

ANALYSIS OF THE RANDOMNESS OF  
THE DEMAND AND PROVISION OF URBAN SERVICES

by

Paul Francois Bursaux

Ingenieur des Ponts et Chaussees

SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF  
MASTER IN CITY PLANNING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 1973

Signature of Author

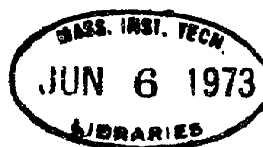
Department of Urban Studies and Planning, May 14, 1973

Certified by

Thesis Supervisor

Accepted by

Chairman, Departmental Committee on Graduate Students



ANALYSIS OF THE RANDOMNESS OF  
THE DEMAND AND PROVISION OF URBAN SERVICES

BY

Paul Francois Bursaux

Submitted to the Department of Urban Studies and Planning on May 14, 1973  
in partial fulfillment of the requirements for the degree of Master in City  
Planning.

ABSTRACT

Our analysis deals with the evolution of the provision of Urban Services and more particularly with the problem of the size of these services. It is obvious that the number of people to whom we want to provide a service is an important parameter of the problem. We want to show that in certain cases the variability of this number will be even more important a parameter.

We shall use extensively the concept of queues. The losses associated with queues naturally lead to economic considerations. Thus as a ground for further discussion we prove in Chapter I that under optimum economic conditions any system should incur congestion costs. This approach to the problem of congestion is limited for several reasons. We therefore try to build an alternate approach through probabilistic computations. This is done in Chapters II, III and IV.

In Chapter II we use a simple model to describe the way in which a certain class of services approaches saturation and the relationship between the capacity of the service and the characteristics of the saturation process. We basically want to show that in many cases the larger the capacity of the service, the closer the optimum level of provision of the service to the breakdown.

In Chapters III and IV we try to develop the following argument: a given way of providing a service is bounded in the amount of service it can provide. For various reasons when we try to increase the number of units providing the service we obtain a decreasing output. Thus when we attempt to relieve congestion by simply increasing the dimensions without changing the structure underlying the operation of the system we face increasing difficulties. We use two examples: the bus system (Chapter III) and the case of Pollution (Chapter IV). In Chapter V after having undertaken a few statistical measurements we derive the consequences and implications of our analysis.

Thesis Supervisor: Professor Aaron Fleisher  
Title: Professor of City Planning

## ACKNOWLEDGMENTS

It is euphemistic to say that I am indebted to Professor Aaron Fleisher for this work. Without the numerous hours he devoted to the discussion of this dissertation, it would simply not exist. Working with him has always been fascinating but certainly not always an easy task, and it is now that I best realize how well he has "advised" me throughout my studies at M.I.T.

I am grateful to Professor Richard Larson for guiding me in the choice of a subject and in helping me clarify my ideas and to Professor Thomas Willemain for his fruitful comments and for helping me with the cumbersome details of the computations. I am also indebted to Ms. Nathalie Robatel for giving me some new insights on the problem which is discussed here.

Given the little amount of time left for typing and editing the manuscript I greatly appreciated the editing of M. Theodore Yoos and the fast and accurate typing of Ms. Ann LeMieux.

My friend Suzanne Weinberg has shared the emotions associated with one year of work, at certain points this certainly deserves the largest credit.

Paul Francois Bursaux  
Boston, May 14, 1973

## TABLE OF CONTENTS

	<u>Page</u>
Abstract	2
Acknowledgments	3
Table of Contents	4
List of Tables	5
List of Figures	6
Introduction	8
Chapter I: Economic Analysis of the Congestion of Urban Services	15
Chapter II: An alternate Approach; Comparing the Optimum Level of Operation of a Service and its Maximum Level of Operation	23
I - Single Channel Services	24
II - Multi Channel Services	32
1. Infinite Queue Allowed	34
2. No Queue Allowed	41
Table of Symbols Used in Chapter II	43
Chapter III: Further Computations for one Specific Service: Buses and the Clumping Problem	45
Chapter IV: Further Computations for a More General Example: the Case of Pollution	66
Chapter V: Statistics, Consequences and Implications of the Analysis	77
Appendix to Chapter V	89
Bibliography	92

## LIST OF TABLES

	<u>Page</u>
1. Probability of Providing a Service as a Function of the Rate of Use for a Given Capacity	27
2. Relative Utilities of One Unit of Capacity for Various Total Capacities (One Channel Services)	29
3. Probability $Q_M$ of Waiting in Line as a Function of the Rate of Use of the System (Multi Channel Services)	38
4. Relative Utilities as a Function of the Utilization Ratio and the Capacity (Multi Channel Services)	38
5. Probability of Clumping as a Function of the Variance and Frequency of Buses	50
6. Waiting Time as a Function of the Number of Buses in the System	61
7. Lower Bound of the Probability of an Infinite Accumulation of Pollutant in the System	74
8. Mean and Standard Deviation of the Utilization Ratio for the Six Groups of Hospitals	78
9. Capacities (Tons/Day), Mean and Standard Deviation of the Three Groups of Incinerators of the State of Massachusetts	80

## LIST OF FIGURES

	<u>Page</u>
1. Probability of Saturation Versus Average Demand for a Given Capacity of Service	11
2. Cost of a Service as a Function of the Number of Customers	18
3. Cost of a Service as a Function of its Capacity	18
4. Law of Supply and Demand for a Service of a Given Capacity	20
5. Relative Utilities of One Unit of Capacity for Various Total Capacities (One Channel Services)	30
6. Economies of Scale and Flexibility as a Function of the Capacity (One Channel Services)	33
7. Probability That All $M$ Channels are Busy as a Function of the Rate of Utilization for Poisson Arrivals. $M$ Exponential Channels, Single Queue	35
8. Relative Utilities per Unit of Capacity of the Service as a Function of the Utilization Ratio and the Number of Channels	39
9. Economies of Scale and Flexibility of the System as a Function of its Capacity (Multichannel Services)	40
10. Distribution of Headways and Probability of Clumping	48
11. Probability of Clumping as a Function of the Variance and Frequency of the Buses	51
12. Expected Waiting Time as a Function of the Probability of Clumping	60
13. Expected Waiting Time as a Function of the Number of Buses in the System	61
14. Utility as a Function of the Number of Buses	64
15. Probability Tree for One Unit of Pollutant	69
16. Probability Tree for the Whole Pollution System	69
17. Lower Bound of the Probability of an Infinite Accumulation of Pollutant as a Function of the Ratio Emission Absorption	73

7.

	<u>Page</u>
18. Lower Bound of the Probability of an Infinite Accumulation of Pollutant as a Function of the Amplification of the System	73
19. Mean Utilization Ratio of the Hospital Beds as a Function of the Capacity of the Hospital System	79
20. Standard Deviation of the Rate of Utilization of the Hospital Beds as a Function of the Capacity of the Hospital System	79
21. Ratio of Extreme Flows to Average Daily Flows Compiled From Various Sources (Sewage)	83

## Introduction

Anybody who remembers that period knows that the year 1040 was supposed to bring a major catastrophe. Several usual catastrophies occurred in the years following this prediction but no particularly significant one in 1040. We, too, want to speculate on the occurrence of catastrophies, limiting, however, our speculations to the more specific field of urban services, but, we shall be more interested by the description of the phenomenon itself than by the exact forecasting of the date of its occurrence.

Thus, this dissertation will deal with the evolution of the provision of urban services and more particularly with the problem of the size of these services. It is obvious that the number of people to whom we want to provide the service or any estimate of this number (i.e., expected demand) is an important parameter of the problem. But, we shall see that in certain cases the variability of this demand will be an even more important parameter. It can eventually become a major threat to the operation of large services.

We shall start by discussing the fact that not only should we accept a certain congestion of the services but that in a certain sense optimality implies congestion<sup>(1)</sup>. Then, given a particular mode of provision of a service we shall consider the relationship between the optimum level of provision of this service and the level of saturation. This will argue that it is impossible to increase the size of services and not change the mode of provision without heading towards catastrophe. We shall use extensively the concept of queues. In fact, we shall see that a queue is

(1) Jerome Rothenberg, The Economics of Congestion an Integrated View  
AER, May, 1970.



not only a phenomenon such as the one we observe in front of movie theaters, but that almost all services have an explicit or implicit waiting system.

Queues can be distributed either in time or in space. A police car which responds immediately to a call randomly located in the city<sup>(2)</sup> and the grocer who takes a random amount of time to serve the customer are two different types of queues, but basically they correspond to the same phenomenon which we can describe as delays introduced by random processes. In fact, our main interest will be in the randomness of the demand for the service.

For certain services, it is possible to contend that queues should not exist. For example, there should not be any delay in the provision of obstetrical services. The emergency capacity of obstetrical departments of hospitals should be large enough so that there is no chance for a queue to form. It is easy to turn this into a general statement and say that we should plan the capacity of all the basic services (hospitals, schools) so that no one has a chance of ever waiting for the provision of these services. But, this is impossible since there is no such condition as no chance. Services do have a finite capacity and there is always a chance that they can get saturated. At the end of the last century, a Dutch scientist was trying to measure with perfect accuracy the frequency of some obscure vibration. He isolated himself in a remote palace where he noticed that there was a wave signal superimposed on the signal he was trying to measure. He concluded that this was the noise of the North Sea,

---

(2)Richard C. Larson, Urban Police Patrol Analysis, M.I.T. Press, Cambridge, Massachusetts, 1972.

and strangely enough he could still notice it even after he had moved to Switzerland. If we do not want to listen to the noise of the North Sea, we should not make such statements as "no chance."

Let us therefore reformulate the no chance problem in other terms. Let us assume that the service we are considering has a finite size. Thus there exists a maximum number of customers that we can serve at a time. We shall call this number  $N$  and refer to it indifferently as being the size or the capacity of the service. If the number of customers is  $n > N$ , we may incur a big loss, a human life, for example in the case of a hospital. If  $\lambda$  is the average demand for the service ( $\lambda = E(n)$ , expected value of  $n$ ), then for any given  $N$  the probability of facing a saturated system is a function of  $\lambda$  monotonically increasing with  $\lambda$ . In Chapter III we shall explore the behavior of this function  $f_N(\lambda)$ . One of the characteristics of congestion phenomena is that they appear suddenly for a certain level of use of the service and then tend to rise sharply for any further increase of the utilization of the service. Thus, in many cases we can assume that  $f_N(\lambda)$  is almost zero until a certain value  $N_0$  and then increases sharply to reach almost one for  $\lambda = N$ . This is described by Figure 1.

This curve is most significant because if we know the losses associated with the queue, it gives the expected losses due to congestion even if the system is normally operated below the limits of its capacity and is not usually saturated. Notice that this is and should be the case of systems the disruption of which is expensive or undesirable. As we have already noticed such systems range from police to hospitals, for which delaying a customer may have serious consequences. If we consider as a cost

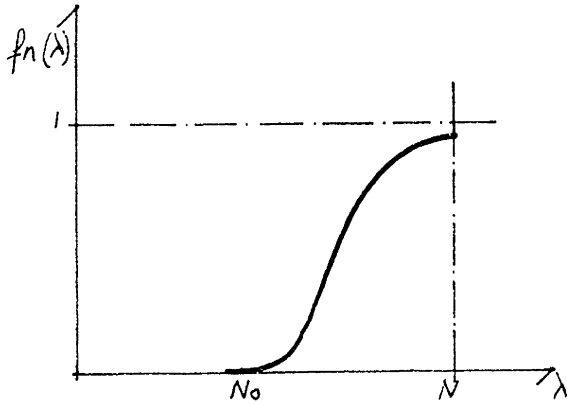


Figure 1

Probability of saturation versus average demand for a given capacity of a service

function the cost of having a queue multiplied by its probability of occurrence, we see that the systems we have just mentioned will have a loss curve increasing very rapidly as soon as  $\lambda$  becomes larger than  $N_0$ . They should therefore be operated very close to  $N_0$ , i.e., with a very small probability of congestion. But, this does not mean that the expected loss is zero. That type of system can be such that it practically never shows any congestion, but still has a significant expected loss due to the highly unlikely but most expensive disruption.

We shall call this expected cost or its equivalent monetary cost a congestion cost. As far as cost versus intensity of use is concerned, the behavior of our system which has no apparent waiting line is exactly the same as the behavior of a system with an apparent waiting line and a small disruption cost and where the only cost which has to be taken into consideration is the time wasted in the queue (it can also eventually add up to very large figures).

Up to now we have described explicit queues and explicit costs associated with queuing, but this is not always the case and a large number

of systems have hidden queues and queuing policies. This is, for example, the case of a university<sup>(3)</sup> which has a finite capacity and which because of this finite capacity rejects a large number of candidates. The admission process is based on an evaluation of the relative ability of the candidates. However, as opposed to a road system or a telephone system, in the case of the university the importance of the social environment is such that the rejected candidates do not question too much the fact of being turned away<sup>(4)</sup>.

Kafka describes the justice system in very symbolic but in fact very similar terms: "Before the law stands a doorkeeper. To this doorkeeper comes a man from the country and prays for admittance. But, the doorkeeper says he cannot grant admittance at the moment. The man thinks it over and then asks if he will be allowed in later. "It is possible," says the doorkeeper, "but not at the moment"<sup>(5)</sup>...The doorkeeper frequently has little interviews with him asking questions about his home and other things, but the questions are put indifferently as great lords always put them and finish with the statement that he cannot be let in yet..." We shall not take such a philosophical standpoint as Kafka and discuss the social and moral values of all rejection processes, but we must keep in mind that not only road networks or telephone systems are concerned with queues, queuing procedures, and rejection policies. These phenomena often

(3)Clark Kerr, The Uses of the University, Harper and Row, 1963.

(4)David Riesman, Constraint and Variety in American Education (New York) Double Day Anchor Book, 1956.

(5)Franz Kafka, Preface to the Trial.

exist at much deeper social levels. The idea that in almost any system there exist somewhere losses due to congestion or rejection will be a constant reference throughout this dissertation.

Our first concern, however, will be an economic analysis of the problem of congestion<sup>(6)</sup>. As a ground for further discussion we want to prove that under optimum economic conditions any system should incur congestion costs. This approach to the problem of congestion is limited for several reasons. We shall therefore try to build an alternate approach through probabilistic computations. This is done in Chapters II, III, and IV.

In Chapter II we use a simple model to describe the way in which a certain class of services approaches saturation and the relationship between the capacity of the service and the characteristics of the saturation process. We basically want to show that in many cases the larger the capacity of the service, the closer the optimum level of provision of the service to the breakdown.

In Chapters III and IV we try to develop the following argument: a given way of providing a service is bounded in the amount of service it can provide. For various reasons, when we try to increase the number of units providing the service we obtain a decreasing output. Thus we attempt to relieve congestion by simply increasing the dimensions without changing the structure underlying the operation of the system, we cannot succeed indefinitely. Our dissertation is developed around two examples. The first

(6) Matthew Edel and Jerome Rothenberg, Readings in Urban Economics, McMillan, 1972 (New York).

one is the bus system and the so-called clumping problem (Chapter III), and the second one is more general and describes pollution (Chapter IV).

