the heart of the congestion problem. Verhoef (2003) analyzes a situation where a bottleneck is created by a decrease in the number of lanes from two to one. However, he assumes that the merging process is "smooth" and that cars from the right and left lanes enter the bottleneck in an alternating fashion (p.540). On actual highways, the merging process tends not to be smooth and vehicles compete for access to the open lane. As we will see below, there is evidence that this competition reduces the rate at which vehicles can exit the bottleneck. During peak periods, the reduction in capacity can result in a significant increase in travel time.

The objective of this paper is to offer a solution to this problem. We will proceed as follows. In section II, we present a description of the travel problem and show that the failure of drivers to coordinate their behavior lies at the heart of the congestion problem. In section III, we derive an expression that can be used to determine the total amount of time drivers spend in the queue. In section IV, we present a solution to the problem, which is a strategy for reducing the scope for strategic behavior and thus creating a more orderly traffic flow. The conclusion discusses how the simple situation considered in this paper can be made more realistic.

DESCRIPTION OF THE PROBLEM

Drivers

We assume that the traffic flow consists entirely of passenger cars, that each car is occupied only by its driver, and that all drivers place the same value on time spent in the traffic queue. Determining the amount of money each driver would be willing to pay to reduce time spent in the queue by one hour is a straightforward process. Button's (1993) survey of empirical estimates of the value of travel time shows that the price that drivers would be willing to pay to save an hour of travel time ranges from 12 percent of the wage rate to 145 percent of the wage rate. There is also evidence that higher income drivers are willing to pay more. Following Mohring (1999), we assume that the value of a travel time saving is equal to half of the wage rate. In 2003, the average hourly wage rate in the private sector was about $15.35 (Council of Economic Advisers, 2005). The value of a reduction in travel time would then be $7.68 per hour. However, research by Small (1982) indicates that drivers are willing to pay more to avoid an hour of driving under congested conditions than under free-flow conditions. Based on their analysis of the behavior of commuters in Sydney, Australia, Hensher *et al.* (1990) find that the ratio of the value of time spent under congested conditions to time spent under free-flow conditions is 1.7 to 1. Applying this ratio to the time value above implies that each driver would be willing to pay $13.05 to reduce time spent in congested conditions by one hour. This value, which we will use in our examples, is close to the value of $12.50 used by Mohring (1999).

Each driver also wishes to travel at a safe distance behind the preceding car. We assume that all drivers have the same conservative notion of what constitutes a "safe" following distance. They wish to be able to avoid a rear end collision in the event that the preceding car stops instantaneously. Given this standard, Pacquette *et al.* (1972) derive an equation which shows that the space occupied by a car depends on the car's length, the driver's reaction time, speed and the car's breaking capacity:

$$SP = L + 1.4666 * b * S + \frac{S^2}{30 * f} \tag{1}$$

where L is the length of the car (in feet), b is driver reaction time (in seconds), S is speed (in miles per hour), and f is the coefficient of friction between the tires and the roadway. The coefficient of friction is a function of speed. As speed increases, the coefficient decreases in an approximately linear manner (Pacquette *et al.*, 1972). At 45 miles per hour (mph), which is the commonly posted speed limit in highway work zones in the United States, f would be approximately .89 on dry pavement. At 65 mph, f would be .84 on dry pavement. Given the amount of space occupied by one car, it is possible to compute traffic density (cars per lane-mile) and the traffic flow (cars per lane-mile per hour).[1]

In our analysis, drivers incur no penalties for arriving at their destinations earlier or later than they would desire. As noted, these penalties are an important element of the congestion model developed by Arnott *et al.* (1990). Their model considers the behavior of a group of drivers trying to get to work. If commuters arrive at work late, they suffer a loss of income. If they arrive early, they spend time at the work site that might be more pleasantly spent at home. Since the trip to work is made repetitively, it is reasonable to expect that drivers will devise strategies that minimize their travel costs. In this paper, we focus on the behavior of drivers making long, non-work related intercity trips. These trips are made over roads with which the driver might not be familiar. Thus, by their nature, these trips cannot be scheduled very precisely. Even if drivers are aware that they will encounter construction delays, the location and severity of these delays are generally not known with precision. For example, the AAA's Trip-Tiks do provide a rough indication of the location of construction sites but do not suggest an alternative route. In effect, we assume that drivers view construction delays on intercity trips as random bad luck and do not find it worthwhile to devise a strategy to deal with the problem.

Finally, we assume that all drivers travel at the same speed and that when both lanes are open to traffic the traffic flow is divided evenly between the left- and right-hand lanes.

