

THE ORIGINS OF TRAFFIC THEORY

DENOS C. GAZIS

26 Lake Road, Katonah, New York 10536, denos@worldnet.att.net

"In the beginning there was the Ford."

With this biblical beginning, I started a paper in 1974 reviewing the development of traffic control theories. I think it is appropriate to start the present paper the same way.

Traffic problems existed, to some extent, even before the Ford, which ushered the age of universal automobile transportation. But the problems were relatively small and isolated, and did not provoke much thought for their solution. The automobile changed all that in a big way, and by the middle of the 20th century traffic problems were big and commanding much attention. By 1950, scientists from many walks of life came forward with attempts to model the movement of traffic, with the ultimate goal of finding amelioration to traffic problems.

Some of the early contributions to traffic modeling were those of Reuschel (1950) and Pipes (1953), on one hand, and Lighthill and Whitham (1955), on the other. Reuschel and Pipes proposed a traffic model describing the detailed movement of cars proceeding close together in a single lane, a "microscopic" model of traffic. Lighthill, a world renowned fluid mechanics theorist, together with Whitham, proposed a "macroscopic" model of traffic, modeling traffic as a continuum akin to a fluid.

The Lighthill-Whitham (L-W) model is based on two premises, the conservation of the number of cars and the existence of an equation of state describing a relationship between traffic flow, measured in cars per hour, and traffic density, measured in cars per mile. The conservation of cars is, of course, a rational assumption. The existence of an equation of state is partially verified by measurements, with much scatter observed in the data, particularly at high traffic densities. Nevertheless, the L-W model provides a pretty good description of some basic phenomena in traffic, such as the propagation of "shock waves" which are generated when traffic shifts from one steady-state pattern to another associated with different density and flow. The shock wave theory of L-W is, however, only good in describing density changes that are separated relatively widely in time, i.e., transitions from one steady-state traffic flow situation to another. Attempts to use the L-W wave theory to describe detailed movement of traffic around intersections have been

badly misguided, because the time of the generation of one shock wave is of the same order as the time of transition to the next, and the inherent approximation of the shock wave produces large errors in the modeling of the movement of traffic.

The Reuschel and Pipes models described the movement of a car following another one in front. They were based on the assumption that the speed of the following car was a linear function of the distance between the lead car and the following car. The models were reasonable in concept, but no experimental verification of their conclusions was pursued for many years.

By the mid-1950s, traffic models had attracted the attention of such notable scientists as Elliott Montroll. Then, in 1956, a confluence of events spurred a rapid development of traffic theory. General Motors hired a new executive in charge of its Research Laboratories. He was Larry Hafstad, a nuclear physicist by background, who had the ambition of making the General Motors Research Labs a leader in "basic science" and not just a development lab as it had been for years. One of the persons hired to bring about this change was Robert Herman. Herman teamed up with Montroll and pursued some basic investigations of car-following which have stood the test of time to this day.

I was hired into the Basic Science department of GM Research in 1957 by Bob Herman. My original assignment had nothing to do with traffic. Herman, in his unrivaled spirit of pursuit of knowledge in every meaningful direction, turned me loose around the Lab to see how I could apply my knowledge of applied mechanics to the modeling and improvement of devices and processes useful to the automobile industry, such as the production of better ball bearings through better understanding of contact stresses between the balls and their casing. However, it was not long before I was attracted to the traffic modeling activities, which ended up keeping me busy for a good part of my life for years to come.

Let me take a moment to stress the influence of corporate management on the research output of an organization. I joined GM Research at the advice and recommendation of my Columbia University Ph.D. mentor, Professor Ray Mindlin. Mindlin told me: "Automobile companies are not

Subject classifications: Decision analysis: driver behavior in traffic flow. Organizational studies: performance and improvement of traffic systems. Professional: comments on.
Area of review: ANNIVERSARY ISSUE (SPECIAL).

0030-364X/02/5001-0069 \$05.00
1526-5463 electronic ISSN

known for their basic research, but GM has a new management there and they seem to be serious about changing things. You ought to give them a try.” When I got to GM I did indeed find a new spirit. Not only was Hafstadt strongly supportive of our research efforts, but an extremely supportive senior executive by the name of John Campbell became our guardian angel, taking personal charge of looking after the needs of our group. We got the best support in the industrial and academic community, including the best computer available at the time anywhere. And Bob Herman had a generous allowance for consulting support, which allowed him to bring over many of his friends either on daily visits, or even on extended “sabbaticals.” This is how Montroll, Potts, and later the Nobel laureate Prigogine ended up being involved in our traffic theory activities.