In Chapter V, we discuss the relevance of our dissertation for real life services. This leads naturally to the discussion of the implications of our argument.

## Chapter I

Economic Analysis of the Congestion of Urban Services

In this chapter, we shall consider an economic aspect of congestion. We shall see that an economically optimum use of a public service exhibits a certain degree of congestion. By a certain degree of congestion, we do not mean that there has to be a permanent waiting line, but that the losses incurred either because of a frequent waiting line or because of the possibility of a waiting line are not insignificant.

In the introduction we described the service by two variables,  $N$  and  $N_0$ .  $N$  was the maximum possible number of customers, and  $N_0$  was the maximum average number of customers served without any significant congestion loss. In the case of a highway,  $N$  would be the maximum possible flow on the highway and  $N_0$  the maximum flow so that the movement of a car is not significantly hindered by other cars. We defined  $N$  as being the capacity of the service. In a certain sense,  $N_0$  is also a measure of the capacity of the service. In fact, as long as we know the correspondence between  $N_0$  and  $N$ , measuring the service in terms of  $N_0$  or  $N$  is indifferent. Thus in this chapter we shall define the services in terms of  $N_0$ , calling this number the effective capacity of the service.

Let us assume that in order to provide the service we face:

- a. A cost,  $CF$ , directly related to the capacity of the service. This is the cost of construction plus the cost of operating the facility independently of the amount of use which is made of

the facility. In the case of a highway CF would be the cost of building the highway and maintaining it for a zero level of use (for example, signs, maintenance, snow removal...) We shall call CF fixed cost  $CF = CF(N_0)$ ,  $N_0$  being the effective capacity<sup>(1)</sup>.

- b. A congestion cost, CG. Let  $n$  be the number of people who use the given service. By definition of  $N_0$ , the congestion cost is insignificant up to  $N_0$ , then rises sharply with  $n$ .
- c. The remainder of the operating costs which have not been taken into consideration in CF depend on the amount of use of the facility. This cost,  $C_0$  (cost per user), is a function of both the capacity of the service and of the number of users. However, if we consider  $C_0$  as a cost per user, its variation with  $N_0$  or  $n$  is likely to be small when compared to the variations of the

<sup>(1)</sup>Meyer Kain and Wohl, The Urban Transportation Problem, Cambridge Harvard University Press, 1965.



congestion cost. We shall therefore use the approximation,  $CO(n, N_0) = CO$  (Constant). In the case of a highway,  $CO$  would typically be the cost associated with the wear of the road surface.

These three costs make up the total cost  $CT(n, N_0)$ . Here  $CT$ , is a cost per user. The two following figures show the variations of these costs; first, when  $N_0$  is fixed and  $n$  is varying; second, when  $n$  is fixed and  $N_0$  is varying.

Thus, we have,

$$CT(n, N_0) = \frac{CF(N_0)}{n} + CO + CG(n, N_0)$$

Let us assume that we know the number,  $n$ , of customers to whom we want to provide the service and let us try to find the optimum size of the facility. The optimum is always such that,

$CT(n, N_0)$  is minimum Since  $n$  is given,

$$\frac{\partial CT}{\partial N_0} = \frac{1}{n} \frac{dCF}{dN_0} + \frac{\partial CG(n, N_0)}{\partial N_0} = 0$$

$\frac{dCF}{dN_0}$  is always positive (the larger the service, the more expensive it is to build it). This is the rate of change of fixed cost with effective capacity. Since we assume that  $n$  is given,  $CG$  is a function of  $N_0$ . If  $N_0 > n$  this congestion cost is zero; but, if  $N_0 < n$  this cost is positive and decreasing with  $N_0$ .

Thus, for any given  $n$ :

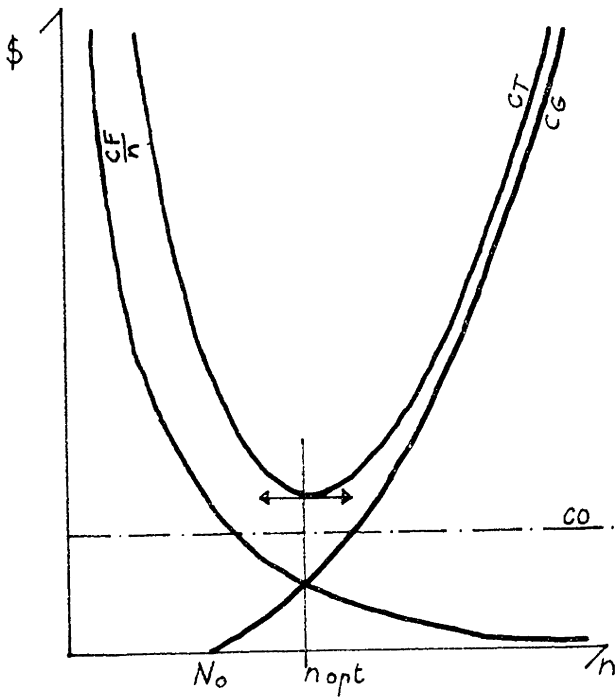


Figure 2

No fixed  $n$  varying: Cost as function of the number of customers ( $n$  fixed)

$$\frac{\partial CG}{\partial N_0} \text{ is negative for } 0 < N_0 < n$$

$$\frac{\partial CG}{\partial N_0} = 0 \text{ for } n \leq N_0$$

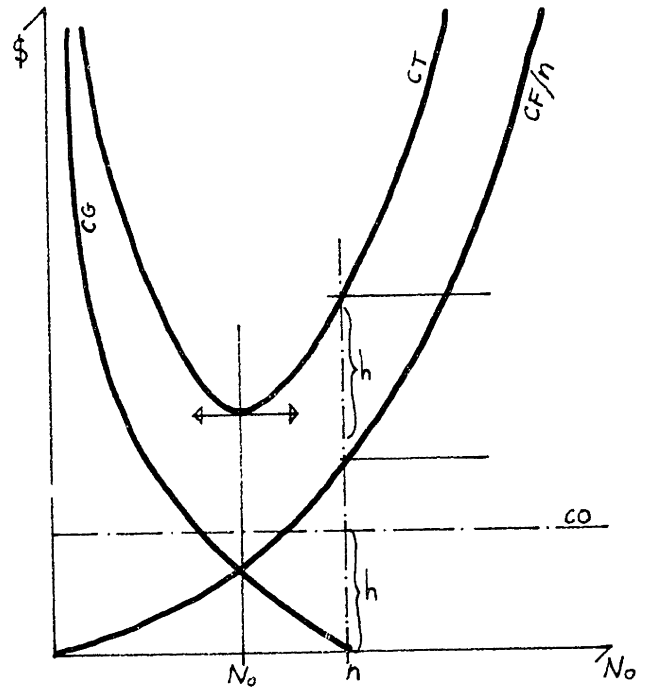


Figure 3

$n$  fixed  $N$  varying: Cost as a function of the capacity ( $n$  fixed)

In most cases, small capacities will be associated with large congestion losses. Thus  $CG(n, \epsilon)$  becomes large when  $\epsilon \rightarrow 0$ . Therefore

$\frac{\partial CG}{\partial N_0}$  is not only negative but very large for small values of  $N_0$ . We see

that there always exists a  $N_0$  such that

$$\frac{1}{n} \frac{dCF}{dN_0} + \frac{\partial CG(n, N_0)}{\partial N_0} = 0 \quad (N_0 < n)$$

Thus there exists a value of  $N_0$  which leads to a minimum cost for the provision of the service, and this value of  $N_0$  is such that  $n > N_0$  which means

that the system does indeed incur congestion costs<sup>(2)(3)</sup>.

The minimum cost for the provision of a public service implies congestion.

Notice that a congestion cost rapidly increasing with  $n$  (or decreasing with  $N_0$ ) leads to an optimum size close to the threshold of congestion (hospitals) whereas a cost slowly increasing with  $n$  (or decreasing with  $N_0$ ) leads to an optimum size much smaller than the one which would cause no congestion (roads).

We have shown that for any demand  $n$  we have an optimum capacity  $N_{opt}$  or conversely that for any capacity we have an optimum use,  $n_{opt}$ . This optimum use corresponds to a certain cost per user,  $p_{opt}$ . If we have the choice of the capacity, we can usually choose  $N$  so that  $(n_{opt}, p_{opt})$  corresponds to the actual demand, which is to say that the point lies on the demand curve. However, in many cases the capacity of the system is a given datum and we cannot modify it easily. In that case there is no reason for the demand curve to intersect the total cost curve at its minimum. This is the case described by Figure 4. If the price paid by the customer is strictly equal to the cost of the service (no tax, no subsidy), the actual system may be such that the price of the service is  $p_{ec}$  and the number of customers  $n_{ec}$ .

(2) Jerome Rothenberg, The Economics of Congestion on Integrated View AER, May, 1970.

(3) William S. Vickrey, Pricing in Urban and Suburban Transportation AER, May, 1963.

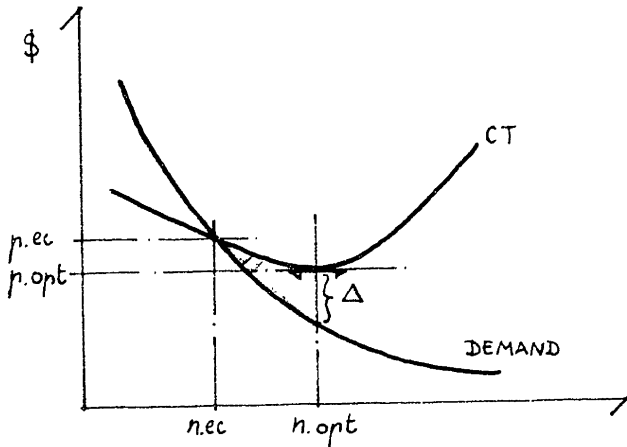


Figure 4

Supply and demand for a given capacity

Two cases are possible:

1.  $n_{ec} < n_{opt}$ . This is the case described by the figure. We immediately see that a subsidy,  $\Delta$ , would bring the system to the point  $(n_{opt}, p_{opt} - \Delta)$ . Since the two curves are decreasing together, in certain cases a rather small subsidy may significantly increase the number of people using the system. Since varying the capacity is difficult and since a small loss may enable a much larger number of people to use the system, the alternate solution of subsidizing the use of the facility has to be considered. If we did not have to give each customer the same subsidy but only the exact amount which would induce him to use the service, the loss would then be equal to the shadowed area of the figure and would obviously be much smaller. In that case, for any given  $n$ , we only need to give to the marginal customer a subsidy equal to the difference between the cost and demand curve. Notice also that since  $N$  is given, the case  $n_{ec} < N_0$  (no congestion loss) and  $n_{opt} > N_0$  is quite possible (overdesigned facilities).

2.  $n_{ec} > n_{opt}$ . The losses due to congestion are much beyond the optimum. If we are not able to modify the system so that its capacity is increased, we probably should impose taxes on the system.

To summarize: public services should normally show a certain degree of congestion; those which do not show any congestion are presumably either oversized or obsolete. However, the congestion we describe is limited. It probably would be more accurate to say that public services should be at the threshold of congestion. This is the case of the bus system we describe later in more detail. The means devised by economists to obtain this optimum is taxes equal to the difference between average and marginal prices. As we have already mentioned, this problem of taxes is standard in economics (see for example Rothenberg<sup>(4)</sup>).

This chapter has tried to give an example of the type of understanding of the congestion problem we could gain through an economic analysis. However we must not forget that we cannot compute the figures needed to define optimum economic policies if we do not know the law of demand and there is no possible means to derive this law from the demand that we actually face at the prevailing price.

Thus we have to define an "a priori" law of demand and the associated variables. It is only through an iterative process and all the associated difficulties that we shall eventually reach the real optimum. Furthermore, even if we have found the optimum, the overall system is not

(4) Matthew Edel and Jerome Rothenberg, Readings in Urban Economics, MacMillan, 1972 (New York).

itself optimum unless we redistribute optimally the money collected through congestion taxes. This is even more difficult and the actual discussions about the redistribution of the money collected through highway taxes gives an example of the type of problems we may run into.

As a whole, classical economic theory assumes that the interaction between people is summarized by a system of prices and that an optimum allocation of resources can be obtained by internalization of all possible externalities. We sincerely doubt that this is possible. We also refuse not to consider the redistributive effect of such or such economic policy. Taxing the marginal use of a transportation facility will certainly hit more strongly the worker who has no choice of the hour at which he should use the facility than the richer man who has a choice of his working hours<sup>(5)(6)</sup>.

Thus, certain policies might lead to a better economic system but a worse social system. There is a trade off between the two that we should systematically consider. The alternate approach used in socialist countries which pretends to define the needs of the population is unfortunately even more difficult to undertake. However, it indicates that there are limitations in the theory used in this chapter. Thus, in the following chapters we shall try to reconsider the problem of congestion rather independently of economics.

---

(5) Maurice Netter, Critique de la tarification au cout marginal social. la vie urbaine no. 2, Paris, 1971.

(6) Alain Juillet, Sur la rente fonciere urbaine la vie urbaine, no. 2, Paris, 1971.

## Chapter II

An Alternate Approach: Comparing the Optimum Level of Operation of a Service and its Maximum Level of Operation

In the preceding chapter we have shown that minimizing the cost of provision of a service implies congestion. However we have said nothing of the congestion process itself. For this reason we shall approach the problem of congestion from a different point of view and in the three following chapters we shall mainly deal with probability theory. In this chapter we shall try to relate the size of the facilities with the optimum level at which these facilities should be operated. We shall consider successively the case of a single channel service and the case of a multi-channel service. In both cases we shall find that the larger the service, the closer the optimum provision of the service to the breakdown. Economies of scale associated with a more efficient use of the available capacity are compensated by higher risks.

Queues in public services can build up because of the variability of the demand, because of the variability of the service, or because of both. In the study of a particular service we would have to make the distinction and evaluate separately the variability of these two components. This is one of the problems we shall have to deal with in Chapter V, which relates the conclusions of this chapter with what we observe in the outside world. In the present chapter, this problem of variability of the demand or variability of the service is taken care of differently for the single channel service and the multichannel service. For the single channel system, we assume that there is no variability in the service rate. We assume that

the service is provided on a periodical basis (period  $T$ ) and at the beginning of every period all the customers are served or rejected. This is a very conservative assumption, since, as we shall see in the following chapter, nonperiodical services turn out to be worse than periodical ones. The same computations would be valid with hardly any modification, if we assumed that the number of customers asking for the service was constant during each period and that the capacity of the service was varying in a Poisson fashion. Unfortunately, if we assume that both the service and the demand are varying, we are not able to perform the computations for this particular model. In the case of the multichannel service, we use a standard queuing model and we take into consideration at the same time the variability of both the service and the demand.

I. The first type of service we consider is provided by a unique facility (no parallel facility) and we assume that the service is provided on a periodical basis (period  $T$ ).