The Road

Under normal conditions, drivers have two lanes (in each direction) available to them. However, as shown in figure 1, a segment of the right lane is closed to traffic for an extended period of time. In the transition area, traffic is channeled from the closed lane into the open lane. The closed section does not contain an entry or an exit ramp. In the work zone, the posted speed limit is 45 mph. Thus, prior to entering the work zone, cars A, B, C and D in figure 1 must reduce speed to 45 mph. In addition, vehicles B and D must merge into the left lane.[2]

Ideally, drivers would like to execute both of the maneuvers without a traffic queue developing. Whether this is possible depends on the relation between the capacity of the work zone (i.e. the rate at which cars can enter the work zone) and the flow of traffic on the highway. If the traffic flow exceeds capacity, then the transition area becomes a bottleneck and a queue will develop. Traffic queues impose a number of costs on drivers. The most obvious is that they increase travel time. Since queued traffic moves in a stop-and-go manner, fuel consumption, vehicle wear and tear and the emission of pollutants will also increase. There is also evidence that work zone queues cause an increase in the rate of accidents (Maze *et al.,* 1999). In this paper, we focus on the time cost of traffic queues. Jiang (1999b) estimated the user costs of lane closures on I-65 in Indiana and found that queuing delay costs represented over 90 percent of total costs.

The theoretical capacity of the open lane can be found using equation (1). Assuming a vehicle length of 16 feet and a reaction time of .4 second, equation (1) implies that the distance between successive cars (front bumper to front bumper) at 45 mph would be about 118 feet.[3] Drivers would probably not want to travel this closely together for any length of time, since the required level of attention would be difficult to maintain. However, the trip through the construction zone is a temporary event. Moreover, the physical set up of construction zones often includes construction barriers on both sides of the road and narrower (or no) shoulders, which force drivers to pay attention. In addition, since passing is not permitted in work zones, drivers don't have to worry about other drivers cutting in front of them. With a space requirement of 118 feet per car, traffic density would be 44.74 cars per lane mile and the traffic flow would be about 2000 cars per lane mile per hour. Of course the theoretical capacity of the open lane depends on a number of site specific considerations including: the width of shoulders, the gradient and alignment of the road in the work zone, and the proximity of construction workers and equipment (Dudek and Richards, 1982). For long-term construction projects, the ideal would be to segregate workers from the traffic flow and to design barriers and shoulders to minimize driver feelings of claustrophobia.

In theory, as long as the traffic flow is less than or equal to 2000 cars per hour, cars should be able to move through the construction zone without a queue developing. For example, assume that the traffic flow is 2000 cars per hour evenly divided between the two lanes and staggered as shown in figure 1. At 65 mph, traffic density would be 15.38 cars per lane-mile, which implies that the space for each car would be 343.2' (d_1 in figure 1). Assuming a reaction time of 1 second, equation (1) implies that each car would require 283' of space. While 343.2' is more than enough to provide an adequate safety margin under normal conditions, it does not provide enough space to allow B to merge into the left lane without forcing C to slow down. If each car were to attempt to maintain a speed of 65 mph until the last possible moment before slowing down to enter the work zone, cars in the right lane would not be able to move into the left lane without triggering a cascade of speed reductions that would transmit a shock wave upstream. However, consider a scenario in which a sign, located 1000' upstream of the transition area, instructs all drivers to reduce their speed to 45 mph immediately and tells drivers in the right lane to merge left when it is safe to do so. By the time car A is 500' from the transition area, all four cars will be traveling at 45 mph. In addition, d_1 will be 237.6' and d_2 (the distance between A and B) will be 118.8'. Thus, B would be able to merge into the left lane without impinging on C. Unfortunately, it is hard to envision that actual drivers would behave in such a highly coordinated fashion. Not all drivers will react to the sign equally quickly or brake at the same rate. Some drivers might be tempted to cheat in order to save time. For example, B might maintain speed in an attempt to pass A. An officer in Wisconsin's State Highway Patrol suggests that it is not in the nature of American drivers to slow down unless they are forced to do so by a traffic tie-up (Wald, 1999).

There is evidence that the orderly flow of traffic will break down well before theoretical capacity is reached. For example, Jiang (1999b) defines work zone capacity as "the traffic flow just before a sharp speed drop followed by a sustained period of low vehicle speed...and a long period of traffic congestion (p. 7)." In essence, Jiang defines capacity as the maximum flow that can be sustained without a traffic queue forming. His analysis of four work zones in Indiana indicates that capacity is about 1600 cars per hour (Jiang 1999a). Dixon *et al.* (1996) employ a similar definition of capacity and find that on rural segments of I-95 in North Carolina where two lanes of traffic are forced to merge into one open lane the capacity of the open lane is approximately 1450 cars per hour. Interestingly, they report that on interstate work zones located in urban areas the capacity of the open lane is much higher, between 1696 and 1873 cars per hour. They account for the difference by pointing out that while I-95 "serves a high proportion of unfamiliar through drivers (p. 30)" its urban segments are used by commuters who, because they make the same trips repetitively, are more familiar with the traffic situation. [For a theoretical analysis of the development of instability in the traffic flow, see Ferrari (1991).]

Importantly, the onset of congestion and queuing is accompanied by a discontinuous drop in the rate at which cars are discharged from the queue into the open lane. This phenomenon is analyzed in two papers by Banks (1990, 1991). Jiang (1999a) reports that at the Indiana work zones he studied the mean queue discharge rate was approximately 1400 cars per hour.