It is important to stress some fundamental principles that guided the work of Herman and his colleagues. The first was based on Herman’s conviction that traffic theory was inherently an experimental science and should be pursued as such. The second one was that the mathematical model should be chosen as the one most suitable for describing a particular phenomenon, rather than trying to fit a phenomenon to a model particularly familiar or attractive to a researcher. I mention this because we have seen in recent years a plethora of publications based on “a solution in search of a problem,” in which scientists use their favorite theory to describe traffic without any due consideration of whether or not the theory is justified on physical grounds. I have said frequently that this reminds me of the presentation at a Physical Society meeting by a young scientist who proposed a model for a process and showed a remarkable fit of his theory with experimental data, correlating highly with a straight line intersecting both the x and the y axis. At the end of his talk, a noted physicist remarked that he was very impressed by the remarkable agreement of theory and data, except for the fact that on physical grounds, which he explained, the line should go through the origin. The moral of the story is that our work at GM Research was motivated by the conviction that it was not just the best fit that counted, but the best fit to a model based on **rational physical considerations**.

The model of traffic pursued by Herman and Montroll, and the people they brought into their work, such as Potts, Rothery, and myself, was similar in concept to the Reuschel and Pipes model, but it had meaningful mathematical differences. It stated that the **acceleration** of the following car was proportional to the **relative speed** between the lead and following cars, with a time-lag between these two quantities. In other words, it was assumed that drivers were trying to catch up with a car pulling away, or slowing down while closing in on a car, but they did it after a certain time-lag which depended on their own reaction time-lag and the physical characteristics of their car. To achieve experimental validation of the model, Herman and Montroll, together with Chandler, carried out some car-following experiments at the test track of General Motors. It might be mentioned here that some of the GM engineers

tried to encourage Herman to sit back and let them design and carry out the experiments according to his general specifications. Herman simply refused to allow such a thing to happen. He insisted that only if the experimental scientist designed and carried out his own experiments would an experimental science advance.

The experiments might be described as somewhat primitive by modern standards. Because technology for remote sensing of the distance between the cars was not quite readily available, the two cars were physically connected with a wire wound around a reel mounted on the front bumper of the following car and hooked to the rear bumper of the lead car. The data were recorded in a device located in the trunk of the following car. Several drivers were asked to drive under varying driving conditions of mean speed. The results showed a high correlation between the acceleration of the following car and its relative speed with respects to the leader, with a time-lag in the neighborhood of one second, and a “gain factor” which appeared to depend on the distance between the cars. The results, published in a paper by Chandler et al. (1958), have served as the foundation of microscopic traffic models until today.

The car-following model, after validation, was used to investigate traffic phenomena. One key investigation concerned the stability of a stream of traffic forming a platoon of cars following one another (Herman et al. 1959). The stability can be investigated by asking what happens to a “signal” as it is passed from one car to another. If the lead car of a platoon undergoes a maneuver, for whatever reason, it starts the signal. The signal may be attenuated or amplified as it is passed from one car to another. It turns out that whether the signal is amplified or attenuated depends on the product of the time-lag times the gain factor of the car-following equation. We worked out several examples showing how the signal could be tolerated by the first few cars but be amplified in passing down the line of cars, causing an accident at some point. In one of the illustrative diagrams which I used in my talks, the accident occurred between the seventh and eighth car of a platoon. I recall meeting a lady a year after she heard my talk. She rushed to me and said: “You don’t know what you’ve done to me, Denos. Every time I get caught into a platoon, I count the cars in front of me. If I happen to be seventh or eighth in line, I get the heck out of there, because I can’t stand the suspense.” I guess that is how superstitions get started, notwithstanding the best intentions of us traffic theorists.

I have another story related to car-following, and it has to do with the influence of the theory on the theorists themselves. Someone asked me if my driving habits had changed as a result of my studying the movement of cars. I said it certainly did, and I do believe that I shaved off some gallons from my gasoline consumption, because I somehow trained myself not to press the gas pedal when I know it is going to be fruitless in the next few seconds. But the best story is a lesson on how to avoid tailgaters. Bob Herman and Ren Potts were driving one day and noticed someone

right up to their tail. They were annoyed but unable to shake him off, until Ren thought of something. He told Bob to turn his lights on (it was day time) and at the same time put his foot on the gas pedal. The tailgater perceived the lights-on as braking, so he put his foot on the brakes, but was flustered to see the car pull away. After a couple of repetitions of this trick, he fell back and stopped tailgating. I do not recommend this as a foolproof method of shaking off tailgaters, but it would not hurt to try it if you get desperate. It certainly is not illegal!

And since I am into stories, let me tell another one related to how personal preferences influence the work of all scientists and in particular traffic theorists. There was a noted scientist whose field I could only describe as mathematical biology, by the name of Rashevski, who once decided to tackle the question of how people drove, and what was the result of their driving characteristics. He conjectured that people tried to drive straight on a highway by aiming down the middle of the road. Because they were human, they made a small error and their car drifted toward the edge of the lane. As they approached that edge, they made a correction, trying to aim down the middle, but again they made another error and started drifting toward the other edge of the lane. This produced a zigzag trajectory of the car. Then Rashevski, assuming some reasonable characteristics for driver reaction time-lags and such, computed the maximum speed that people could possibly achieve, regardless of the power of their car.