The number of customers is random. We assume that this randomness can be described by a Poisson process<sup>(1)</sup> ( $\lambda T$  customers on the average asking for the service during time  $T$ ). We define by  $N$  the capacity of the facility. If fewer than  $N$  customers ask for the service during the period  $T$ , they are all provided with the service in an equivalent fashion. If more than  $N$  customers ask for the service, we assume that no service is provided at all. This is a very pessimistic assumption. It is probably a good approximation for the telephone system. For other services it might

<sup>(1)</sup>Alvin W. Drake, Fundamentals of Applied Probability Theory, McGraw Hill, 1967 (New York).



not be such a good approximation. However, the next two chapters will "a posteriori" show that this is not such an unrealistic assumption and that when the number of customers approaches the capacity of the service, internal phenomena tend to reduce considerably the possibilities of operating the system.

We have assumed that the demand was a discrete Poisson process with parameter  $\lambda T$ . The probability that the number of customers asking for the service during the period  $T$  is smaller than  $N$  is

$$p = \sum_{k=0}^N e^{-\lambda T} \frac{(\lambda T)^k}{k!} \quad (\text{probable number of Poisson arrivals } \leq N)^{(1)}$$

$$p = f(\lambda T, N)$$

Which we shall write in an alternate way:  $p = f_N(\lambda T/N)$

The arrival rate is not expressed in absolute terms; but, given a capacity it is expressed as a fraction of that capacity.

If we call  $\rho = \frac{\lambda T}{N}$  the utilization ratio, we now have  $p = f_N(\rho)$

For any given capacity  $N$ ,  $p$  is a function of the rate of use of the service. Values of  $p$  are given in Table 1.

At this point, we must emphasize again that the model we use is a very specific one. In fact, every set of assumptions we may choose, corresponds to a different model. We might, for example, have worked on a more optimistic ground and assumed that even when there are too many customers the facility is providing some service and that the penalty is only a function of the overflow. In that case we would have to compute the

(1) Alvin W. Drake, Fundamentals of Applied Probability Theory, MacGraw Hill, 1967 (New York).

value of the overflow. In fact, if we use the same variables as before this overflow turns out to be:

$$\lambda T \left[ 1-p + \frac{(\lambda T)^N}{N!} \right]$$

Then we would have to compare the previous number with the average number of people being served which would turn out to be:

$$\lambda T \left[ p - \frac{(\lambda T)^N}{N!} \right] + N (1-p)$$

The rest of the analysis would be the same as for our previous model. We should notice too that the only term which is significantly different is  $N (1-p)$  which would most probably not introduce any perturbation in our conclusions.

We could also allow a queue to build up. Using a standard queuing model<sup>(2)</sup> we could then compute the size of the queue and define the penalty as a function of either the probability of queuing or the size of the queue. Whether we should allow a queue or not depends only on the particular service we are describing. It changes the nature of the losses, but not the argument. Notice though that in one case the rate of utilization has to be smaller than one; otherwise an infinite queue builds up, whereas in the other case we can operate at any level of utilization.

(2) Philip Morse, Queues Inventories and Maintenance, New York, Wiley, 1958.

$N$	$\rho$	.1	.2	.3	.4	.5	.6	.7	.8	.9	1	1.1
2		.999	.992	.977	.953	.920	.990	.834	.783	.731	.677	.570
5			1	.995	.983	.959	.916	.858	.785	.703	.616	.241
10				1	.997	.986	.967	.911	.816	.706	.583	.010
20					1	.998	.988	.953	.868	.731	.56	0
100									.985	.860	.54	0
$10^3$										$1-10^{-3}$	.515	0
$10^4$											.505	0

$$p = f_N(\rho)$$

Table 1

Probability of providing the service as a function of the rate of use for a given capacity

We shall now make some very simple assumptions about the utility of the system. We assume that this utility is  $\alpha$  per customer if we serve this customer and  $-\alpha$  per customer if the customer is turned away. In other words, we gain  $\alpha$  each time we serve a customer and we lose  $\alpha$  each time we reject a customer. Thus, the expected utility of the service for one customer is equal to  $\alpha$  weighted by the probability that this customer is served plus  $-\alpha$  weighted by the probability that this customer is not served.

$$u = \alpha p + (-\alpha) (1-p)$$

$$u = \alpha (2p - 1)$$

According to these assumptions, a service which serves randomly one customer out of two has a zero utility (casino). The total utility of the system is equal to the average utility per customer multiplied by the number of customers; for one period of time this is:

$$U = \lambda T u = N \alpha \frac{(\lambda T)}{N} (2 p - 1)$$

$$U = N \alpha \rho [2 p(\rho) - 1]$$

Since  $\alpha$  is simply a number of dollars (utility of serving one customer), we can arbitrarily set  $\alpha = 1$  and refer to the utilities as being relative utilities (dimensionless). If we want to consider a particular service we shall just have to multiply by the value of  $\alpha$  for this service. We are more particularly interested in the utility per unit of capacity which we shall call  $g_N$ . Given the fact that we already know  $p$  as a function of  $\rho$  we can now compute:

$$g_N(\rho) = \frac{U}{N} = \rho [2 p(\rho) - 1]$$

The results are given in Table 2.

However, this is not the most general way of expressing the utility of the system. In general, the losses associated with the rejection of one customer will not be equal to the benefits derived from the provision of the service to another customer. Let us call  $\alpha$  the benefit and  $\beta$  the loss. Let us also call  $K$  the ratio  $\beta/\alpha$  and indicate by a superscript this ratio in the otherwise identically defined utility functions:

$$u^K = \alpha p - \beta [1 - p]$$

$$u^K = (\alpha + \beta) p - \beta$$

$$\frac{U^K}{N} = \rho [(\alpha + \beta) p - \beta]$$

As previously we can arbitrarily set  $\alpha = 1$ . The relative utility per unit of capacity is then:

$$g_N^K = \rho [(K + 1) p - K]$$

$$g_N = \rho [2p - 1]$$

$$\frac{g_N^K + K\rho}{g_N + \rho} = \frac{K + 1}{2}$$

$$g_N^K = \frac{K + 1}{2} g_N + \frac{1 - K}{2} \rho$$

If we consider the plane  $[p, g_N]$  this transform is an affinity parallel to  $g_N$  axis. The invariant axis is  $g_N = \rho$  and the ratio is  $\frac{K + 1}{2}$ . Therefore the  $g_N^K(\rho)$  will be obtained from the  $g_N(\rho)$  by this very simple transform and we need not perform the computations for the most general case.

$N^p$	.1	.2	.3	.4	.5	.6	.7	.8	.9	1	1.1
2	.1	.197	.286	.362	.420	.455	.467	.454	.415	.353	.154
5	.1	.2	.297	.387	.460	.499	.501	.456	.365	.232	-.570
10	.1	.2	.3	.4	.486	.549	.562	.505	.370	.166	-1.07
20	.1	.2	.3	.4	.5	.6	.633	.588	.419	.112	-1.1
$10^2$	.1	.2	.3	.4	.5	.6	.7	.776	.648	.08	-1.1
$10^3$	.1	.2	.3	.4	.5	.6	.7	.8	.9	.03	-1.1

Table 2

Relative utilities of one unit of capacity for various total capacities (one channel services)

Given Table 2 we can draw the curves describing the variation of  $g_N(\rho)$  for various values of  $N$ . This is done in Figure 5. Let us assume that the service is operated at its maximum utility ( $g_N(\rho)$  maximum) and let us call  $\bar{\rho}$  the corresponding utilization ratio. In other words, the maximum utility

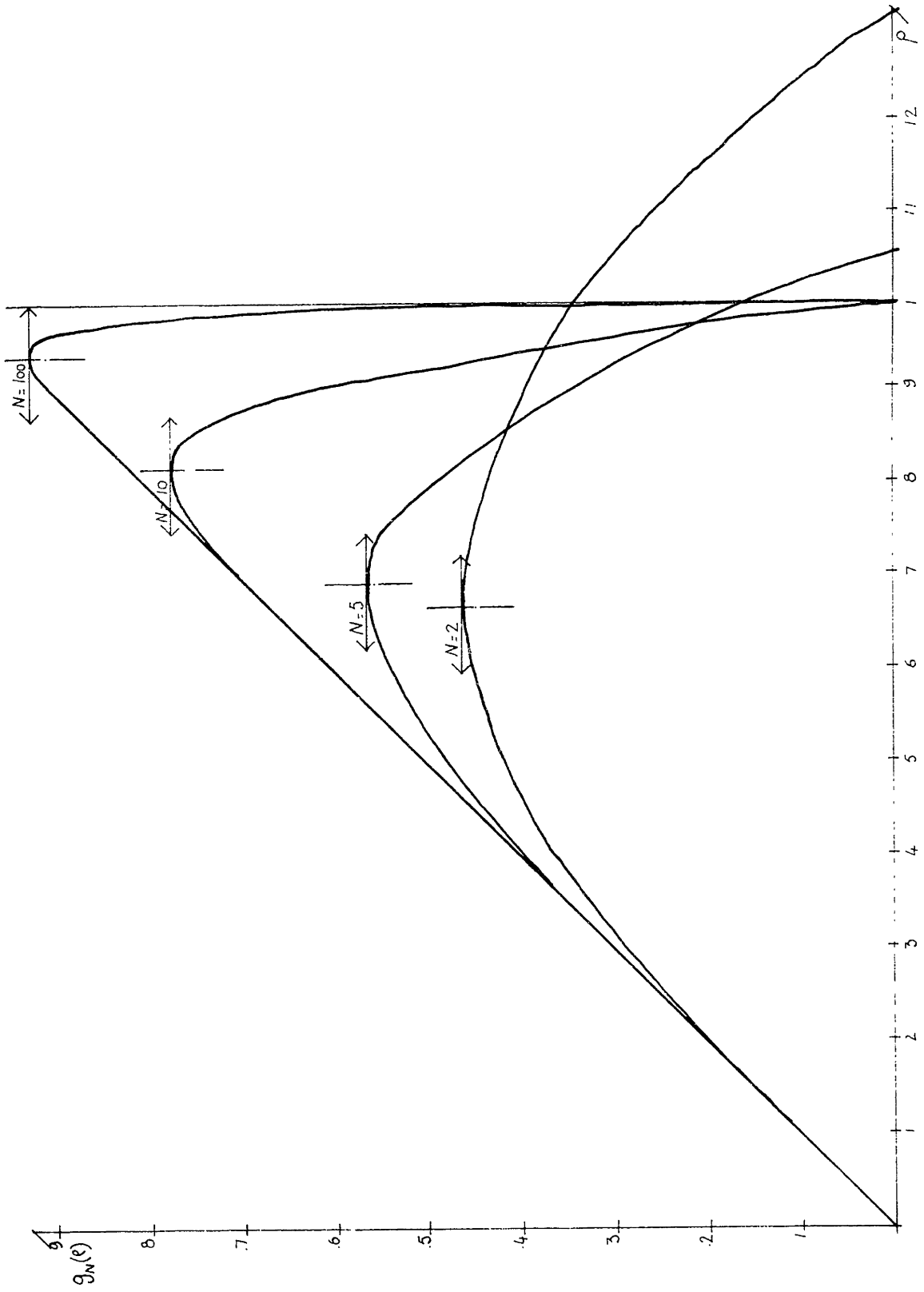


Figure 5

Relative utilities of one unit of capacity for various total capacities ( $g_N(\rho)$ )

operation of the service is such that for a given  $N$ ;  $\rho = \bar{\rho}$  and  $g_N(\rho) = g_N(\bar{\rho})$ . The curve giving  $g_N(\bar{\rho})$  as a function of  $N$  describes the economies of scale in the provision of the service (relative utility of one unit of capacity of the service versus total number of units of provision of the service). This curve is given in Figure 6. However, the percentage of capacity which is still available when the service is operated at its maximum utility level is  $\rho_0 - \bar{\rho}$ ,  $\rho_0$  being the value of  $\rho$  for which the  $g_N$  curve intersects the horizontal axis. It describes the difference between the maximum utility utilization and a zero utility utilization of the system. We can consider it as the inverse of the risk associated with the operation of the system. Let us call flexibility  $F_T$  of the service the quantity  $(\rho_0 - \bar{\rho})$ . Figure 6 shows the variation with  $N$  of the flexibility.

Assuming that we operate the service at its maximum utility level, we see that economies of scale tend to become insignificant for large values of  $N$  whereas the flexibility approaches zero. Moreover, not only do we see that the risk becomes very high for large values of  $N$ , but we notice the loss of utility which occurs when the system passes its maximum is much more severe for larger values of  $N$ . Given the fact that in real life it is very difficult to know when we have passed the maximum, the large service is indeed more dangerous than the small one. In the general case, we would be able to define exactly in the same way  $\rho^K$  and  $\rho_0^K$  as being  $\rho$  such that  $g_N^K$  is respectively maximum or zero. Notice that we would have:

$$g_N(\rho_0^K) = \frac{1-K}{1+K} \rho_0^K$$

$\rho_0^K$  is the intersection of  $g_N(\rho)$  with  $g = \frac{1-K}{1+K} \rho$

In the same way:

$$\frac{d}{d\rho} g_N(\rho^{-K}) = \frac{1-K}{1+K}$$

$\rho^{-K}$  is the point of  $g_N(\rho)$  where the tangent is parallel to  $g = \frac{1-K}{1+K} \rho$

Looking at the curves of Figure 5 we see that our conclusions would most likely be exactly the same.

II. We could therefore try an alternate solution and instead of increasing the service rate of a single channel, we could try to improve the capacity of the service by adding channels in parallel. This procedure has some advantages since it reduces the size of the queue in comparison to the number of units in the system. From the point of view that we have chosen for the first part of this chapter, this type of system behaves exactly like the single channel system; the larger the service, the closer the optimum to the breakdown.

In order to arrive at this conclusion, we shall use a certain number of results of queuing theory which can be found directly in Morse's "Queues, Inventories and Maintenance."<sup>(2)</sup> Let us assume that the service can be described by  $M$  parallel channels. Each channel is exponential (mean service rate  $\mu$ ). Units arrive randomly with a Poisson probability density function (mean arrival rate  $\lambda$ ). We characterize the state of the system by the total number of units,  $n$ , present in the system. When  $n$  is smaller than  $M$  there is no queue, when  $n$  is larger than  $M$  there is a queue  $n-M$  long.  $P_n$  describes the probability that the service is in state  $n$ . We

<sup>(2)</sup>Morse, op. cit.