It is reasonable to ask what causes this abrupt capacity drop. Banks (1991) suggests that "merge conflicts ...lead to the general flow breakdown" when the merging process is uncontrolled (p. 89). This hypothesis is plausible. Consider a relatively benign merger protocol in which each car in the left lane allows a car from the right lane to cut in before entering the work zone. With this arrangement, the flow of traffic into the open lane would not be continuous because there would be small time intervals in which traffic in neither lane would be moving. At actual construction sites, the merger process can be much more confrontational. Maze *et al.* (1999) observe that some drivers will attempt to travel in the closed lane as long as possible before merging into the open lane. They point out that drivers who have been waiting patiently in the queue for their turn to enter the bottleneck have been known to straddle both lanes to discourage such behavior. Thus, cars traveling in the closed lane might have to force their way into the open lane. In addition to reducing the capacity of the open lane, such behavior also contributes to the increased accident rate at work sites mentioned above. In addition, some drivers will switch lanes several times before reaching the bottleneck in an effort to be in whichever lane of traffic is moving. Each lane switch uses up capacity and doubtless contributes to the frustration level of drivers in the queue.

There is evidence to support the hypothesis that the abrupt drop in capacity is caused by uncontrolled merging. For example, at work zones in Texas, Dudek and Richards (1982) report that when two lanes of traffic are forced to merge into one open lane, the capacity of the open lane (what Jiang would term the queue discharge rate) is 1340 vehicles per hour. In situations where three lanes of traffic are forced to merge into one lane, the capacity of the open lane is only 1130 vehicles per hour. A reasonable explanation for the decrease is that the merging problem would be more complicated in the second case. Additional support for the hypothesis comes from Pavis *et al.* (1995) who analyze a bus metering scheme at the Lincoln Tunnel in New York City. They study a situation where a lane of cars and a lane of buses have to merge into one lane. They report that when merging was unmanaged the traffic flow was "turbulent" and that the average throughput was 1050 vehicles per hour (350 buses, 700 cars) (p. 35). A metering scheme in which two cars were allowed to enter for each bus was imposed. Pavis *et al.* indicate that metering created a "more uniform flow into the tunnel" and increased the entry rate to approximately 1200 vehicles per hour (p. 37).

While individual drivers might perceive it to be in their interest to behave strategically, this behavior reduces the capacity of the open lane and therefore increases travel time for all drivers. Economic theory suggests that drivers do not pay attention to the external costs of their actions, which leads to a suboptimal outcome. The implication is that controlling the merging process can reduce the scope for strategic behavior and thus increase the capacity of the open lane. In section IV, we consider a possible method the highway authority can use to achieve this objective. Of course, the basic economic question is whether the benefits of achieving a more orderly traffic flow are likely to exceed the costs. Therefore, in the next section, we will present a method of computing the total amount of time that drivers spend in the traffic queue.

DETERMINATION OF TIME SPENT IN THE QUEUE

We use ER to represent the rate at which cars can enter the bottleneck under uncontrolled conditions. We assume that the travel day contains a period T_1 hours in length when the traffic flow on the highway exceeds ER. After this period, we assume that the traffic flow on the road and thus the rate at which cars arrive at the bottleneck is less than ER. We use AR_1 and AR_2 to represent the rates at which cars arrive at the bottleneck during these periods ($AR_1 > ER > AR_2$). To simplify the analysis, we assume that AR_1, AR_2, and ER are constant. The relation between the arrival rates and the entry rate is shown in figure 2.

When the arrival rate exceeds ER, the queue gets steadily larger increasing from 0 at time t_0 to its maximum length at time t_1. After t_1, the rate at which new cars join the queue is less than the rate at which cars enter the bottleneck. The queue then starts to dissipate and will eventually disappear.

As McShane *et al.* (1998) and Mun (1994) point out, the Lighthill and Whitham shock wave equation can be used to determine the rate at which the edge of the queue moves upstream during the buildup period:

$$S_B = \frac{AR_1 - ER}{D_1 - D_J} \tag{2}$$

where D_1 is traffic density (cars per mile) under free-flow conditions and D_J is jam density in the traffic queue. Since AR_1 is greater than ER and D_1 is less than D_J, S_B is negative. Since AR_1, ER, D_1 and D_J are constant, S_B is constant, which means that during the buildup period the edge of the queue will move upstream at a constant rate. We assume that the queue will occupy both lanes upstream of the transition area.

Equation (2) enables us to determine the maximum length of the queue (in cars) at t_1:

$$Q_{\mathrm{MAX}} = - S_B * T_1 * D_J \,. \tag{3}$$

Given (3), the maximum wait in the queue (for the car that enters the queue at time t_1) is:

$$TQ_{\mathrm{MAX}} = \frac{- S_B * T_1 * D_J}{ER} \tag{4}$$

Since the length of the queue increases in a linear fashion from 0 at time t_0 to its maximum length at t_1, the average driver's wait in the queue will be:

$$AvTQ = \frac{- S_B * T_1 * D_J}{2 * ER} \tag{5}$$

After t_1, the arrival rate is less than ER and the queue starts to dissipate. The rate at which the queue shrinks is given by:

$$S_D = \frac{ER - AR_2}{D_J - D_2} \tag{6}$$

Since ER exceeds AR_2 and D_J is greater than D_2, S_D is positive which means that during the second period the edge of the queue is moving closer to the work zone. Given our assumptions, S_D is also constant, which means that during the second period the queue will decrease at a constant rate. This implies that for drivers who join the queue after t_1 the average wait in the queue will also be given by equation (5).