I was in Herman's office one day, and we were discussing Rashevski's driving theory – I admit with a great deal of skepticism. A mutual friend, who also knew Rashevski, walked into the room and asked for an explanation of the diagrams on the blackboard. We told him about Rashevski's theory, and his eyes lit up.

"By God," he exclaimed, "that's exactly how Rashevski drives!" The moral of this story is that it is not only professional experience that may bias a researcher's thinking, but also his or her personal habits.

MACROSCOPIC VERSUS MICROSCOPIC TRAFFIC FLOW THEORIES

One may ask which of the two approaches, the macroscopic or microscopic one, is the most useful in modeling traffic and seeking its improvement. Indeed, this question was raised by some macroscopic traffic theorists who felt that their approach was the first and the best in modeling traffic. The true answer lies, of course, in the type of traffic phenomena one tries to model, just as physical models of solids move from continuum to lattice models in investigating distinct classes of behavior. Nevertheless, some form of rivalry was manifested in the presentation of some macroscopic models, such as that of Greenberg (1958) in which an improved relationship between flow and density was obtained on the basis of an equation of continuity of a compressible fluid.

An interesting merging of the two approaches took place in 1959, when a macroscopic relationship was obtained

from the microscopic car-following model. This signaled my own formal entry into the family of traffic theorists. At the time, Renfrey (Ren) Potts was spending some time in our GM department on a visiting appointment away from his base in Australia. Potts, Herman, and I (Gazis et al. 1959) asked the question of what could be revealed from the car-following equations regarding the transition from one steady-state traffic flow situation to another. By integrating the car-following equation we obtained such a transition equation. Interestingly enough, we obtained the Greenberg macroscopic relationship from a car-following model in which the "gain factor," or sensitivity as we used to call it at the time, was inversely proportional to the distance, very much in accordance with our own experimental car-following data. This paper of ours, together with the other two seminal papers on car-following theory (Chandler et al. 1958, Herman et al. 1959) were published in the *Operations Research* journal, which provided a home for our traffic theory publications right from the start. In 1959, the collection of these three papers received the Lanchester Prize of Operations Research. Because of the relatively large number of participants in authoring these three papers, this was probably the smallest per capita allocation of the prize in the history of ORSA/INFORMS. I did get a raise at GM, though.

In a follow-up paper, Herman, Rothery, and I (Gazis et al. 1961) generalized the results of the above paper by assuming a car-following model in which the gain factor was proportional to the speed of the following vehicle raised to some power, m , and inversely proportional to the relative distance raised to some power, l , resulting in what some people started calling the "L&M model." Integration of this equation could yield just about any sort of flow versus concentration equation anyone could desire. Of course, we did not promulgate superiority of any of these models, but just showed compatibility of these models with macroscopic models.

I should mention here that a few years later I showed how the car-following could be derived by the desire of a driver to follow a rule of maintaining a distance from the leader depending on the car speed, such as the "California rule" of allowing a car length of separation for every 10 miles per hour. If we assume that the driver tries to minimize the integral of the square of the deviation from this desired pattern, subject to appropriate constraints, we obtain a Lagrangean derivation of the car-following equation through differentiation of the objective integral. The exact form of the derivative car-following equation will depend on the car-following rule assumed.

SOME APPLICATIONS, GOOD AND BAD, OF TRAFFIC FLOW MODELS

One of Herman's contacts was Leslie (Les) Edie of the New York Port Authority, which operated the tunnels and bridges linking New York and New Jersey as well as the major airports of the metropolitan New York area. One of

the tunnels, the Lincoln Tunnel, was the victim of heavy congestion during the rush hour. A somewhat puzzling aspect of this congestion was that the capacity of the tunnel seemed to drop substantially during congestion, frequently below 1,200 cars per lane per hour, as compared to values as high as 1,800 obtained on normal highways. After many discussions, it was concluded that somehow the tunnel was forced into the low flow, high density range of the flow versus concentration relationship, because too many cars found themselves trapped inside the tunnel. It was decided to try and limit the entry of cars into the tunnel when the density inside the tunnel started rising, with the hope of shifting the operation of the tunnel near the maximum throughput. Some early experiments appeared to be encouraging on the application of this concept of "increasing throughput by limiting input!" But a conclusive answer to the question had to wait for a few years and will be given later in this paper.

Another attempted application of the macroscopic traffic flow model was less successful, although it was promulgated by one of the leading members of the growing population of traffic theorists, Reuben Smeed of the Road Research Laboratory in London, England. Smeed was leading a prominent group of researchers at the Road Research Laboratory, including Wardrop and Tanner who made substantial contributions to traffic theory. Smeed himself had a passion for collecting information about traffic that he could use to develop his models, and he traveled with notebooks full of this information in his pockets. He was particularly active in developing "realistic" designs of networks of urban centers which could accommodate traffic demands in an optimal fashion following route patterns designed to minimize conflicts. Applications of Smeed's designs have been by necessity limited due to limited capability of altering the roadway patterns of urban centers, but his work remains as a model of imaginative thinking in transportation planning.