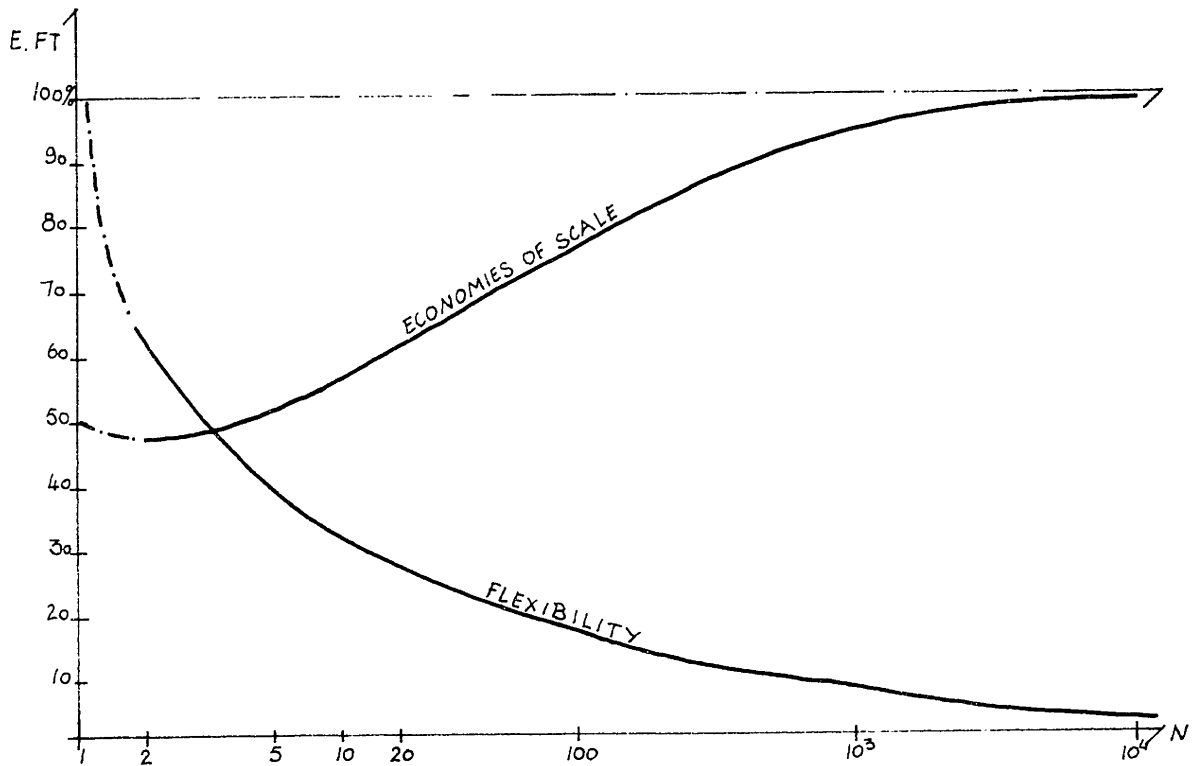


Figure 6

Economies of scale and flexibility as a function of  $N$  (one channel) shall define  $M_\mu$  as being the capacity of the service. It is the equivalent of the quantity  $N/T$  of the first part of this chapter (number of customers which can be served per unit of time). Notice though, as we have already mentioned, that we do not describe the same type of service as in the first part.

Two different cases are possible depending on whether or not we allow a queue to form. If we allow a queue, we must put a penalty on the queue and if we do not allow a queue, we must put a penalty on the rejection of customers.

### 1. Infinite Queue Allowed

Let us define the following functions:

$$E_M(x) = \sum_{n=0, m} \frac{x^n}{n!} e^{-x}$$

$$e_n(x) = \frac{x^n e^{-x}}{n!}$$

$$D_m(x) = E_{m+1}(x) - \frac{x}{m+1} E_m(x)$$

In this particular case where an infinite queue is allowed, we find in Morse\* the probability that the queue is as long or longer than a given length  $K$ .

$$\begin{aligned} Q_{M+K} &= \sum_{n=0}^{\infty} P_{M+K+n} \\ &= \rho^{1K} e_m(\rho^{1M}) / D_{M-1}(\rho^{1M}) \end{aligned}$$

\*Chapter 8, page 104. op. cit. (2)

We are particularly interested in  $Q_M$ , the probability that all  $M$  channels are busy as a function of  $\rho^1 = \frac{\lambda}{M\mu}$ . Figure 7 shows the behavior of  $Q_M$  for several values of  $M$ . (Morse, page 105)

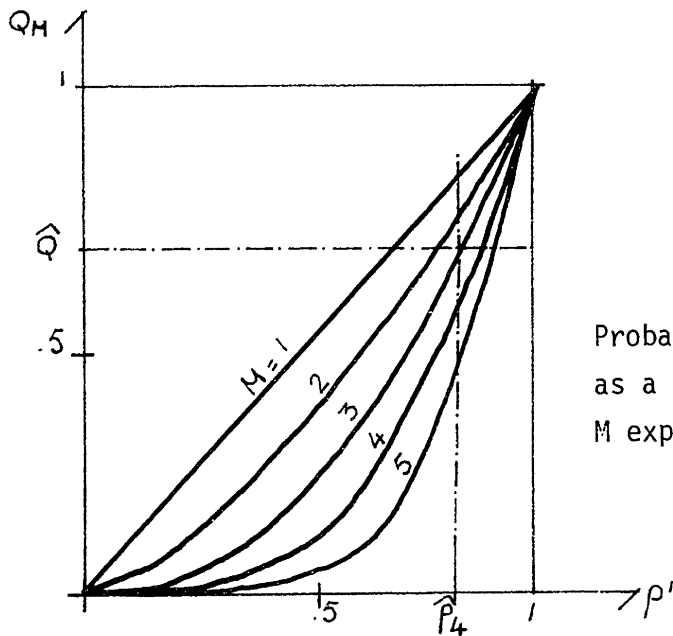


Figure 7

Probability that all  $M$  channels are busy as a function of  $\rho^1$  for Poisson arrivals,  $M$  exponential channels, single queue<sup>(2)</sup>

We notice that as  $M$  is increased,  $\rho^1$  can get nearer to one before the channels become saturated. Let us assume, for example, that the system is operated according to the following policy: the number of channels is such that the customers do not have more than a certain probability  $\hat{Q}$  to wait in queue. In certain cases this probability will be very low (emergency

<sup>(2)</sup>Morse, op. cit.

hospitals); in certain other cases a much larger margin will be accepted. Thus for each  $M, \rho^1$  must be smaller than  $\hat{\rho}_M^1$  as defined on the above figure. The figure immediately shows that as  $M$  increases  $\hat{\rho}_M^1$  gets closer and closer to one, which is the point where the queue becomes infinite. For example if  $Q = .5$  for  $M = 1$ ,  $\hat{\rho}_M^1 = .5$  and for  $M = 10$ ,  $\hat{\rho}_M^1 = .8$ .

We reach conclusions very similar to the conclusions we reached for the single channel system. Let us therefore reformulate the problem in similar terms and assess costs and benefits to the operation of the system. Assume that the utility of providing the service to one customer is  $\alpha$ , and the loss of utility of having one customer waiting in queue is also  $\alpha$ . We mentioned in the first part that having identical figures on both the loss and the profit side was a certain loss of generality but we also explained how to deal with this loss of generality and go back to the general case from a more specific one. The same argument is valid here. We know that the probability that a customer faces a queue is  $Q_M$ , and that the probability that he does not have to face a queue is  $1 - Q_M$ . Therefore, the utility of the system for that particular customer is:

$$u = - Q_M \alpha + (1 - Q_M) \alpha$$

$$u = \alpha [1 - 2 Q_M]$$

the total utility per unit of time is:

$$U = \lambda \alpha [1 - 2 Q_M] = M \mu \alpha \rho^1 [1 - 2 Q_M]$$

Since the capacity of the system is  $M \mu$  per unit of time, the relative utility ( $\alpha = 1$ ) of the service per unit of capacity is:

$$\frac{U}{M\mu} = g_M(\rho^1) = \rho^1 [1 - 2 Q_M]$$

the two following tables give the values of  $Q_M$  and  $g_M(\rho^1)$  for various values of  $M$ .

Table 4 enables us to draw the curves describing the variations of  $g_M$  as a function of  $\rho^1$  for various values of  $M$ . This is done in Figure 8. Notice that the utility of the system can become negative, because we have allowed a queue to form. When the queue becomes large, the system operates at a loss. For  $\rho^1 = 1$ , the queue becomes infinite.

If we assume that the service is operated at its maximum utility level [ $g_M(\rho^1)$  maximum] and if we call  $\bar{\rho}^1$  the corresponding utilization ratio, we have, as in the first part of this chapter, a curve giving the economies of scale ( $g_M(\bar{\rho}^1)$  as a function of  $M$ ) and a curve giving the flexibility of the system  $\rho_0^1 - \bar{\rho}^1$  as a function of  $M$ ,  $\rho_0^1$  being defined in the same way as in the first part of the chapter). These two curves are shown in Figure 9.

Our conclusions are basically the same as for the single channel service: the larger the service, the closer the optimum to the breakdown; economies of scale for large services increase more slowly with the size of the service and the flexibility of the service tends to zero.

However, we must notice two particular features of this second type of system. First, our conclusions are only valid for large values of  $M$ . For small values of  $M$ , the economies of scale are increasing at a constant rate and the flexibility tends to increase. Multichannel services are usually believed to be more stable and less congestion-prone than single channel services and for small values of  $M$  this point argues in the same way.

$M^{\rho 1}$	.1	.2	.3	.4	.5	.6	.7	.8	.9	1	
1	.1	.2	.3	.4	.5	.6	.7	.8	.9	1	
2	.018	.067	.138	.225	.324	.450	.575	.0711	.855	1	
5		.004	.020	.060	.131	.236	.380	.555	.765	1	
10				.009	.036	.103	.292	.618	.668	1	
20					.004	.024	.095	.256	.550	1	
50							.301	.204	.118	1	
	$M^{\rho 1}$										
					.75	.80	.85	.90	.95	1	
					100	.004	.020	.074	.200	.480	1
					200			.023	.095	.365	1

Table 3

Probability  $Q_M$  of waiting in line as a function of the rate of use of the system

$M^{\rho 1}$	.1	.2	.3	.4	.5	.6	.7	.8	.9	1	
1	.80	.120	.120	.080	0	-.120	-.280	-.480	-.720	-1	
2	.96	.173	.217	.220	.176	.06	-.105	-.338	-.550	-1	
5	.1	.198	.288	.352	.369	.317	.168	-.088	-.477	-1	
10	.1	.2	.3	.393	.464	.476	.389	.131	-.302	-1	
20	.1	.2	.3	.4	.496	.571	.567	.398	-.08	-1	
50	.1	.2	.3	.4	.5	.6	.6846	.6608	.245	-1	
	$M^{\rho 1}$										
					.75	.80	.85	.90	.95	1	
					100	.744	.768	.724	.590	.038	1
					200	.75	.8	.811	.729	.257	1

Table 4

Relative utilities as a function of the utilization ratio and the capacity ( $g_M(\rho)$ )

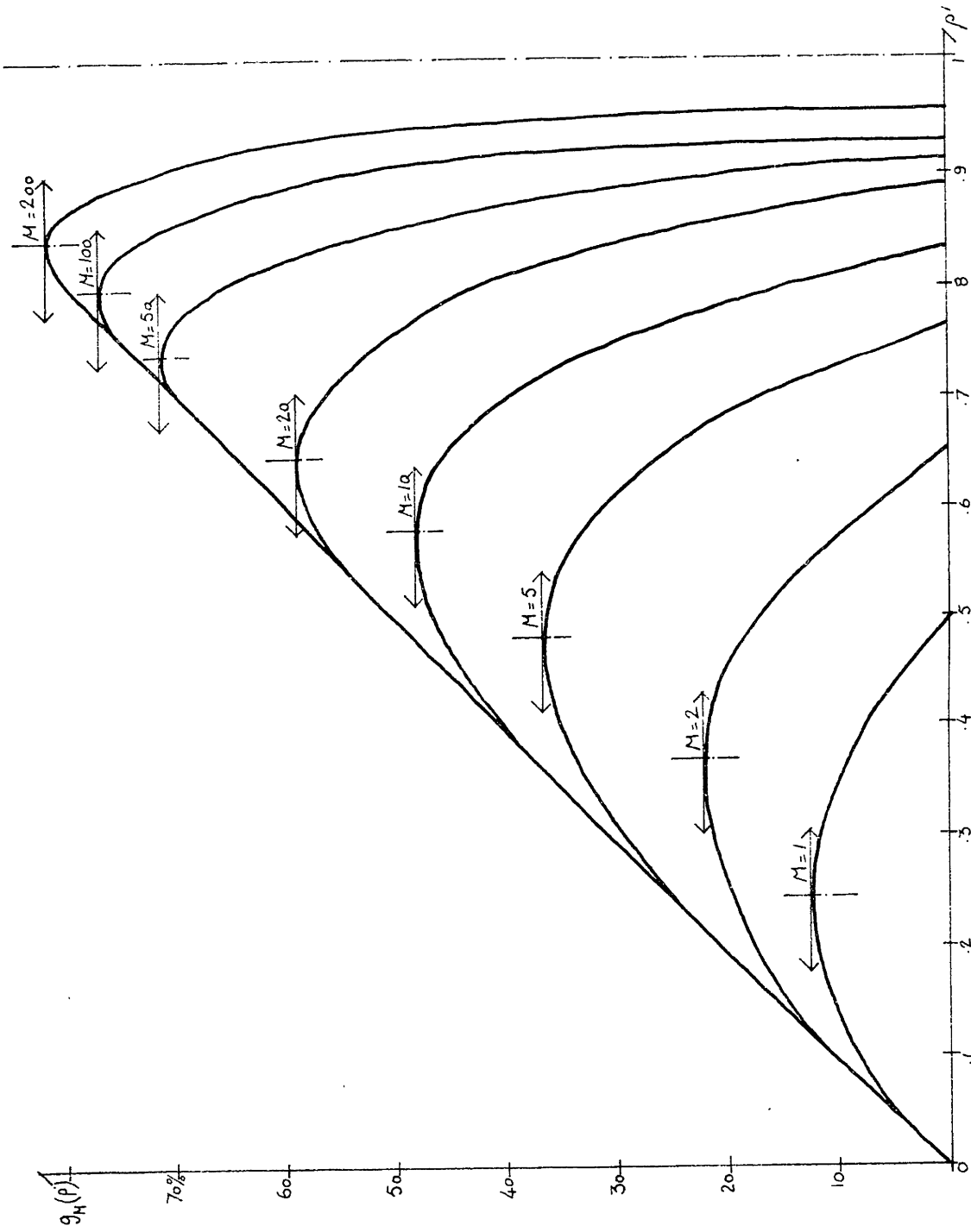


Figure 8

Relative utilities per unit of capacity of the service as a function of the utilization ratio and the number of channels

Second, we can conjecture that the curves obtained in this case are asymptotically identical to the curves obtained in the first part of this chapter. Even though we have no further way of proving this result we can feel on the ground of the central limit theorem that even though each channel has an exponentially distributed service, when we have a very large number of channels, the number of customers served by the system during a given period of time tends to become less and less variable and the multi-channel system tends to become more and more identical to the one channel system described in the first place.