The total amount of time spent in the queue can be found by multiplying the average time spent in the queue by the number of drivers entering the queue during the two periods. The number of drivers entering the queue during the buildup period is:

$$T_1 * AR_1. \tag{7}$$

The number of drivers during the second period is the length of the second period multiplied by AR_2. To find the length of the second period (T_2), it is first necessary to determine how long it will take for the queue to disappear:

$$T_2 = \frac{-S_B * T_1}{S_D} \tag{8}$$

The numerator of (8) is the length of the queue in miles.

Using equations (5), (7) and (8), we can then determine the total amount of time that drivers spend in the queue:

$$TTQ = [\frac{-S_B * T_1 * D_J}{2 * ER}][T_1 * AR_1 + \frac{-S_B * T_1 * AR_2}{S_D}] \tag{9}$$

Substituting for S_B and S_D using equations (3) and (6), equation (9) can be rewritten as:

$$TTQ = k * \frac{AR_1}{ER} * \frac{AR_1 - ER}{D_J - D_1} + k * \frac{AR_2}{ER} * [\frac{AR_1 - ER}{D_1 - D_J}]^2 * \frac{D_J - D_2}{ER_1 - AR_2} \tag{10}$$

where k is $(T_1^2 * D_J)/2$. The first and second order partial derivatives of (10) with respect to AR_1 are both positive which means that an increase in the arrival rate during period 1 will cause the time delay incurred by drivers in the queue to increase at an increasing rate. The first order partial derivative with respect to ER is negative while the second order partial is positive which means that an increase in the entry rate causes TTQ to decrease at an increasing rate. Multiplying TTQ by $13.05 yields the value of time spent in the queue.

The only variable in (10) which is under the direct control of the highway authority is the entry rate. As a measure of the magnitude of the benefit of an increase in ER, we use the elasticity of TTQ with respect to ER:

$$e = \frac{\partial TTQ}{\partial ER} * \frac{ER}{TTQ}.$$

To give an idea of the magnitudes involved, Table 1 provides a numerical example. Table 1 assumes that T_1 is 2 hours, AR_1 is 2000 (cars per hour), AR_2 is 1000 cars and that ER is 1400 cars per hour. The entry rate is based on Jiang's (1999a) findings. In the traffic queue, we assume that each car occupies 30'. Therefore, the jam density is 176 cars per lane mile. The first column of the table shows that the maximum length of the queue is 1315 cars which implies that the driver entering the queue at time t_1 would encounter the edge of the queue 3.73 miles from the work zone. It is interesting to note that the length of the period during which the queue dissipates exceeds the length of the buildup period. The value of the elasticity coefficient implies that a one percent increase in the entry rate would cause total delay time to decrease by 5.9 percent. The second two columns examine the impact of an increase in the entry rate. As the elasticity coefficient implies, increases in the entry rate would lead to more than proportional decreases in delay time. The second two columns show this to be case. An increase in the entry rate from 1400 to 1800 cars per hour, an increase of 28 percent, causes total delay time to fall by about 2800 hours, a decrease of about 83 percent. The value of the time savings is approximately $36,000. An increase in ER reduces total delay by reducing the maximum length of the queue, which in turn reduces the average delay. Since the maximum length of the queue is reduced and the entry rate is higher, the length of the second period is drastically reduced, which means that fewer drivers are caught in the queue after time t_1. In this example, the length of the second period is reduced by about 2.5 hours. Table 2 shows the impact of an increase in AR_1 to 2200 cars per hour. The table shows that an increase in the arrival rate during the first period produces a disproportionate increase in the delay cost. Given the higher value of AR_1, the benefits of an increase in the entry rate are higher. For example, an increase in the entry to 1800 cars per hour reduces delay cost by about $53,000.

SOLUTION

The examples presented in tables 1 and 2 show that an increase in ER can yield a sizable benefit. This then raises the question of what policies the highway authority can implement to increase ER. The least cost solution would be to require drivers to slow down to 45 mph and merge left in an orderly manner as outlined in section II. Drivers who are inclined to cheat could be discouraged by fines. The problem is that in construction zones it is difficult for police to find a safe space to detect speeders and pull them over (Wald, 1999). In theory, it would be possible to install an automated detection system which would photograph the license plates of offenders and send them tickets in the mail. However, this system, which is used in Europe and other places, is not widely used in the US. In addition, intercity highways are used by drivers from a variety of states and Canadian provinces. There is some question as to whether the prospect of getting a ticket in the mail from New York would encourage a driver from Ontario to slow down. If relatively few cheaters could be ticketed, then in order to discourage cheating the fines levied on those few would probably have to be quite large.