In 1967, Smeed had a look at the L-W model with an eye toward its possible message regarding the consequences of conflicts in traffic. He published a paper in which he claimed that under certain circumstances a driver could arrive at his destination, along a certain route, earlier by leaving later. His conclusion was based on the fact that by entering into a highway link, a driver in effect was increasing the density of this link, and therefore on the basis of the flow versus concentration relationship was decreasing the flow and the speed on this link. By delaying his entry, he might maintain a lower density, hence a higher flow and speed, and therefore cover the distance fast enough to compensate for the slight delay in entering. All this was documented fully through proper integration of the pertinent differential equations, resulting in the "Smeed paradox." The problem was that Smeed, in pursuing this approach, was in effect propagating traffic signals forward by assuming that the entry of a vehicle into a link instantaneously resulted in a decrease of the flow rate in that link. I remember that

this misconception persisted for a few years, and I was discussing the "paradox" with Wardrop, another leading member of the traffic theory community and associate of Smeed, who was defending Smeed's conclusions. I finally pointed out to Wardrop that if at the time the driver decided to wait for a later entry we allowed a "phantom" car, or marker, to enter the stream, without of course affecting the flow in any way, this "phantom car" would proceed to complete the trip and the driver would never be able to catch up with it. Wardrop chuckled and admitted that now that I put it this way, the "Smeed paradox" did not make sense. I mention this example to illustrate the possible pitfalls in applying carelessly a traffic flow model, something that unfortunately has happened repeatedly over the years. In recent years, some electrical engineers entering the field of traffic theory started talking about the "impedance" of a traffic link increasing as a function of its density, an obvious allusion to the increased impedance of a wire with increasing amperage. Unfortunately, traffic links are not like electrical wires, and careless use of the impedance concept can result in conclusions analogous to the "Smeed paradox." Let me conclude this discussion by reemphasizing that, notwithstanding the error of his paradox, Smeed was a true leader in traffic theory modeling, who left his mark in many ways through his personal research as well as leadership in research management at the Road Research Laboratory and teaching at the University College of London.

MORE FLOW THEORIES

Bob Herman showed leadership in more than choosing and carrying our research projects. He was convinced that for our work to have any beneficial effects, it had to reach out to others who had influence on matters related to traffic. All of us joined appropriate committees of the Highway Research Board, the precursor of today's Transportation Research Board, to bring together transportation practitioners and traffic theorists. We formed the Transportation Science Section of ORSA, the first such Section of the society, to be followed by many others. And in 1960, we organized and held the First International Symposium on Traffic Theory at the GM Research Labs in Warren, Michigan. This symposium has been held every three years since then, around the world.

The first symposium brought together an impressive gathering of leaders in the now growing field of traffic theory, including Smeed and Wardrop from the Road Research Laboratory of England, Potts from Australia, and one other unusual participant. He was Ilia Prigogine, a prominent leader in Statistical Mechanics who had expressed interest in modeling traffic as an ensemble of units with statistically distributed properties interacting with each other, analogous to the Boltzmann model of solids. Prigogine presented his ideas at the symposium, after which he continued discussing them with Herman. This led to the collaboration of the two in developing a Boltzmann-like model of traffic flow (Prigogine and Herman 1971) on multilane highways.

The Boltzmann-like model filled a lot of gaps in traffic flow modeling. First of all, previous models inherently described single-lane traffic. Second, these previous models did not appear to describe in any meaningful way very light traffic in which interactions between cars were rare or absent altogether. The Boltzmann-like model covers both of these shortcomings. It was based on the assumptions that the ensemble of cars is associated with a distribution of "desired speeds" of the drivers, who drive at those desired speeds as long as they are not impeded by other cars. But when a fast car reaches a slow car and is prevented from passing it immediately because of the presence of other cars in the vicinity, it slows down for a while until it finds a passing clearance, when it passes and resumes its desired speed. Herman and Prigogine proposed a differential equation for the speed distribution as a function of space and time, with terms accounting for the effect of interference between slow and fast cars chosen for mathematical convenience and plausibility. True to his tradition as a leading expert in statistical mechanics, Prigogine referred to this interference term as a "collision" term, a rather ominous term when used in conjunction with traffic.

The model produced remarkably reasonable results. At very low densities, cars moved relatively freely, the mean speed was the mean of the speed distribution, and the flow increased linearly with density. As the density increased, interference between cars produced a decrease of the flow below the straight line, leading up to a maximum and then a decrease of flow with increasing density. Herman and Prigogine then defined a transition from individual movement to collective movement of cars in which cars had to follow one another because passing was virtually impossible. There have been numerous attempts to validate the model with experimental data, some obtained through aerial photography, with generally satisfactory results. One thing that can be said about experimental results is that they exhibit a sizable scatter, particularly in the range of high densities, due to many random events that are not accounted for in any model. I shall discuss some of these random events later in this paper.