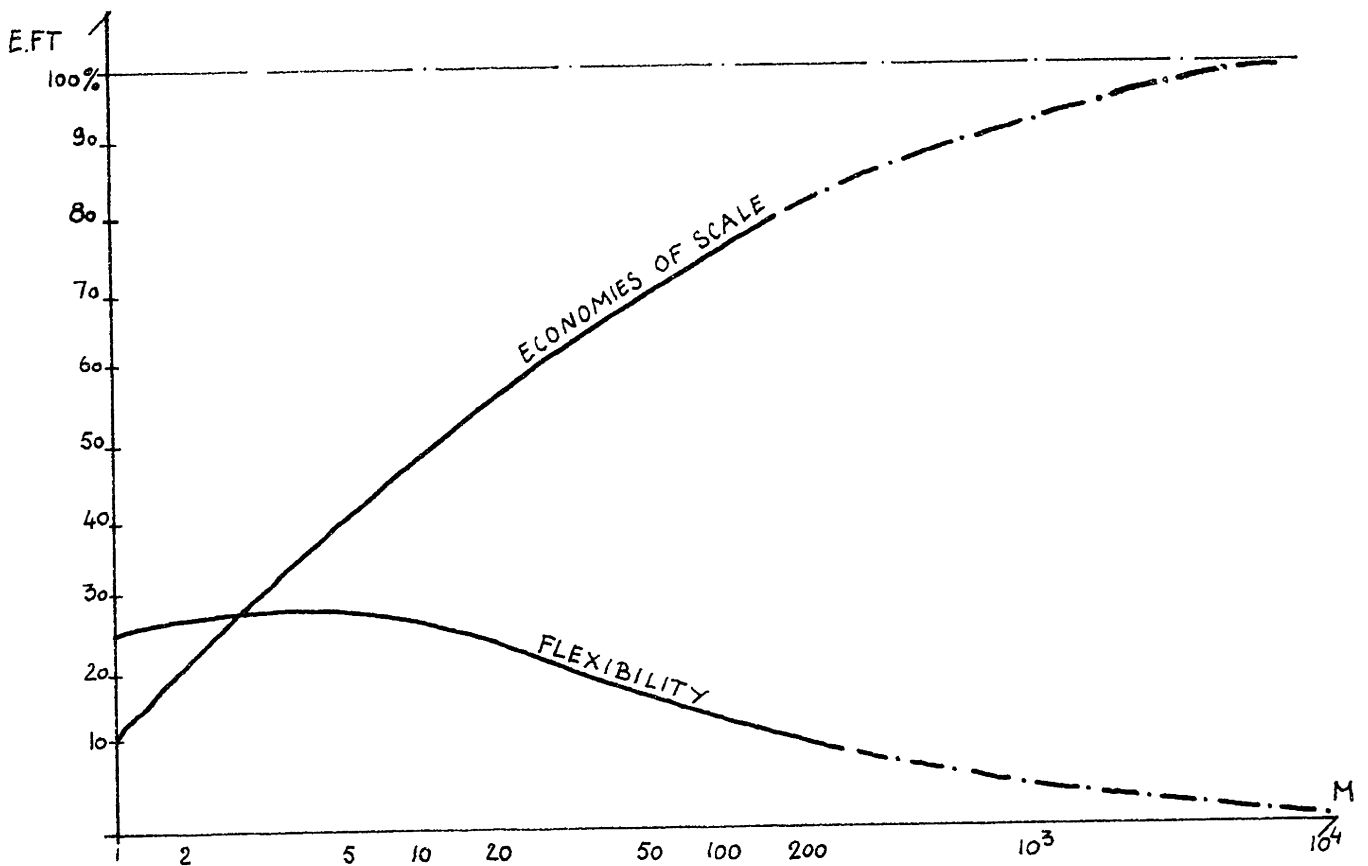


Figure 9

Economies of scale and flexibility of the system as a function of its capacity (Multichannel system)



## 2. No Queue Allowed

$E_M \cdot e_M$  and  $D_M$  being defined in the same way as in the previous case, we are now interested by the probability that a customer is going to be rejected. We find in Morse (Chapter IV, page 31):

$$P_M = \frac{e_M(\rho M)}{E_M(\rho M)}$$

If we substitute  $P_M$  for  $Q_M$ , the rest of the analysis is strictly identical to the infinite queue system. The only distinction lies in the fact that  $\rho^1 = 1$  is not a breaking point of the system. The system breaks down when the losses due to the rejection of customers are equal to the gains accruing because of the provision of the service. However, this model is not likely to represent any real life system. In the first part, we assumed that the service was provided every  $T$  units of time ( $T$  fixed). If  $T$  were large enough, it was therefore possible to assume that a customer who was not served would disappear. In the case we are describing now, when the number of channels becomes large, the service rate of the system becomes infinitely small, and it is therefore most unlikely that a rejected customer will not try again to obtain the service: therefore most of the time the infinite queue model will be more appropriate. Thus, we are not interested in arguing our point on this particular model and we shall not make the computations in this case.

Before concluding this chapter, we should notice that throughout the chapter the closeness to disruption was expressed in relative terms. The distinction between relative and absolute is fundamental to the point we have tried to make. When we want to describe the service or the facility in itself, it is its size or capacity which is important. This is an

absolute variable (number of seats, places, employees...). On the other hand the change of the rate of demand is typically described by relative variables. This is true when we describe the average increase of the rate of the demand for a service (% of increase per year) or when we describe random variations due to non-forecasted events (% of variation). In the same way, if we had mainly focused our discussion on the variability of the service rate, this would have been a relative variable.

At this point we still face several major questions: To what extent does this theoretical analysis apply to real life services? What are the consequences and implications of the analysis? To what extent can we increase the rate of service to face an increasing rate of demand? On the ground of a small statistical analysis we try to answer these questions in Chapter V.

Thus, this chapter has theoretically proved that large services are more fragile than small ones. In this chapter we have considered the noise of real life but, in fact, we shall have to listen to this noise even more carefully in the next chapter.

## Table of Symbols Used in Chapter II

One Channel Service

T	period
N	capacity
$\lambda$	mean arrival rate
p	probability that the numbers of customers is smaller than N $p = f(\lambda T, N) = f_N(\lambda T/N)$
$\alpha$	utility of the service per customer
u	expected utility of the service per customer
U	expected utility of the service for all customers
$g_N$	$g_N(\lambda T/N) = U/N \cdot \alpha$ relative utility of the service per unit of capacity
$\rho = \frac{\lambda T}{N}$	utilization ratio
$\bar{\rho}$	maximum utility utilization ratio
$\rho_0$	zero utility utilization ratio
E	economies of scale
FT	flexibility
$\beta$	loss associated with the rejection of one customer
$K = \frac{\alpha}{\beta}$	ratio utility/disutility

Multichannel Service

$\lambda$  mean arrival rate (Poisson)

$\mu$  mean service rate of one channel (exponential)

$M$  number of channels

$n$  number of units in the system

$P_n$  probability that there are  $n$  units in the system

$\rho^1$  utilization ratio

$Q_k$  probability that there are  $k$  or more units in the system

$$E_k \quad E_k(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} e^{-x}$$

$$e_k \quad e_k(x) = \frac{x^k}{k!} e^{-x}$$

$$D_k \quad D_k(x) = E_{k+1}(x) - \frac{x}{k+1} E_k(x)$$

$\hat{Q}$  given value of  $Q$

$\hat{\rho}_M^1$  for a given  $M$ , maximum value of  $\rho$  so that  $Q_M < \hat{Q}$

$u$  expected utility of the service per customer

$U$  expected utility of the service for all customers

$g_M$   $g_M(\rho^1) = U/M\alpha$  relative utility of the service per unit of provision of the service

$\rho^1$  maximum utility utilization ratio

$\rho_0^1$  zero utility utilization ratio

$E$  economies of scale

$FT$  flexibility

## Chapter III

Further Computations for One Specific Service:  
Buses and the Clumping Problem

In the preceding chapter, we have discussed the problem of increasing the size of urban services assuming that the environment was totally random. This enabled us to evaluate the advantages and the risks associated with an increase of the size of public services. We now want to show by an example that "real" conditions can be even worse than totally random ones and that what we have estimated as being the advantages of larger systems is in fact a very optimistic view of reality. Our example will be a bus system and the reason which makes this system even worse than a purely random one is the so-called "clumping phenomenon"<sup>(1)</sup>. We say that there is clumping, or that two buses are clumped, when they immediately follow each other. If there is no need for extra seating capacity, the second bus is useless. The clumped condition is a most stable one. There is very little chance for anything to happen which could separate two clumped buses. Red lights, pedestrians, cars pulling out, and other events which previously affected the buses independently now affect both of them at the same time. The two buses are running in identical conditions. Even if we allow the buses to pass each other, the first bus will always have to spend a longer time at the stations to pick up passengers than the second

---

(1) Paul F. Bursaux, Scheduling the Buses at Rush Hour (M.I.T.)

bus. This will stick them together. Notice that this phenomenon of clumping is likely to happen along routes where the frequency of vehicles is relatively high<sup>(2)</sup>.

This chapter will be divided in two parts. In the first part, we shall try to evaluate the probability of clumping. In order to do so, we shall use very simple models to describe the two main causes of clumping: the random incidents along the route and the queues at the stations. We shall consider that these two causes are independent, and with the help of our models we shall try to relate the probability of transition from a random distribution along the route to a clumped distribution with the density of traffic and the frequency of buses. In the second part, we shall use the results of the first part to assess utilities to the bus system as a function of the number of buses. As opposed to the purely random systems described in Chapter II, we shall see that, for increasing values of  $N$ , the relative utility of the bus system is not tending towards 100% but towards a lower figure. Thus, in this case, the nonindependence of the variables of the real life problem reinforce the conclusions of our preceding chapter.

### I. Probability of Clumping

#### Clumping Because of Random Incidents Delaying a Running Bus

We want to model the small incidents affecting the buses along the route. We assume that as long as the buses are not clumped the incidents affecting them are independent. For example, if a bus is slowed

<sup>(2)</sup>John Bourne, Towards the Quantification of Transit Scheduling Procedures.

down by a pedestrian, the following bus is not affected. This is not a totally realistic assumption. In fact, if the traffic is seriously slowed down by an incident there can be a certain amount of slowness at the same point long after the incident occurred. Our model might therefore provide for more clumping than there really exists and we might have to restrict ourselves to small incidents.

As soon as the buses are clumped, the incidents affecting them are the same and the buses remain clumped for the rest of the route. The best way to model this is to assume that the number of incidents is large ( $N > 10$  is a very good approximation). Then, we can use the general solution to the problem of the random walk<sup>(3)</sup>. The probability density function of the deviation  $R$  of a bus from its normal position is

$$W(R) = \frac{1}{\sqrt{2\pi} \sqrt{N\sigma^2(\xi)}} e^{-\frac{(R - N E(\xi))^2}{2N\sigma^2(\xi)}}$$

$N$  number of incidents

$E(\xi)$ : expected value of the deviation caused by one incident.

In fact this is a normal distribution,  $N(R; \lambda, \sigma)$

with mean,  $\lambda = NE(\xi)$  and variance,  $N\sigma^2(\xi)$ .

We can easily introduce time.  $N$  here represents the number of incidents.

If we face  $n$  incidents per unit of time and if the vehicle starts at time 0,

we have a normal distribution with mean,  $\lambda = n t E(\xi)$ , and variance,

$$\sigma^2 = n t \sigma^2(\xi).$$

(3) S. Chandrasekar, Stochastic Problems in Physics and Astronomy in Selected Papers on Noise and Stochastic Processes, Nelson Wax (editor), Dover Publications, New York, 1959.

If we now consider the distance between two successive buses, it will also be normally distributed. Assume that the two successive buses start at time 0 (small bias, since the second bus will in fact be submitted to the perturbations of the network only a few minutes later than the first). However, this is more than compensated for by the variability of the departure of the second bus. The headway (time interval) between the two buses will also be normally distributed (difference of two normal variables) with mean  $h$  (theoretical headway between two buses) and with a variance equal to the sum of the variance of the two buses.

$$\sigma_0^2 = 2\sigma^2 = 2nt\sigma^2(\xi)$$

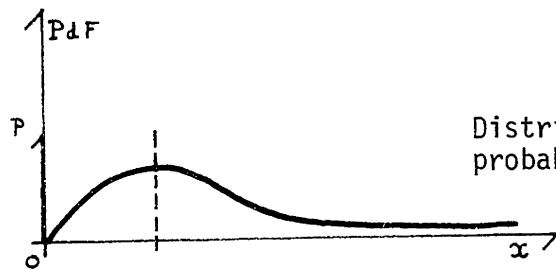


Figure 10

Distribution of headways;  $p$  is the probability of clumping

The headway between the two buses is oscillating like a particle submitted to a Brownian movement<sup>(3)</sup>. However, our model is not complete since we have said that whenever a bus catches up to another one they get stuck and remain so. In the terms of our comparison, there is an absorbing wall located at position 0 on the  $x$  axis (see figure). When the particle hits this wall (i.e., the distance between the two buses becomes zero) clumping occurs.

We find in Chandrasekar<sup>(3)</sup> the probability density function of our particle (distance between the two buses) in the presence of the absorbing wall.

<sup>(3)</sup>Chandrasekar, op. cit.



$$w(x) = N(x; h, 2\sigma^2) - N(x; -h, 2\sigma^2) \quad \text{for } x > 0$$

$$w(x) = 0 \quad \text{for } x < 0$$

In order to find the probability of clumping or in other terms the fraction of the particles absorbed by the wall, we just have to make the difference:

$$1 - w(x)$$

$$\text{since: } N(x; h, 2\sigma^2) = N(-x; -h, 2\sigma^2)$$

$$\text{we have: } p = 2 \int_{-\infty}^0 N(x; h, 2\sigma^2) dx$$

By changing the variables this can also be written:

$$p = 2 \int_{\sigma/\sqrt{2}}^{\infty} \frac{h}{\sigma\sqrt{2}} N\left(\frac{x-h}{\sigma\sqrt{2}}; 0, 1\right) dx \quad (\text{standardized})$$

Table 5 gives the probability,  $p$ , of clumping for a certain number of values of  $\sigma/h$

This simple model shows that as soon as  $\sigma$  is larger than one half of the headway the buses have at least 15% chances to clump; if the variance  $\sigma$  is of the same order of magnitude as the headway, the buses have about one chance out of two to clump. Figure 11 represents the probability of clumping as a function the frequency  $1/h$  for a given variance,  $\sigma$ , or as a function of the variance,  $\sigma$ , for a given frequency. The frequency  $1/h$  describes the number of buses in the system. Twice as many buses means a frequency multiplied by 2.

50.

$\sigma/h$	$h/\sigma$	$h/\sigma \sqrt{2}$	p
.1	10	7.07	$\sim 10^{-9}$
.2	5	3.54	$\sim 10^{-3}$
.4	2.5	1.77	.077
.5	2	1.41	.16
.6	1.66	1.18	.24
.8	1.25	.88	.38
1	1	.71	.48
1.5	.666	.47	.64
2	.5	.35	.73
4	.25	.18	.86
6	.17	.11	.92

Table 5

Probability of clumping as a function of  $\sigma/h$ .