Since the ideal solution is not practical, we propose that prior to the transition area a barrier be erected between the two lanes which would prevent drivers from changing lanes. This barrier would prevent cars B and D in figure 1 from moving into the left lane. (The length of the barrier will be discussed below.) The barrier would not have to be elaborate or continuous. It just has to convince any driver contemplating a lane switch that s/he is running the risk of incurring a sizable bill at a body or muffler shop. Data provided by Richards and Dudek (1984) indicate that the cost of dividing the two lanes using traffic tubes would be on the order of $100 per day per mile. Just before the transition area, a traffic signal would be placed in front of each lane to control the entry of cars into the open lane. The cost of renting two of ADDCO’s portable solar traffic signals would be approximately $400 per day (Mueller, 2005). An electronic message sign located upstream of car D would inform drivers about the presence of a construction zone and tell them to stay in lane and be prepared to stop. The daily rental cost of the sign would be about $100. Small (1983) describes a signaling system for a bus priority lane that could be adapted to our purposes. During periods when the traffic flow is less than 1400 vehicles per hour, the sign could direct drivers to merge left. The light at the end of the left lane could be set on yellow. Drivers who remain in the right lane would face a flashing red light at the end of the lane, which would force them to wait for a safe gap in the traffic flow before entering the transition area.

This scheme prevents drivers from behaving strategically and ensures that they approach and enter the bottleneck in an orderly fashion. Cars from the two lanes would be allowed to enter the open lane in alternating platoons. In making the transition between lanes, a certain amount of time is lost. When the light changes to green, there is a delay while the first driver reacts to the signal change and begins to accelerate. The reaction and acceleration times of the second driver are faster, because they overlap with those of the first driver (McShane *et al.*, 1998). As more drivers enter the bottleneck, the time gap between entries stabilizes. At 45 mph, given our assumptions, cars can enter the bottleneck at the rate of one car every 1.8 seconds. The headways between the first 3 or 4 cars will exceed 1.8 seconds. In the transportation engineering literature, one approach to dealing with this problem is to assume that the headway between cars is constant but that 3 seconds are "lost"-i.e. during the first three seconds after the light change no cars enter the bottleneck and after that cars enter at a constant rate (McShane *et al.*, 1998). In addition, there is a transition period when both lights are red ("the all red phase") to minimize the possibility that cars from both lanes will attempt to enter simultaneously. We assume that this period is also 3 seconds. Thus, during each interval, 6 seconds are lost. For example, if the interval length ("cycle length") were 60 seconds, a platoon of 30 cars could enter during the remaining 54 seconds. Since there are 60 intervals in an hour, the hourly entry rate would be 1800 cars. While this is less than the theoretical capacity rate of 2000 cars, it is a considerable improvement on the unmanaged entry rate.

As table 1 shows, the maximum queue length would be reduced to 438 cars (2.48 miles). The length of the maximum queue would govern the length of the barrier between the lanes and the location of warning sign. In the simple situation we are considering where the arrival flows and entry rate are constant; the length of the barrier would be 1.24 miles. In a more realistic situation where the arrival rate and entry rate are stochastic, the length of the barrier would have to be longer to allow for times when the arrival rate during period 1 exceeds its expected value or the entry rate is less than its expected value.

The 6 seconds that are lost are analogous to a fixed cost. With a longer cycle length, the time lost would be spread over more cars and the entry rate would increase. For example, if the cycle length were 120 seconds, the hourly entry rate would increase to 1900 vehicles. The longer segment length increases capacity by reducing the number of gaps in the traffic flow. In a more realistic situation where the traffic flow includes trucks, whose performance characteristics are generally worse than those of cars, the number of gaps in the traffic flow might be important if the bottleneck includes an upgrade. If a truck in one platoon were to slow down, the gap between platoons would prevent the slowdown from affecting following platoons.

While increasing the cycle length to 120 seconds would yield an objective improvement as shown in tables 1 and 2, there is some question whether drivers would perceive a psychological improvement. A segment length of 120 seconds means that drivers will be motionless for two minutes at a time. There is evidence that travelers dislike sitting still. For example, Quarmby (1967) finds that urban commuters in the United Kingdom are willing to pay a much higher price to reduce time spent waiting for a bus or subway car than to reduce time spent in transit.

The scheme works because it imposes order on the behavior of drivers. There are parallels. For example, in theory superconductive transmission wires can carry a greater flow of electricity because the reduction of resistance prevents electrons from interfering with each other. Greenberg and Daou (1960) demonstrate that allowing cars to enter the Holland tunnel in platoons rather than individually eliminates shock waves within the tunnel and thus increases traffic flow. There is sizable literature, of which the Pavis *et al.* (1995) study cited above is but one example, that metering entry rates into highways can increase highway capacity and reduce travel times by creating a more orderly traffic flow. One question is whether drivers would be willing to cooperate with the scheme. Pavis *et al.* write that initially car drivers attempted to violate the metering protocol but were forced to cooperate when bus drivers exercised their rights to enter. The evidence presented in tables 1 and 2 indicates that the benefits of the scheme are sizable enough to cover the cost of stationing a trooper at the transition area during peak periods to ticket violators.