Herman and Prigogine went on to pursue another favorite goal of Herman's, devising something equivalent to *Ohm's Laws* of traffic flow. They proposed the *Two-Fluid Model* (1971) of town traffic, which describes traffic in towns as comprising some moving vehicles and some stopped ones. The two-fluid model ends up describing traffic in terms of two parameters, the percentage of cars that are "immobilized," and the average speed of the moving cars, which depends on the percentage of immobilized cars. A few observations suffice to determine these parameters for a given city. The two-fluid theory showed a remarkable agreement with urban traffic data and shows promise of becoming a valuable tool for transportation management and planning.

THE AMBER LIGHT PROBLEM

I will discuss the Amber (Yellow) Light Problem as an example of one that does not fit any particular general the-

ory but introduces another interesting feature, the interaction of traffic characteristics and human intervention in the form of traffic laws. It is also one that was generated by personal experience of one of its protagonists.

Late one morning, we were waiting for Herman to arrive at GM. He was later than his usual of 30 minutes after the starting time, and we did not know what had happened to him. He suddenly walked into the office in a state of fury, and started ranting about that idiot police officer who had given him a ticket for going through a red light and refused to accept his explanation. "I bet I know what you told him, Bob", I said. "You told him that because of your car's speed, the relativistic red shift changed the perceived color of the light!" Herman did not think it was funny, and continued ranting and raving. To humor him, we pursued the discussion, and soon enough we came up with some interesting elements of one of the most widespread problems in traffic. It turns out that according to traffic ordinances drivers are, in principle, allowed to "proceed with caution" (not "accelerate with caution" as some people do) when they do not think they can stop when they see a yellow light. But most ordinances also state flatly that under no circumstances can they be within the intersection during the red phase. Our investigation revealed that there are many cases when the intersection is so wide that it cannot be cleared during the yellow phase when a driver is past the point at which he can stop safely when he sees the yellow light. I was writing this paper for submission to *Operations Research* when a friend saw it and protested that this was not operations research, just algebra and a little calculus. Of course, we disagreed with his definition of what constitutes operations research, and fortunately so did the editorial board of *Operations Research*. As for the ticket, Herman paid it, but he had the satisfaction of going on record instructing the world of the need to pay attention to possible incompatibilities of traffic control design and traffic ordinances. As we wrote, with tongue in cheek, in our paper (Gazis et al. 1960), "with an adequate amber phase it would be easier to separate the violators from the non-violators, insofar as traffic is concerned ..."

TRAFFIC ASSIGNMENT

Traffic assignment addresses the assignment of routes from various origins to various destinations in an urban network. It involves the problem of choosing routes for each origin-destination pair so that the quality of travel is improved for each driver. The contention for finite roadway space obviously causes delays to all drivers. Ideally, one would like to choose the travel routes so that the travel time would be minimized for all drivers. By the time traffic theorists started tackling this problem, network flow theory had been developed for other purposes such as telephone networks, and there was a library of algorithms for obtaining such things as shortest routes between nodes, maximum capacity between nodes, etc. This library of algorithms was then

applied to traffic networks, taking into account the specific properties, including possible nonlinearities, of traffic networks.

One of the most notable early contributions to this field was the work of Wardrop, of the British Road Research Laboratory. As early as 1952, Wardrop (1952) published a paper in which he proposed two different principles for assigning traffic to a network on a system basis. The first one was the principle of equal travel times, stating that travel times were identical along any used routes connecting a pair of an origin and a destination, and less than or equal to the travel time on all unused routes. The second one was the principle of minimizing average travel time, or total travel time of all travelers on the network. The first principle was an obvious one that might be pursued by a knowledgeable traveler interested in minimizing his travel time. The second one was a sensible one for a manager of a network possessing the capability of assigning routes to different travelers. Wardrop, and others after him, observed that the individual choice of routes was by no means guaranteed to satisfy the second principle, and in most cases did not. The "Wardrop principles" have been the starting point of development of modern traffic assignment theories.

Over the years, a great deal of work on traffic assignment has been carried out, including refinements of the early theories to include nonlinear effects, perturbations, and instabilities. For many years, all this work addressed a single steady state of traffic assignment. More recently, attempts have been made to address the problem of "dynamic traffic assignment," in which the traffic generation changes as a function of time, dictating an appropriate time-dependent traffic assignment. I have a word of warning for those wishing to use such algorithms of dynamic traffic assignment, and it is related to the mistake that led to the Smeed paradox. In changing from one traffic assignment pattern to another, one has to make sure that proper account is taken of the transition from one steady state to another, and the requisite time to achieve such a transition. Another way of putting it is that in solving a differential equation, or evaluating an integral, one rightly assumes that a function changes instantaneously with a change of its variables. This, however, is not the case if the function is such a quantity as the flow along a link of a network. A change of this flow may be caused by a change in the origin-destination matrix of the network, but it takes minutes or even hours to be felt, and it is subject to the unidirectional nature of propagation of traffic signals, which move only backward in spite of the futile attempts of tailgaters.