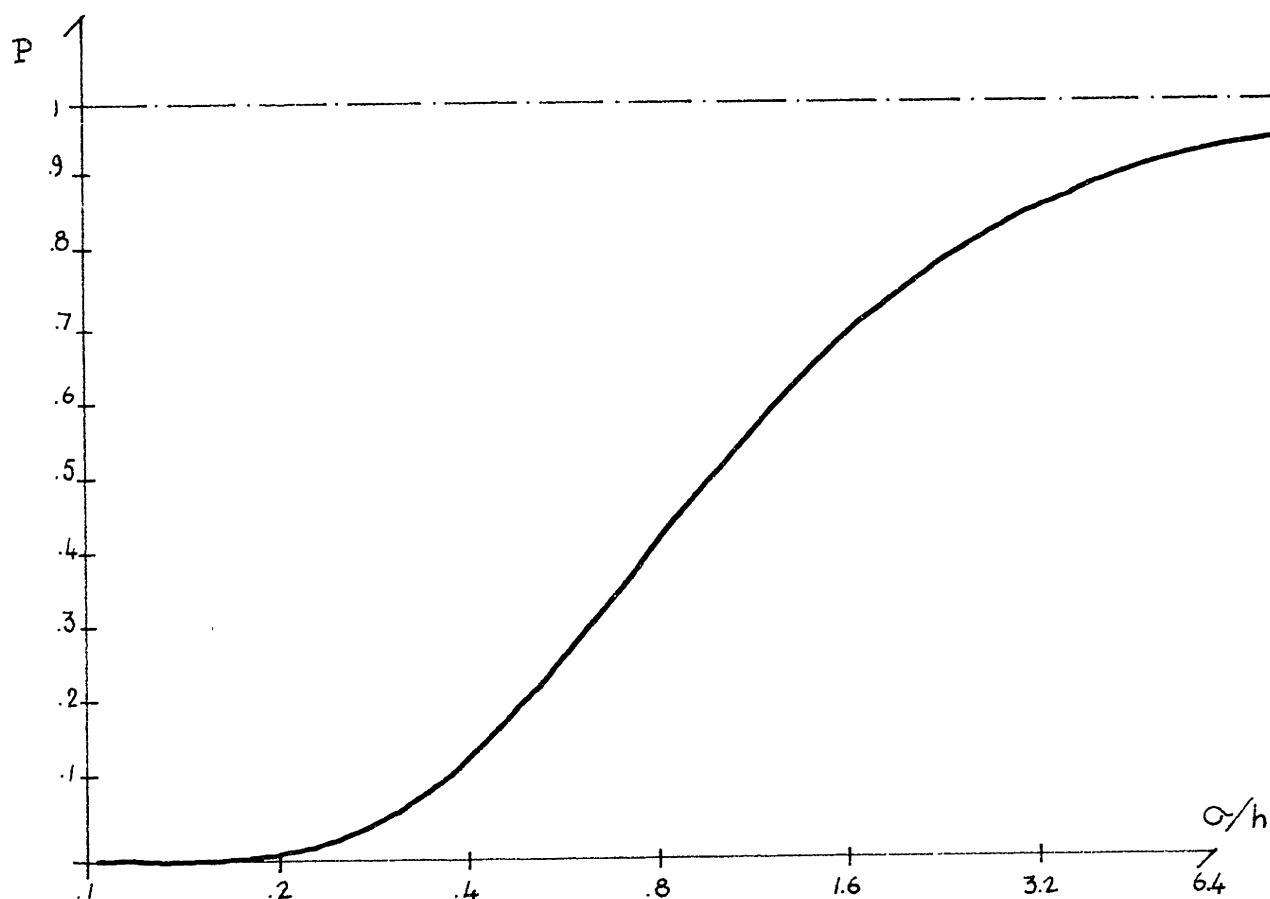


Figure 11

Probability of clumping as a function of the variance and frequency.

We must provide in the model for the observable phenomenon that the probability of clumping increases sharply during the rush hour period<sup>(4)</sup>. If we assume that the number of incidents is proportional to the number of cars in the street and if we assume that the incidents have the same individual effects, then after the same amount of time  $t$  spent in the network

<sup>(4)</sup>Pour les Autobus dans Paris, Paris Project 8, Paris 1970.

the standard deviation  $\sigma = \sqrt{n t \sigma^2(\xi)}$  will be proportional to the square root of  $t$ . A usual approximation is that the peak hour traffic is ten times the average traffic, which introduces a factor of 3. Second, the speed in the network is much smaller. Thus, to reach the same station, the amount of time,  $t$ , taken by the buses will easily be multiplied by a factor of 2. On the other hand the frequency of buses at rush hour time is also currently multiplied by a factor of 3 (headways decreased from 15 to 5 minutes).  $\sigma/h$  can therefore be multiplied by a factor of the order of 20. It is more than enough to hit the potential barrier shown on the figure. Clumping is a common feature of the rush hour period.

#### Clumping Because of the Queues at the Stations

The incidents occurring along the route are not the only causes of clumping. The passengers waiting at the stations introduce a disequilibrium in the system. If a bus is late, the number of passengers waiting for it at the stations will be larger than the average causing the bus to fall even more behind its schedule. The phenomenon is amplified until finally the bus clumps with the bus behind.

Let us assume that the normal headway between two buses is  $h$  and that the rate of arrival of passengers at the station is  $\alpha$  per unit of time. We shall also assume that the time  $S$  necessary to pick up  $N$  passengers at the station is

$$S = a + b N$$

If the headway is constant the normal time a bus will have to stop to pick up passengers is

$$S_0 = a + b \alpha h$$

Let us assume that the bus is delayed by a time  $\delta$ . The time this bus will have to stop is now:

$$S = a + b \alpha (h + \delta)$$

$$S = S_0 + b \alpha \delta$$

The effect of the station is to multiply the lateness by a factor  $(1 + b \alpha)$ . Thus from station to station the lateness of the bus will increase exponentially. After  $i$  stations the lateness of the first bus will be:

$$L_i^1 = \delta (1 + b \alpha)^i$$

The subscript is the number of the station and the superscript the number of the bus. Let us now consider the following bus (bus number 2). For this bus the fact that the preceding one was late leads to a decrease in the number of passengers to pick up. The gain in stopping time is obviously the loss of the preceding bus (i.e.,  $b \alpha \delta$ ). Therefore, this bus will be early by:

$$E_1^2 = b \alpha \delta$$

or indifferently late by

$$L_1^2 = -b \alpha \delta$$

At the second station we shall have

$$L_2^2 = -2 b \alpha \delta (1 + b \alpha)$$

After  $i$  stations the cumulative lateness will be

$$L_i^2 = (-1)^{i-1} 2^i b \alpha \delta (1 + b \alpha)^{i-1}$$

For bus number  $j$ , we could prove in the same way:

$$L_i^j = (-1)^{j-1} C_i^{j-1} (b \alpha)^{j-1} (1 + b \alpha)^{i-j+1} \delta$$

At this point we see that the buses are not independent as we assumed in our former models. Any difference from the schedule is not only amplified as the bus moves along the line but propagated to the other buses. However this tends to die out very quickly because of the coefficient  $b\alpha$ , which is very small as compared to one (ratio arrival rate/number of passengers a bus can load per unit of time).

Let us now consider the first two buses. If the distance between these two buses is the scheduled headway, the fact they go through a station will not change this headway. If it is not the scheduled headway the difference will be multiplied by  $(1 + 2 b\alpha)$ . In the first paragraph we have proved that the distribution of the headways was normal around its mean. Thus, after the station, this distribution will still be normal but with standard deviation:

$$\sigma_1 = (1 + 2 b\alpha) \sigma$$

After  $k$  stations the standard deviation we should consider is

$$\sigma_k = (1 + 2 b\alpha)^k \sigma$$

Thus, the results of the first part remain perfectly valid but at time  $t$

we should not use the standard deviation  $\sigma = \sqrt{nt \sigma^2(\xi)}$  but if we are located at the  $k^{\text{th}}$  station the standard deviation  $\sigma^1 = \sqrt{nt \sigma^2(\xi)} (1+2b\alpha)^k$

On the graph of Figure 11 (logarithmic coordinates) this corresponds to a translation  $T = i \log (1+b\alpha)$

Since  $b\alpha \ll 1, T \sim ib\alpha$

The barrier of potential is even steeper if we consider the stations.

Notice that  $\alpha$  is increased by the rush hour and the very conditions which were already pushing  $\alpha/h$  in the wrong direction.

The main conclusion to be derived from this first part is that the buses are not independent. In order to avoid clumping, we must break down this interdependence or find a feedback control more powerful than the forces creating the disequilibrium. The second part will relate this problem of interdependence of the buses with the scale of the whole system.

## II. Clumping and the Utility of the Bus System

In order to relate the bus system to the systems described in the previous chapters, we must assess a utility to this system<sup>(5)</sup>. Since we are interested in congestion and delays, we shall define this utility as a function of the waiting time for the buses. Therefore, we shall first compute the expected waiting time for the buses under various assumptions.

### Expected Waiting Times

Let us assume that the length of the line is  $\ell$  and that there are  $n$  buses along the line.

#### 1. Perfect Scheduling, Headways Constant

If the distance between the buses is constant it will be equal to  $\ell/n$ . If  $v$  is the average speed of the buses the time interval between two buses will be  $\ell/n V$ . The expected waiting time of a passenger arriving randomly at a station is:

$$E(W_1) = \frac{\ell}{2nV}$$

<sup>(5)</sup>Choix Entre transports publics et transports individuels en Region Parisienne. Cahiers de l'IAURP numero 26, Paris, 1969.

## 2. Purely Random Buses

Let us assume that the buses are distributed in a purely random fashion along the line (flat probability density function).

If a random passenger arrives at time  $t$  at a station located at the end of the line (this does not introduce any loss of generality since we could divide the line and consider that any station is the end of a part), then the probability density function describing the position of the next bus which will arrive is<sup>(6)</sup>:

$$f(x) = n \left( \int_0^x 1/\ell \, dx \right)^{n-1} / \ell$$

$$f(x) = n(x/\ell)^{n-1} / \ell$$

therefore the expected waiting time will be

$$E(W_\ell) = \int_0^\ell \frac{(\ell-x)n(x/\ell)^{n-1}}{\ell v} \, dx$$

$$E(W_2) = (\ell/v) - \frac{n\ell}{(n+1)v}$$

$$E(W_2) = \frac{\ell}{(n+1)v}$$

$$\text{if } n = 1 \quad E(W_1) = E(W_2)$$

$$\text{if } n = \infty \quad E(W_1) = \frac{1}{2} E(W_2)$$

In the same way we can compute the variance of  $W_2$

<sup>(6)</sup>Harold Freeman, Introduction to Statistical Inference, Addison Wesley, Reading, Massachusetts.



$$E(W_2^2) = \int_0^{\ell} \frac{(\ell-x)^2}{v^2} \frac{n}{\ell} \left(\frac{x}{\ell}\right)^{n-1} dx$$

$$E(W_2^2) = \frac{2\ell^2}{v^2(n+1)(n+2)}$$

$$\sigma^2(W_2) = \frac{\ell^2}{v^2} \left[ \frac{2}{(n+1)(n+2)} - \frac{1}{(n+1)^2} \right]$$

$$\sigma^2(W_2) = \frac{n}{n+2} \frac{\ell^2}{v^2(n+1)^2}$$

### 3. Clumped Buses

If  $i$  clumps occur, the service is equivalent to a service with  $n-i$  random buses along the line. The conditional average waiting time is:

$$E(W_{3i}) = \frac{\ell}{v(n-i+1)}$$

$$\frac{1}{n-i+1} = \frac{1}{n+1} \frac{1}{1-\frac{i}{n+1}} = \frac{1}{n+1} \sum_{k=0}^{\infty} \left(\frac{i}{n+1}\right)^k$$

The number of clumped buses is the outcome of a Bernoulli trial ( $n$  buses each of them having a probability  $p$  of clumping). This can be written as:

$$p(i \text{ clumps}) = \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i}$$

therefore:

$$E(W_3) = \frac{\ell}{v(n+1)} \sum_{k=0, \infty} \sum_{i=0, n} \left(\frac{i}{n+1}\right)^k n! \frac{p^i (1-p)^{n-i}}{i!(n-i)!}$$

We do not know how to compute this series but we can find a lower bound.

First, notice that since  $n$  is large we can write  $n=n+1$ .

Therefore, we can write:

$$E(W_3) = \frac{\lambda}{v(n+1)} \sum_{k=0, \infty} \sum_{i=0, n} \left(\frac{i}{n}\right)^k n! \frac{p^i}{i!} \frac{(1-p)^{n-i}}{(n-i)!}$$

Now we want to prove that:

$$\begin{aligned} \left(\frac{i}{n}\right)^k \frac{n!}{i!} &\geq \frac{(n-k)!}{(i-k)!} \\ \frac{n!}{(n-k)! n^k} \frac{(i-k)! i^k}{i!} &\geq 1 \\ \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{i^k}{i(i-1)\cdots(i-k+1)} &\geq 1 \\ \frac{1(1-\frac{1}{n})\cdots(1-\frac{k+1}{n})}{1(1-\frac{1}{i})\cdots(1-\frac{k+1}{i})} &\geq 1 \end{aligned}$$

$$\text{Since } i \leq n \quad \frac{\lambda}{i} \geq \frac{\lambda}{n} \quad 1 - \frac{\lambda}{i} \leq 1 - \frac{\lambda}{n}$$

The fraction is the product of  $k$  fractions individually larger than 1. QED

Dropping the first  $k$  elements of the second summation we can now write:

$$E(W_3) \geq \tilde{E}(W_3) = \frac{\lambda}{v(n+1)} \sum_{k=0}^{\infty} p^k \sum_{i=k}^n (n-k)! \frac{p^{i-k}}{(i-k)!} \frac{(1-p)^{n-i}}{(n-i)!}$$

which trivially reduces

$$\tilde{E}(W_3) = \frac{\lambda}{v(n+1)} \sum_{k=0}^{\infty} p^k$$

$$\tilde{E}(W_3) = \frac{\lambda}{v(n+1)} \frac{1}{1-p}$$

Since for any  $n$  larger than one,  $\frac{n}{n+2} \geq \frac{1}{3}$

we have:

$$\sigma^2(W_3 i) \geq \frac{1}{3} \frac{\ell^2}{v^2} \frac{1}{(n-i+1)^2}$$

$$\sigma(W_3 i) \geq \frac{\sqrt{3}}{3} \frac{\ell}{v} \frac{1}{n-i+1}$$

and the same computations as the one leading to  $\tilde{E}(W_3)$  would lead to a lower bound estimate of  $\sigma(W_3)$

$$\sigma(W_3) \geq \tilde{\sigma}(W_3) = \frac{\sqrt{3}}{3} \frac{\ell}{v(n+1)} \frac{1}{1-p}$$