CONCLUDING COMMENTS

In the context of the stylized transportation problem considered in this paper, we have shown that a sizable benefit can be obtained at a relatively modest cost by controlling the merging process at construction bottlenecks on multi-lane highways. The actual benefits are probably larger than our results indicate. The benefits of a smoother traffic flow would also include reductions in pollution emissions and accident rates. Dougherty (1997) indicates that if a car expects to idle for more than 20 seconds it makes sense to turn off the engine since the fuel savings and reduced pollution emissions exceed the cost of restarting the engine. In our scheme, this threshold is easily met since vehicles will be motionless for (at least) a minute at a time. In addition, some vehicles will have more than one occupant. One can only wonder how much a couple traveling with a carload of children would be willing to pay to avoid a lengthy wait in a traffic queue. While a simplified set of circumstances are useful for establishing the value of an idea, there are a number of interesting ways in which the analysis can be made more realistic.

For example, on real highways, the traffic flow in the right lane generally exceeds the flow in the left lane. This imbalance is important because it influences the length of the barrier and the placement of the warning sign. It also influences the management of the traffic signals. Assuming no lane switching, the queue in the left lane would be shorter than the queue in the right lane. Thus, if the duration of the alternating entry intervals were the same for both lanes, cars in the left lane would have a shorter wait. This knowledge would give at least some drivers in the right lane an incentive to switch lanes which is the sort of behavior our proposal is intended to discourage. One method of dealing with this problem is that the signals could be managed so that there would be no competitive advantage from switching lanes. In other words, the average wait would be the same for drivers in both lanes. An alternative strategy would be to use an electronic sign to assign drivers to either the left or right lane and thus equalize queue lengths.

The analysis could also be adapted to include trucks in the traffic flow. The presence of trucks would alter my computation of the cost of the delays caused by the queue. It is reasonable to assume that the value of an hour of delay would be higher for a truck than for a car, since the computation would be based on the full wage of the driver and might include costs associated with the late arrival of the truck's cargo. The presence of trucks would also influence the management of the signal and thus the rate at which vehicles could enter the bottleneck. For example, if the bottleneck contained a steep upgrade, a truck might be forced to reduce speed. This speed change would send a shock wave upstream and could cause a traffic tie-up within the bottleneck. With our scheme, the intent is that vehicles wait to enter the bottleneck but move at a constant speed once they do. With trucks, it might be necessary to increase the length of time both lights are red. This would introduce longer gaps between the platoons. Longer gaps would reduce the risk of a traffic tie-up but would also reduce the entry rate.

Because our scheme is not price-based, it does not allow for differences between the willingness of drivers to pay for a reduction in queuing time. However, since an increasing number of vehicles are equipped with transponders similar to those used by the EZ-Pass toll collection system, adapting our scheme for differences in driver willingness to pay would be relatively easy. The metering strategy could be altered so that drivers in the left lane would have a shorter wait to enter the transition area but would have to pay for the privilege.

Finally, our analysis assumes that the arrival and entry rates are constant. In reality, these rates are random variables. Therefore, it would be an interesting exercise to adapt the analysis to take the stochastic nature of vehicle flows into account.

NOTES

1. Traffic density would equal 5280 divided by the amount of space per car given by equation (1). The traffic flow would equal traffic density multiplied by speed.

2. Figure 1 depicts a partial lane closure. The same basic analysis would also apply to a crossover situation in which all lanes on one side of a divided highway are closed and drivers traveling in both directions use the open side.

3. Estimates of driver reaction time vary widely (Drew, 1968). The engineers at Consumers Union place driver reaction time at .2 of a second (Caruso, 1998). Paquette *et al.* (1972) assume a reaction time of 2.5 seconds to allow for drivers whose reaction times are above normal. Following Edie (1974), we assume a reaction time of one second under normal driving conditions. However, for reasons discussed in the text, we assume that drivers are more alert in construction zones. Thus, a shorter reaction time seems appropriate.

REFERENCES

Arnott, R., A. de Palma, and R. Lindsey. 1990. Economics of a bottleneck. *Journal of Urban*

Economics 27 (1): 111-130.

Banks, J. H. 1990. Flow processes at a freeway bottleneck. *Transportation Research Record 1287:* 20-28.

Banks, J. H. 1991. Two-capacity phenomenon at freeway bottlenecks: a basis for ramp metering. *Transportation Research Record 1320*: 83-90.

Button, K. J. 1993. *Transportation economics*, 2nd ed. Aldershot, U.K.: Edward Elgar.

Caruso, T. 1998. Personal correspondence.

Council of Economic Advisers. 2005. *Economic report of the President, 2005*. Washington, D.C.: U.S. Government Printing Office.

Dixon, K. K., J. E. Hummer, and A. R. Lorscheider. 1996. Capacity for North Carolina freeway work zones. *Transportation Research Record 1529*: 27-34.

Dougherty, M. 1997. A block queueing system for slow moving traffic. *Transportation Research* 2D (4): 259-270.

Drew, D. R. 1968. *Traffic flow, theory and control.* New York: McGraw-Hill.

Dudek, C. L., and S. H. Richards. 1982. Traffic capacity through urban freeway work zones in Texas. *Transportation Research Record 869*: 14-18.

Edie, L. C. 1974. Flow theories. In *Traffic science*, ed. D. C. Gazis, 9-16. Wiley: New York.

Ferrari, P. 1991. The control of motorway reliability. *Transportation Research* 25A (6): 419-427.

Greenberg, H., and A. Daou. 1960. The control of traffic flow to increase the flow. *Operations Research* 8 (4): 524-532.