TRAFFIC CONTROL

In late 1961, I left GM to join IBM Research. The move was again inspired by my mentor, Ray Mindlin, who told me of the opportunity that IBM held in launching a rapid growth of its research activities, and recommended me to IBM in his consulting capacity. I left GM with a great

deal of sorrow for leaving a friend like Herman, but we maintained contact and frequently collaborated on various topics until his passing in 1997.

I was not hired by IBM for my record as a traffic theorist, but rather for my expertise in applied mechanics and solid-state physics. Interestingly enough, I walked into IBM Research only to find that Montroll had preceded me and was running the General Science Department, while I was a member of the Math Department. Fate quickly made corrections to the assignment I had been given. It turns out that the IBM sales force was very actively pursuing the idea of selling a computer to New York City in order to computerize its traffic control in a way that had never been done before, and no one knew how to do. They found out that I was an expert on traffic and came after me to help carry out their noble objective. Soon enough, I met with the traffic commissioner of New York, the famous Barnes, inventor of the "Barnes Dance," stopping all traffic at intersections and letting people cross in all directions at once. Montroll was also drafted to come along to that meeting. It took a while before a computer sale to New York City materialized, but my reentry into traffic theory was fast and decisive. We eventually tested a computerized traffic control system in San Jose, California, similar systems were deployed around the country, and eventually a collection of them was installed in New York City.

In the course of pursuing these activities, I tackled some basic traffic control problems. While going over problems of timing of arterial traffic lights and their implementation in a computerized system, I addressed the problem of optimizing the control of "critical intersections" through a properly timed traffic light. The problem of optimizing the timing of a signalized intersection had been treated by Webster (1958). Webster derived enduring rules for setting the duration of green phase for conflicting streams of traffic in a way that minimizes the delay to users. However, the case treated by Webster involved intersections in which all queues were served during the green phase. I addressed a key intersection control problem in which such timing algorithms did not make sense, the problem of the Over-saturated Intersection, in which demand over a period of congestion exceeds its capacity, resulting in long queues which are only dissipated at the end of the rush period. I developed the first theory of optimal control of such an intersection with the objective of optimizing the aggregate waiting time of the users of the intersection. Coincidentally, Ren Potts came to our lab as a guest of Montroll, and we collaborated in this investigation.

The key to the solution of the problem was the observation that, in such a system, the best one could do was to maximize output from the system as soon as possible. This led to the somewhat unexpected result that the best control strategy was to give the maximum possible green light to the stream associated with the maximum "saturation flow" across the intersection, e.g., the stream with the greater number of lanes, and switch the strategy near the end of the rush period. An appropriate timing of this switching

could serve out the queues, at the same time minimizing the aggregate delay. The first paper on this topic (Gazis and Potts 1965) was presented at the Second International Symposium on Traffic Theory held in London in 1963.

When I presented my results at another meeting attended by control theorists, I was told that I was promulgating a "bang-bang" solution in traffic control, a rather alarming nomenclature when used in conjunction with traffic. I was also inspired to look into the work of Pontryagin, a famous control theorist, and recast my formulation (Gazis 1964) in terms of the "Pontryagin maximum principle." I was pleased that I had preserved my record, and Herman's, of not trying to fit a chosen theory into a problem, but rather letting the appropriate theory find its way into the treatment of a problem.

I went on to extend the treatment of the oversaturated intersection into a theory of "Store and Forward Networks," involving a network of oversaturated intersections. In its ultimate form, such a problem involves optimum routing of traffic together with optimum timing of all traffic lights during the rush period. But even assuming that the routing is given, one can pose and solve a meaningful problem of minimizing aggregate delay to users. In recent years, I have proposed that the challenge in managing congested systems is one of providing the best possible "Service Channels," or time-space trajectories, to the aggregate of users in a way that minimizes their aggregate delay. It is such an approach that holds the greatest promise of accomplishing the objectives of the Intelligent Transportation Systems (ITS), which are pursued worldwide today. Some of my thoughts regarding this promise are contained in a book chapter I authored recently (Gazis 1998) on the possible use of Intelligent Agents and the Internet for accelerating the deployment of ITS.

I should reemphasize, in connection with my traffic theory activities at IBM Research, the importance of corporate attitudes in promoting advanced research. I was blessed at the time with having the support of my management in pursuing my research with no strings attached. The support was particularly strong when Ralph Gomory took over first the management of the Math Department and then the management of the entire Research Division. I was not given the opportunity to hire more people to join my activities because such resources were more urgently needed in order to pursue the wide-ranging research for the advancement of mathematical sciences. But I was given by Gomory the ability to call on existing facilities and personnel in order to pursue my objectives, which he told me he considered very worthwhile and well done.