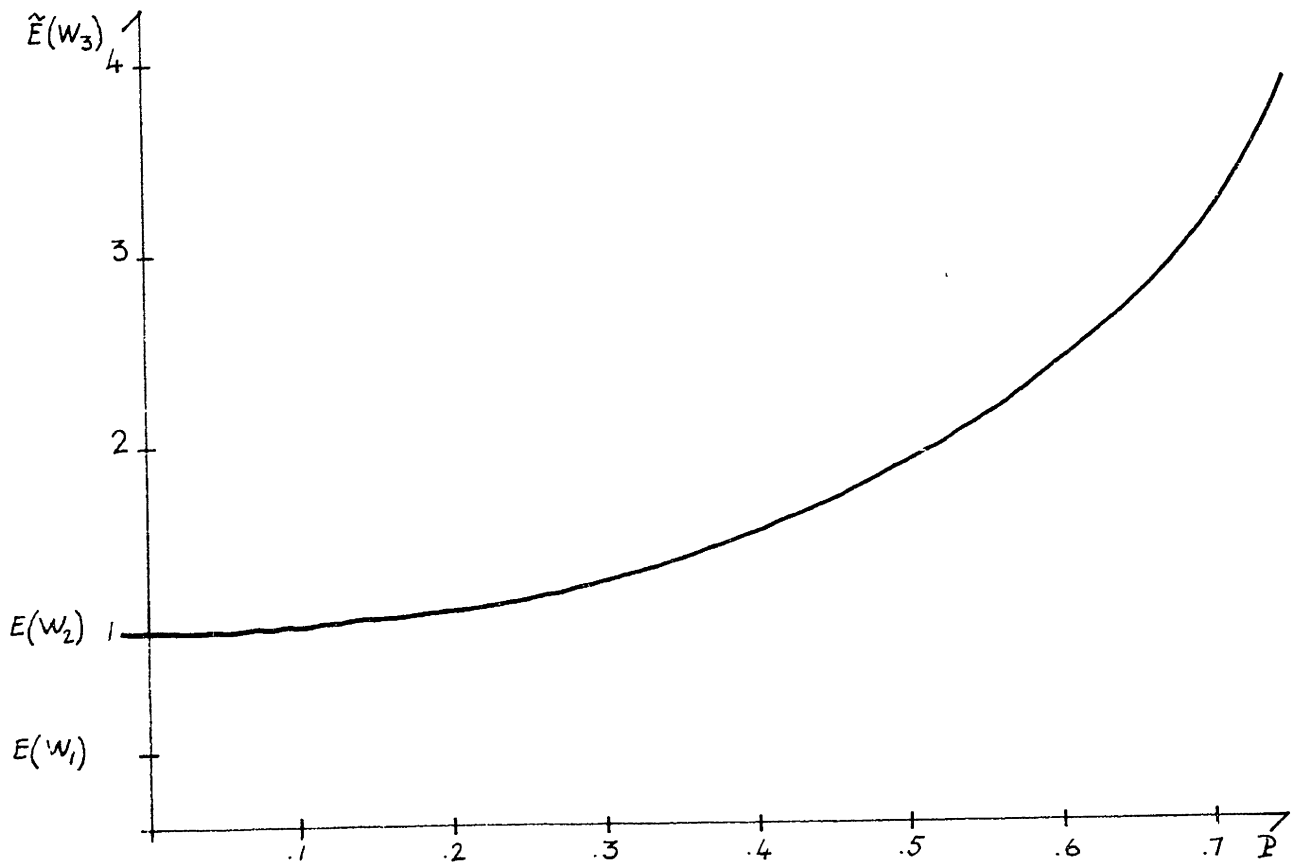


Figure 12

$\tilde{E}(W_3)$  as a function of the probability of clumping

These results are shown in Figure 12 which describes  $E$  as a function of  $p$ . We see that the expected waiting time tends to  $\infty$  when  $p$  tends towards 1. Since  $p$  is a function of the number of buses used on the particular route, we now want to find the variation of the expected waiting time when we increase the number of buses. We shall use our estimate of  $p$  as a function of  $\sigma/h$  and the fact that  $n h = \text{cst}$ . We shall use our lower bound of  $W_3$  as an estimate of the expected waiting time and assume that  $n$  is large, so that  $n = n + 1$ . We shall also assume that with  $N_0$  buses used on the route,  $\sigma/h = .1$ , and the expected waiting time is  $W_0$ . Then the first table of the

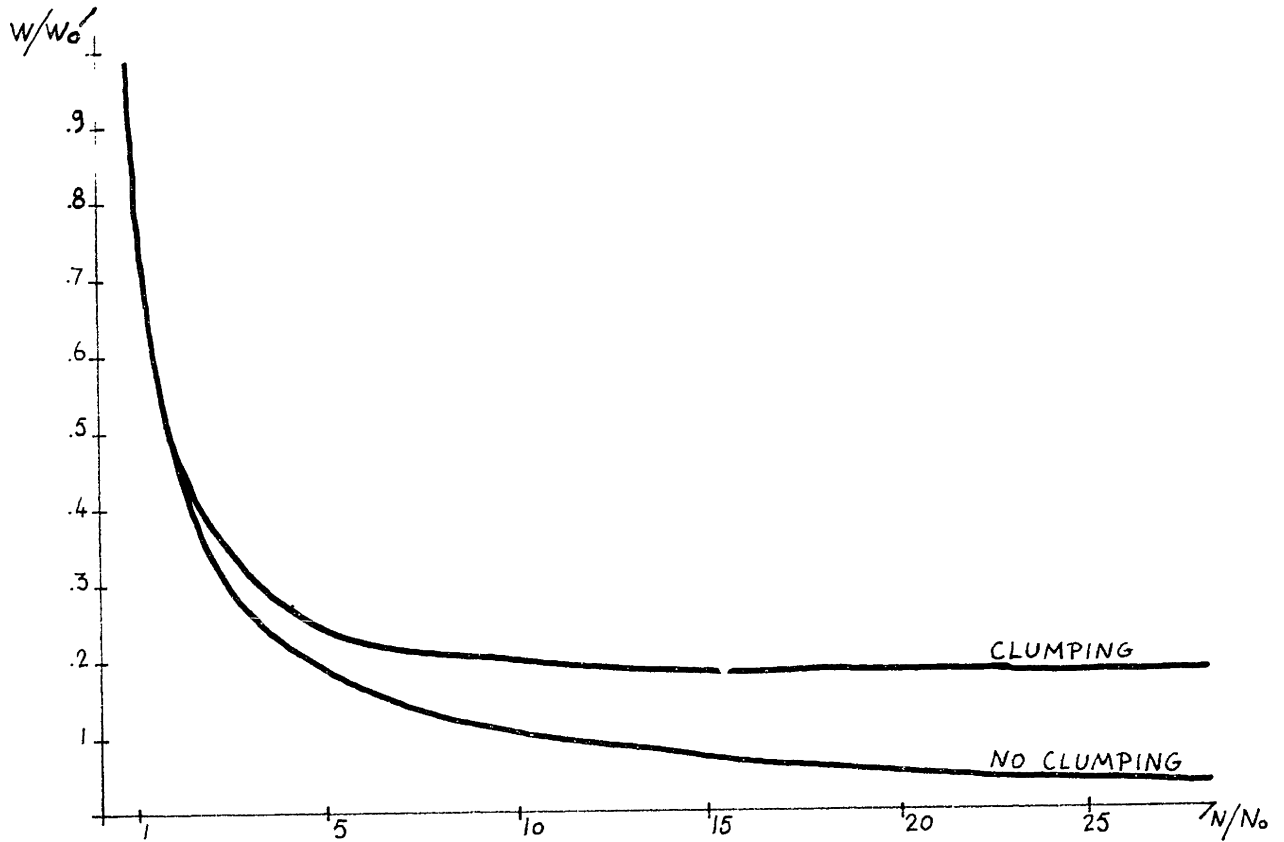


Figure 13

Waiting time as a function of the number of buses in the system

$n/N_0$	$\sigma/h$	$P$	$w/w_0$
1	.1	$10^{-4}$	1
2	.2	$10^{-3}$	.5
4	.4	.077	.275
5	.5	.16	.240
6	.6	.24	.220
8	.8	.38	.201
10	1	.48	.192
15	1.5	.64	.185
20	2	.73	.182
40	4	.86	.180
60	6	.92	.180

Table 6

Waiting time as a function of the number of buses in the system

chapter and the formula giving  $E(W_3)$  enable us to compute the table which is illustrated by Figure 13.

Notice that even with an infinite number of buses the waiting time does not become zero.

### Utilities

The utility of increasing the number of buses is difficult to assess since the computations of this chapter can only be expressed in relative terms. We have only been able to give the waiting times as a function of a reference waiting time and in the same way we shall only be able to define utilities as a function of the utility of a reference system. Let us define by  $N_1$  the number of buses in the reference system and by  $W_1$  the corresponding waiting time. By definition  $N_1$  is arbitrary. We can, for example, choose  $N_1 = N_0$ ,  $N_0$  being defined in the same way as in the preceding paragraph, as being the number such that the standard deviation of the buses around their mean is 1/10 of the headway. This is a good approximation for the actual operating conditions of a suburban line (a bus every hour, standard deviation approximately 5 minutes). Given the reference level, it becomes easy to define the utility of having  $N$  buses in the system as a function of the savings in waiting time:

$$U(N) = f(W_1 - W)$$

If we want to scale the utility between zero and one, we can define the zero level as being the reference level and choose a linear function for  $f$ :

$$U(N) = \frac{W_1 - W}{W_1}$$

In this particular case the marginal utility is equal to the marginal savings in waiting time:

$$\frac{dU}{U} = \frac{d(W_1 - W)}{W_1 - W}$$

In the case,  $N_1 = N_0$ , if we use for  $W$  the same lower bound estimate as in the preceding paragraph,  $U$  is described by the curves of Figure 14. Notice that in the absence of clumping we can obtain any level of utility by increasing the number of buses. With clumping we cannot go beyond the level .82 which corresponds to waiting times of the order of 11 minutes.

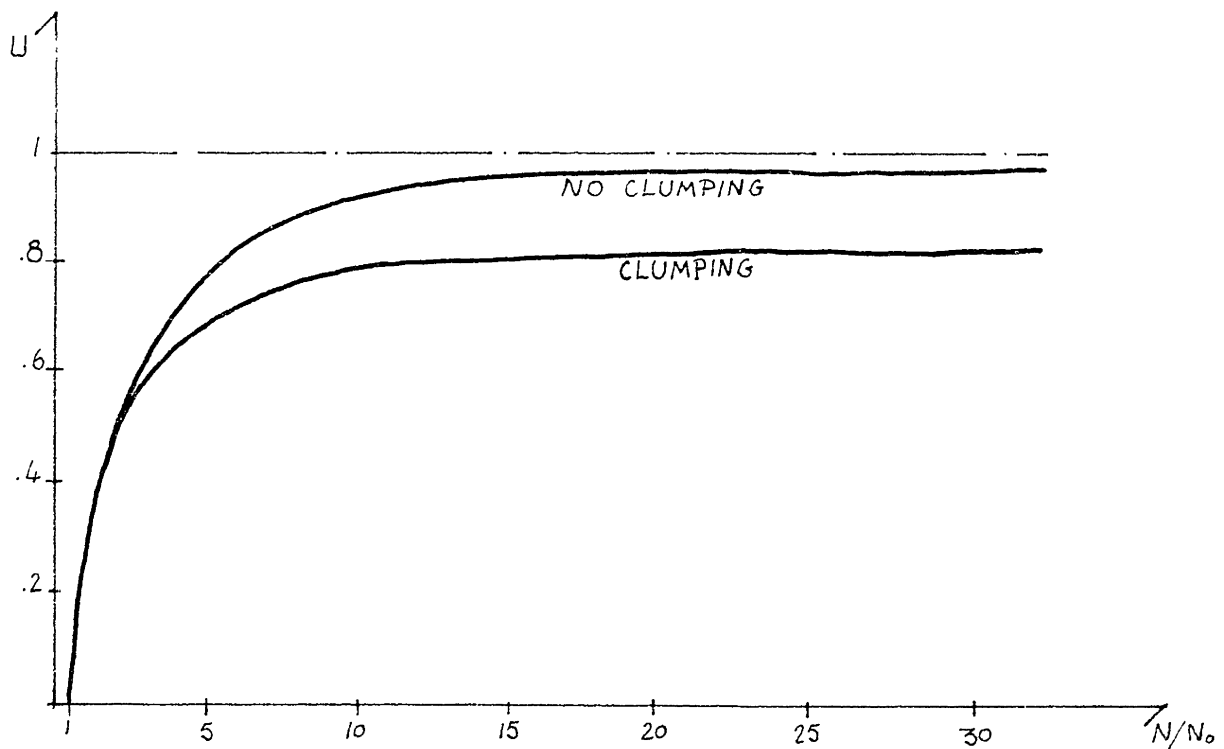


Figure 14

Utility as a function of the number of buses

If we choose a reference  $N_1$  such that  $\sigma/h = .5$  which is a good approximation for a central city line at the non-rush hour time (headways 10 min., standard deviation 5 min.), at the reference level the expected waiting time is 12 min. (2 min. lost because of clumping). With an infinite number of buses the waiting time is still of the order of 9 min. and the maximum level of utility we can reach is only .25. This means that even in non-rush-hour conditions, it is impossible to operate this bus system so that the expected waiting time at the stations is less than 9 minutes. With different figures the same results will be true for any bus system and it is impossible to improve these systems by simply increasing the number of buses.

Some figures describing real systems are given in Chapter V.

However, we can already notice that this example shows that we have been



rather optimistic in the previous chapter with our description of the economies of scale of larger systems. In the case of buses, it seems "a priori" rather easy to reduce the waiting time between buses to any desired level. In real life we are blocked by a structural impossibility. I tend to believe that the same problem is true of other systems as well. Chapter IV tries to formalize this idea.

## Chapter IV

Further Computations for a More General Example the Case of Pollution

In the previous chapter which dealt with clumping, we have shown that buses along a line were not independent and that this actual dependence was a major cause of trouble for the system. We are therefore interested in considering another example very easy to generalize of a congestion process very similar to the clumping phenomenon. This chapter somehow escapes a major difficulty, which is to know whether service rates increase or decrease when congestion begins to appear. In the case of buses we have proved it was decreasing; here we assume it is decreasing too, but we do not prove it. This is certainly the case of many systems which range from the welfare system to the ecosystem (absorption of pollutants), but since it is only an assumption it is not enough to show that our second chapter is always a conservative description of reality. This chapter is therefore more an exploration of the congestion process of services than a general statement. Thus, it is a complement to Chapter II. But the more general discussion is only resumed in Chapter V.

It is easier to describe a model by reference to an example. The actual concern about pollution<sup>(1)</sup> leads us to choose pollution as the reference example. We can assume that the pollutant consists of discrete units emitted in a Poisson fashion ( $\lambda$  units of pollutant emitted per unit of time). Let us describe the absorption process of the pollutant in a

<sup>(1)</sup>Edwin S. Mills, Economic Incentive in Air Pollution in Harold Wolozin Ed. The Economics of Air Pollution (New York, W.W. Norton and CO.).

little more careful fashion. We shall assume that if there is only one unit of pollutant in the atmosphere it will disappear after a period of time  $t$ ,  $t$  being a random variable exponentially distributed:

$$f(t) = \mu e^{-\mu t}$$

$\mu$  : absorption rate ( $\mu$  units per unit of time).