Hensher, D.A., F. W. Milthorpe, N.C. Smith, and P. O. Barnard. 1990. Urban tolled roads and the value of travel time savings. *The Economic Record* 66 (193): 146-156.

Jiang, Y. 1999a. A model for estimating excess user costs at highway work zones. *Transportation Research Record 1657*: 31-41.

Jiang, Y. 1999b. *Traffic characteristics and estimates of traffic delays and usser costs at Indiana freeway work zones.* West Lafayette, IN: Indiana Department of Transportation Research Division.

Maze, T.H., S. D. Schrock, and K. S. VanDerHorst. 1999. *Traffic management strategies for merge areas in rural Interstate work zones.* Ames, IA: Iowa Department of Transportation.

Mc Shane, W, R., R. P. Roess, and E. Prassas. 1998. *Traffic engineering,* 2nd ed. Upper Saddle River, NJ: Prentice-Hall.

Mohring, H. 1999. Congestion. In *Essays in transportation economics and policy: a handbook in honor of John Meyer,* eds. J. Gomez-Ibanez and W. B. Tye, and C. Winston, 181-221. Washington, D.C.: Brookings.

Mueller, J. 2005. Phone conversation on March 28, 2005. Mr. Mueller is the Northern Region manager for ADDCO, which is a manufacturer of traffic control devices.

Mun, S. 1994. Traffic jams and the congestion toll.   *Transportation Research* 28B (5): 365-375.

Paquette, R. J., N. Ashford, and P. H. Wright. 1972. *Transportation engineering: planning and design.* New York: Ronald Press.

Pavis, A., A. Saracena, H. N. Yagoda, and A. Bauer. 1995. Congestion management through bus metering at the Lincoln Tunnel. *Transportation Research Record 1496*: 35-40.

Quarmby, D. A. 1967. Choice of travel mode for the journey to work: some findings. *Journal of Transport Economics and Policy* 1 (3): 273-314.

Richards, S. H., and C. L. Dudek. 1984. *Selection of work zone channelizing devices using the value engineering approach.* College Station, TX: Texas Transportation Institute, Texas A&M University.

Small, K. A. 1982. The scheduling of consumer activities: work trips. *The American Economic Review* 72 (3): 467-479.

Small, K. A. 1983. Bus priority and congestion pricing on urban expressways. *Research in Transportation Economics* 1: 26-74.

Son, Y. T. 1999. Queuing delay models for two-lane highway work zones. *Transportation Research* 33B (7): 459-471.

Verhoef, E. T. 2003. Inside the queue: hypercongestion and road pricing in a continuous time-continuous place model of traffic congestion. *Journal of Urban Economics* 54: 531-565.

Wald, M. L. 1999. Danger, under construction. *The New York Times,* January 8, sec. F.

Table 1. Impact of a Change in the Entry Rate on Delay Time

ER	1400	1800	1900
S_B(mph)	-1.87	-.62	-.31
Q_{MAX} (Vehicles)	1315	438.32	219
Max Delay (hrs)	.94	.24	.115
Av. Delay (hrs)	.47	.12	.058
S_D (mph)	1.19	2.38	2.67
T_2 (hrs)	3.14	.52	.233
Arrivals during T_2	3144	524	244
Total Delay Time	3355	550.81	244.12
Total Time Cost	$43,780.96	$7188.11	$3185.80
E	-5.9	-11.30	-21.16

Table 2. Impact of an Increase in AR_1 to 2200 cars per hour

ER	1400	1800	1900
S_B(mph)	-2.49	-1.24	-.93
Q_{MAX} (Vehicles)	1753	877	657
Max Delay (hrs)	1.25	.49	.346
Av. Delay (hrs)	.626	.24	.173
S_D (mph)	1.19	2.38	2.67
T_2 (hrs)	4.19	1.05	.698
Arrivals during T_2	4192	1048	699
Total Delay Time	5379.78	1326.61	882.16
Total Time Cost	$70,205.81	$17,312.29	$11,512.15
E	-5.3	-6.798	-8.49