The IBM company at large also gave me the nod of approval. In the mid-1960s, IBM launched its first public advertisements of the company. The first series of published advertisements involved magazine centerfolds featuring IBM individuals who made unusual contributions for the benefit of our customers or society at large. One of these advertisement, for example, featured John Backus, the inventor of FORTRAN. Another one featured me. I was

pictured on centerfolds, in proper IBM attire of course, in front of a traffic light prop, and with a story under the heading: "This man is trying to improve your driving experience. What is he doing at IBM?" A couple of years later, I starred in a three-minute TV advertisement (2:57 min to be exact, to avoid my having to join the Actors' Guild and be paid appropriately), which detailed my work on traffic theory and was shown at the beginning of a one-hour TV show sponsored by IBM. Those were the days!

REVISITING THE LINCOLN TUNNEL, AND TRAFFIC FLOW INSTABILITIES

Sometime in 1967, I received a telephone call. Bob Foote of the New York Port Authority asked me if I could help with the control of the Lincoln Tunnel, which he was pursuing. I had been frequently in touch with the Port Authority people, in particular Les Edie and Bob Foote. I knew that Foote and his people had built a special-purpose computer device, the "Green Box" as they called it because of its color, with which they were trying to regulate the input into the tunnel in order to improve its throughput along the lines mentioned previously in this paper. The Green Box did not appear to make the desired progress, and Foote asked me if I had an old computer that he could use. I suggested that instead of resurrecting Smithsonian Museum pieces I would try to use a modern computer. To make a long story short, I ended up using the main computer of the Yorktown Research Center, connected to the Lincoln Tunnel, 45 miles away, by a leased telephone line. Later, I used a computer specifically designed for real-time control, known as the IBM 1800 system. We went on to study the traffic phenomena in the Lincoln Tunnel in some detail, and carry out some very successful control experiments, dealing with the special configuration of the Lincoln Tunnel roadway.

The Lincoln Tunnel was a notable demonstration of the inadequacy of conventional traffic flow theories, such as the L-W theory. Such theories intrinsically describe traffic on a straight, flat road. However, the influence of road geometry can be dramatic. The Lincoln Tunnel, connecting Manhattan and New Jersey, is approximately one and a half miles long and consists of a sequence of three almost equal segments, a downgrade, a flat, and an upgrade one. It was this geometry that was the root cause of the degradation of throughput. In particular, degradation most of the time started at the foot of the upgrade section. A car, often a truck, would have to slow down at that point because of a slowdown or stoppage ahead. Once stopped, many such cars and particularly trucks could not accelerate sufficiently fast going uphill. This created large gaps and a consequent large decrease in throughput. The asymmetry between acceleration and deceleration behavior had been observed or stipulated by others before and had been described as a causative factor of a "hysteresis" effect in traffic flow, a departure from the L-W theory. In the Lincoln Tunnel we had a most dramatic demonstration of this deceleration-acceleration asymmetry, attributed to the geometry of the roadway.

The early control experiments at the tunnel consisted of limiting the input of traffic to about 22 cars per minute, hoping to maintain a steady throughput of 1,320 cars per hour. This “open loop” control worked rarely, because once congestion set in, the input of 22 cars per minute was high enough to sustain it. In our computerized, “closed loop” control, we based our decision on the premise that there existed a safe domain in the three-dimensional space of the density values in the three sections of the tunnel. We verified this premise through observations and then adopted a strategy of limiting the input into the tunnel when the three densities crossed a surface in this three-dimensional space, and removing the input constraints when the densities dropped below another surface, closer to the origin. An additional adaptive factor was the adjustment of these control surfaces as a function of the speed at the foot of the upgrade, which was known to be the point of initiation of most of the bottlenecks. The results, first reported in 1969, were quite impressive (Gazis and Foote 1969). The average throughput in the tunnel increased by at least 10%, and frequently by as much as 12%. Moreover, the average speed within the tunnel was about 30 miles per hour instead of about 10 mph observed during stop-and-go conditions. This had the additional salubrious effect of requiring decreased ventilation of the tunnel and preventing frequent car breakdowns due to overheating.

I should mention that in carrying out the Lincoln Tunnel experiment I saw a dramatic demonstration of the fact that 90% of the contribution to the solution of the traffic problem comes from knowing where the cars are. A considerable percentage of our effort was spent in accurately estimating the densities of traffic in the three sections of the tunnel. We achieved this by detecting vehicle length patterns passing through successive sensors, and counting the cars that entered the section after a distinctive pattern, (e.g., a truck followed by a car), passed the upstream sensor, until that pattern reached the downstream sensor. Later, having these accurate estimates of traffic densities, we demonstrated the feasibility of using *Kalman Filtering* techniques to obtain accurate estimates of densities, even when such information regarding vehicle length patterns is not available.