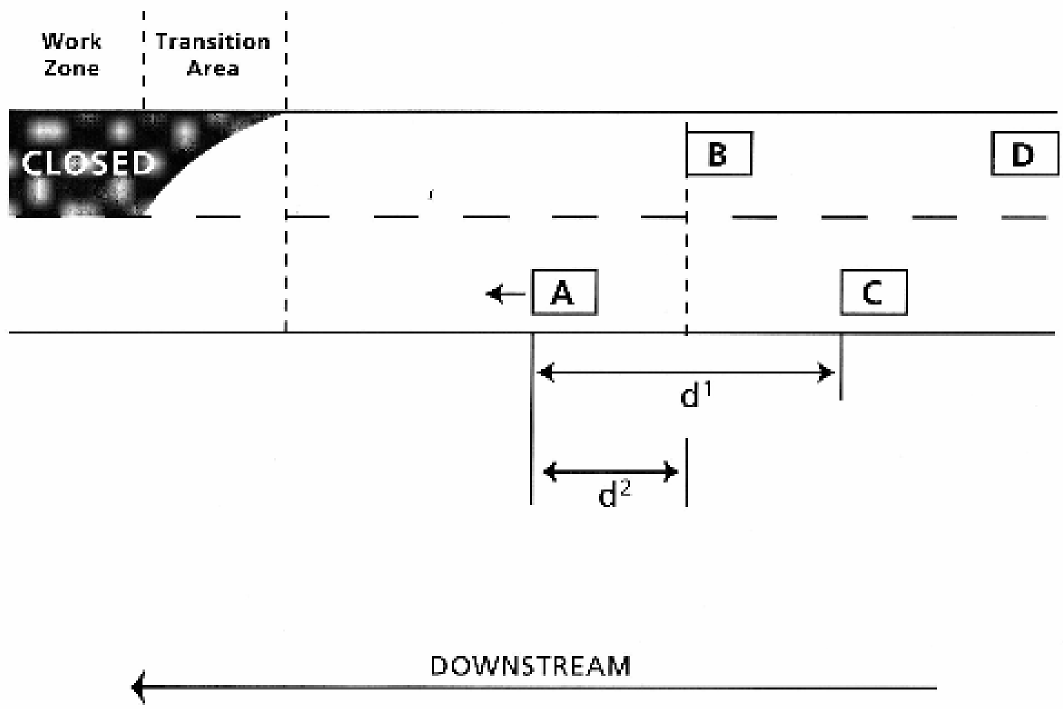

Figure 1. Structure of the Roadway

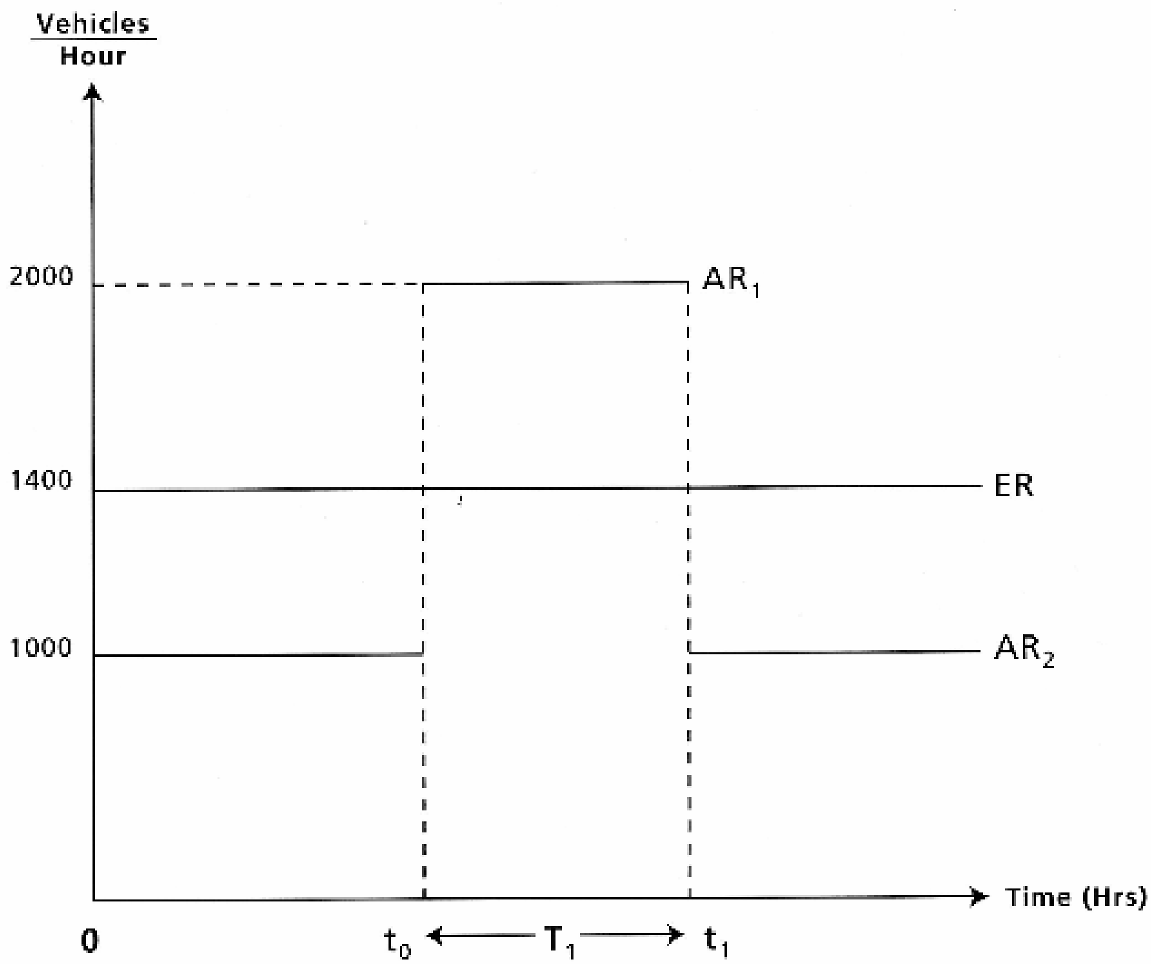

Figure 2. The Relation Between Demand and Capacity