The conclusion of the Lincoln Tunnel experiment offers a demonstration of the negative effects of organizational management on research. The Port Authority management refused to spend something of the order of \$100,000 which would be required to purchase an IBM 1800 system. Given the prorated benefit of an increase of the tunnel capacity by 10%, with the cost of a tunnel in hundreds of millions of dollars, this management decision must be judged as rather unreasonable. Apparently, the Port Authority management felt that it was unnecessary to spend any money they did not have to spend in order to deliver a service that the commuters had to buy anyway. Thus, our experiment was delegated to the history of traffic theory, but yielded no lasting benefits to commuters.

CONCLUDING REMARKS

I have presented an overview, in the form of personal recollections, of some key early developments in traffic theory. I did not cover many examples of good contributions to traffic theory that I considered derivative of the topics I discussed. I ventured into adding my perspectives regarding two important factors contributing to the development of good traffic theory. One is the proper attitude of the researchers entering the field, including the recognition of the experimental nature of traffic science and the importance of avoiding the “solution in search of a problem” attitude. As is the case in all branches of science, the signal-to-noise ratio has drastically decreased with increasing participation in the field by newcomers anxious to apply their disciplines in addressing traffic problems. The other important factor is the attitude of the management in promoting and supporting good research. Sadly, financial considerations over the last several years have diminished the willingness or ability of corporate management to support such research that has no direct influence on the corporate bottom line. Can the government step in to make up for the corporate lagging? Undoubtedly the answer is yes, but government initiatives must be governed by appropriate objectives managed by knowledgeable bureaucrats. The last phrase is a loaded one, posing requirements of **appropriateness** and **knowledge** not necessarily characteristic of government programs. I recall that in one major government project in which I participated, the responsible government representatives had experience in the management of Department of Defense projects but none on traffic. And the criteria in judging the research output appeared to depend more on measuring report pages per thousand dollars of contract money, than on a rational evaluation of the quality of research. There is no doubt in my mind that we need more enlightened patrons of science like Campbell and Hafstad of GM, or Gomory of IBM, in positions of power than our modern society provides today.

REFERENCES

- Chandler, R. E., R. Herman, E. W. Montroll. 1958. Traffic dynamics: Studies in car following. *Oper. Res.* 6 165–184.
- Gazis, D. C. 1964. Optimum control of a system of oversaturated intersections. *Oper. Res.* 12 815–831.
- . 1998. PASHAs: Advanced intelligent agents in the service of electronic commerce. D. Leebaert, ed. *The Future of the Electronic Marketplace*. MIT Press, Cambridge, MA, 145–171.
- , R. S. Foote. 1969. Surveillance and control of tunnel traffic by an on-line digital computer. *Transportation Sci.* 5 255–275.
- , R. B. Potts. 1965. The over-saturated intersection, *Proc. 2nd Internat. Sympos. Theory Traffic Flow*. O.E.C.D. Publications, Paris, France, 221–237.
- , R. Herman, A. Maradudin. 1960. The problem of the amber signal light in traffic flow. *Oper. Res.* 8 112–132.

- , ——, R. B. Potts. 1959. Car following theory of steady state traffic flow. *Oper. Res.* **7** 499–505.
- , ——, R. W. Rothery. 1961. Non-linear follow-the-leader models of traffic flow. *Oper. Res.* **9** 545–567.
- Greenberg, H. 1958. An analysis of traffic flow. *J. Oper. Res. Japan* **2** 11.
- Herman, R., I. Prigogine. 1979. A two fluid approach to town traffic. *Science* **204** 148–151.
- , E. W. Montroll, R. B. Potts, R. W. Rothery. 1959. Traffic dynamics: Analysis of stability in car following. *Oper. Res.* **7** 86–106.
- Lighthill, M. J., G. B. Whitham. 1955. A theory of traffic flow on long crowded roads. *Proc. Roy. Soc.* **A229** 317–345.
- Pipes, L. A. 1953. An operational analysis of traffic dynamics. *J. Appl. Physics* **24** 274–281.
- Prigogine, I., R. Herman. 1971. *Kinetic Theory of Vehicular Traffic*. American Elsevier Publishing Co., New York, 17–54.
- Reuschel, A. 1950. Fahrzeugbewegungen in der Kolonne, *Oesterr. Ingr. Arch.* 193–215; Fahrzeugbewegungen in der Kolonne bei gleichförmig beschleunigtem oder verzögertem Leifahrzeug, *Z. Oesterr. Ingr. Architekt. Vereines* **95** 59–62, 73–77.
- Wardrop, J. G. 1952. Some theoretical aspects of road traffic research. *Proc. Inst. Civil Engrg.* 325–378.
- Webster, F. V. 1958. Traffic signal settings. Department of Scientific and Industrial Research, Road Research Technical Paper, No. 39. H.M. Stationery Office, London, U.K.