We shall now assume that if a second unit of pollutant is emitted while the first one is not yet absorbed, the absorption time will be  $K$  times larger ( $K > 1$ ). In other terms the pollutant disappears after a time  $t$  randomly distributed according to the probability density function:

$$f(t) = \frac{\mu}{k} e^{-\frac{\mu}{k} t}$$

If a third unit is emitted while the first two are still present assume that the absorption time is multiplied by  $K^2$  etc... We shall call  $K$  the amplification factor of the system.

Assume that at time zero one unit of pollutant is emitted. Thus the absorption time for this first unit of pollutant will be  $t$  distributed with probability density function:

$$f(t) = \mu e^{-\mu t}$$

The emission time for the second unit of pollutant will be  $t'$  distributed with probability density function

$$f(t') = \lambda e^{-\lambda t'} \quad (\text{Poisson process})$$

The probability that  $t'$  is larger than  $t$  for any given  $t$  is:

$$\int_t^{\infty} \lambda e^{-\lambda t'} dt'$$

Thus the overall probability that  $t'$  is larger than  $t$  is:

$$p = \int_0^{\infty} \mu e^{-t\mu} \int_t^{\infty} \lambda e^{-\lambda t'} dt' dt$$

$$p = \int_0^{\infty} \mu e^{-(\lambda+\mu)t} dt$$

$$p = \frac{\mu}{\lambda+\mu}$$

Therefore with probability  $p = \frac{\mu}{\lambda+\mu}$  the absorption time of the second unit will be  $t$  distributed according to the probability density function

$$f(t) = \mu e^{-\mu t}$$

and with probability  $(1-p) = \frac{\lambda}{\lambda+\mu}$  the absorption time of the second unit will be  $t$  distributed according to the probability density function:

$$f(t) = \frac{\mu}{k} e^{-\frac{\mu}{k} t}$$

Now, if we are in the second case, we could prove exactly in the same way that the probability that the third unit will be emitted while the first two are still present is

$$P_2 = \frac{\lambda}{\lambda + \mu/k}$$

In a more general way, given the fact that  $i$  units are present in the system, then the probability that the  $(i+1)$ th unit will be emitted while  $i$  units are still present is

$$P_i = \frac{\lambda}{\lambda + \mu/k} i^{-1}$$

We therefore have the following scheme for one unit<sup>(2)</sup>

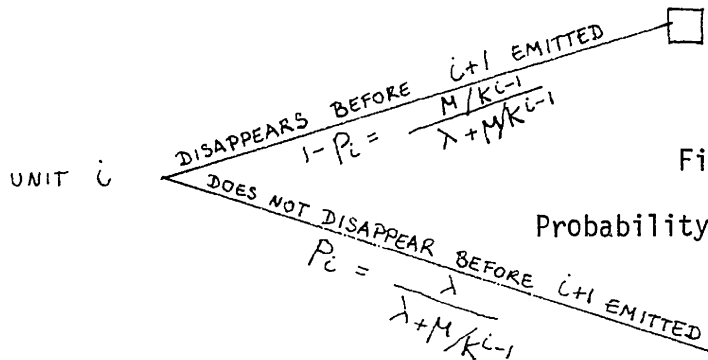


Figure 15

Probability tree for one unit

and for the whole system:

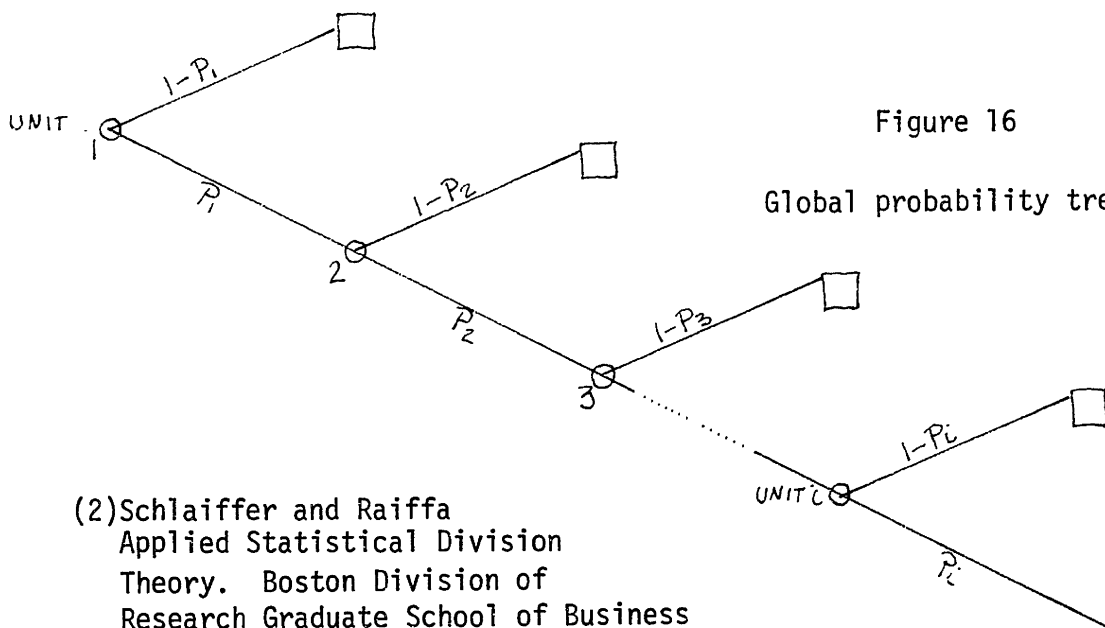


Figure 16

Global probability tree

(2) Schlaiffer and Raiffa  
 Applied Statistical Division  
 Theory. Boston Division of  
 Research Graduate School of Business  
 Administration, Harvard University, 1961.

Thus we are interested in the product

$$\Pi = p_1 \cdot p_2 \cdots p_i \cdots$$

$$\Pi = \frac{\lambda}{\lambda + \mu} \cdot \frac{\lambda}{\lambda + \mu/k} \cdots \frac{\lambda}{\lambda + \mu/k^i} \cdots$$

$\Pi$  is the probability that the units of pollutant will accumulate indefinitely.

Let us write  $\rho = \frac{\lambda}{\mu}$ .

$$\rho = \text{absorption ratio} = \frac{\text{emission rate}}{\text{absorption rate}}$$

$\rho$  can vary from zero to infinity and we are interested in defining  $\Pi$  as a function of  $\rho$

$$\Pi = \frac{1}{1 + \frac{1}{\rho}} \times \frac{1}{1 + \frac{1}{\rho k}} \times \frac{1}{1 + \frac{1}{\rho k^2}} \times \frac{1}{1 + \frac{1}{\rho k^3}} \cdots \frac{1}{1 + \frac{1}{\rho k^i}} \cdots$$

It is more convenient to express  $\Pi$  in terms of  $\text{Log } \Pi$

$$-\text{Log } \Pi = \text{Log} \left(1 + \frac{1}{\rho}\right) + \text{Log} \left(1 + \frac{1}{\rho k}\right) + \cdots + \text{Log} \left(1 + \frac{1}{\rho k^i}\right)$$

If we consider individually the elements of this sum we have:

$$\text{Log} \left(1 + \frac{1}{\rho}\right) = \frac{1}{\rho} - \frac{1}{2\rho^2} + \frac{1}{3\rho^3} - \cdots$$

$$\text{Log} \left(1 + \frac{1}{\rho k}\right) = \frac{1}{\rho k} - \frac{1}{2\rho^2 k^2} + \frac{1}{3\rho^3 k^3} - \cdots$$

.....

$$\text{Log} \left(1 + \frac{1}{\rho k^i}\right) = \frac{1}{\rho k^i} - \frac{1}{2\rho^2 k^{2i}} + \frac{1}{3\rho^3 k^{3i}} - \cdots$$

$$-\sum_{j=0, \infty} \frac{1}{\rho k^j} < \text{Log } \Pi < \sum_{j=0, \infty} \frac{1}{\rho k^j} + \sum_{j=0, \infty} \frac{1}{2\rho^2 k^{2j}}$$

$$-\frac{1}{\rho} \frac{1}{1 - \frac{1}{k}} < \text{Log } \Pi < \frac{1}{\rho} \frac{1}{1 - \frac{1}{k}} + \frac{1}{2\rho^2} \frac{1}{1 - \frac{1}{k^2}}$$

which reduces to

$$e^{-\frac{1}{\rho} \frac{k}{k-1}} < \Pi < e^{-\frac{1}{\rho} \frac{k}{k-1}} + \frac{1}{2\rho^2} \frac{k^2}{k^2-1}$$

Therefore the probability that the units of pollutant will continuously accumulate in the atmosphere is a strictly positive number and a lower bound for this probability is

$$I = e^{-\frac{1}{\rho} \frac{k}{k-1}}$$

where  $\rho$  is the absorption ratio and  $k$  the amplification factor.

An upper bound for this probability  $\Pi$  would be

$$U = e^{-\frac{1}{\rho} \frac{k}{k-1}} + \frac{1}{2\rho^2} \frac{k^2}{k^2-1}$$

Let us therefore consider the variations of  $I$  with  $\rho$  and  $k$ .

First Case I as a function of  $\rho$

$$I(\rho) = e^{-\frac{\alpha}{\rho}}$$

$$\frac{dI}{d\rho} = -\frac{\alpha}{\rho^2} e^{-\frac{\alpha}{\rho}}; \frac{\partial^2 I}{\partial \rho^2} = e^{-\frac{\alpha}{\rho}} \left[ \frac{-\alpha^2}{\rho^4} + \frac{2\alpha}{\rho^3} \right]$$

The inflection points occurs for  $2\alpha = \rho_c$

$$\rho_c = \frac{2k}{k-1} \quad I(\rho_c) = e^{-\frac{1}{2}}$$

Second Case I as a function of k

$$I(k) = e^{-\alpha \frac{k}{k-1}}$$

$$\frac{dI(k)}{dk} = \frac{\alpha}{(k-1)^2} e^{-\alpha \frac{k}{k-1}}$$

$$\frac{\partial^2 I(k)}{\partial k^2} = e^{-\alpha \frac{k}{k-1}} \left[ \frac{\alpha^2}{(k-1)^4} - \frac{2\alpha}{(k-1)^3} \right]$$

$$\frac{\partial^2 I(k)}{\partial k^2} = 0 \quad \text{for } k_c = 1 + \frac{1}{2\alpha}$$



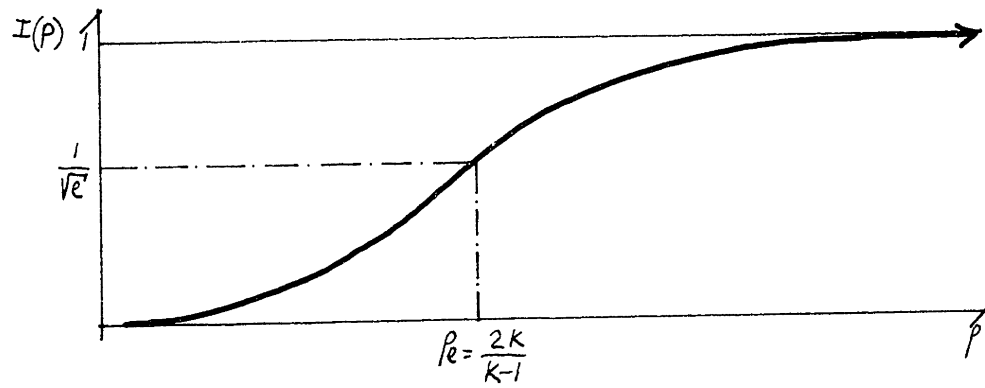


Figure 17

Lower bound of the probability of an infinite accumulation as a function of  $\rho$

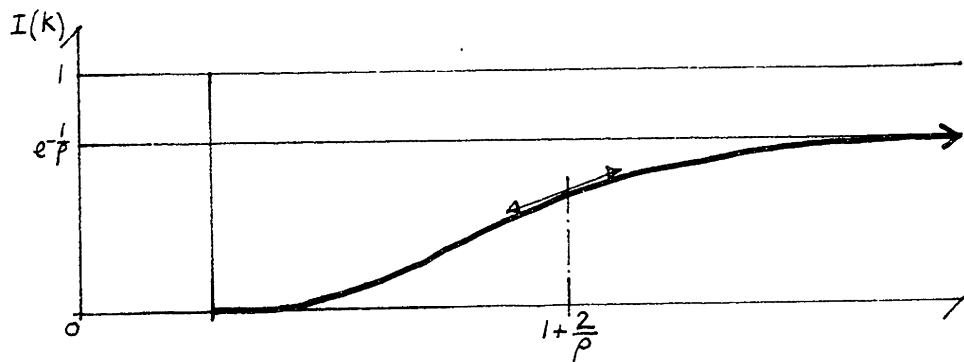


Figure 18

Lower bound of the probability of an infinite accumulation as a function of  $k$

If  $K$  were equal to one, the service rate would be independent of the demand rate, and this model would be a standard M/M/1 queuing model. Thus we would have found the probability of having the pollutant continuously accumulating is zero for any value of  $\rho$  smaller than one. With our description, we find that this probability may well be a strictly positive number. This number is a function of  $K$  and the values of this function are given in Table 7 for various rates of utilization. Thus  $K$  (amplification factor) is the key parameter to the model. It describes how much the service is affected by the presence of several interacting units and its values affect drastically the outcome of the model.

$\rho$	$k=1.1$	$k=1.2$	$k=1.3$	$k=1.4$	$k=1.5$	$k=2$	$k=4$
.5	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$2.5 \cdot 10^{-3}$	$1.8 \cdot 10^{-2}$	$7.4 \cdot 10^{-2}$
.8	$< 10^{-3}$	$< 10^{-3}$	$4.5 \cdot 10^{-3}$	$1.2 \cdot 10^{-2}$	$2.3 \cdot 10^{-2}$	$8.2 \cdot 10^{-2}$	.19
1	$< 10^{-3}$	$2.5 \cdot 10^{-3}$	$1.3 \cdot 10^{-2}$	$3 \cdot 10^{-2}$	$5 \cdot 10^{-2}$	.13	.27
2	$4 \cdot 10^{-3}$	$5 \cdot 10^{-2}$	.11	.18	.22	.37	.52
10	.33	.55	.65	.70	.74	.82	.88

Table 7

Lower bound of the probability of an infinite accumulation

In fact, this model is very easy to generalize and describes the operation of a large number of systems for which the service rates depends on the number of people asking for the service. This is, for example, the well-known problem of maximum flow through a tunnel<sup>(3)</sup> <sup>(4)</sup>. If we let everybody in a tunnel beyond a certain threshold the service rate (flow through the tunnel) decreases for an increasing demand ( $K > 1$ ). A better solution consists of letting people wait at the entrance of the tunnel while maintaining the flow through the tunnel at its maximum value  $K = 1$ ). This has been experimentally proved and is now a common practice.

Thus, the conclusion of this chapter is that we must try to design systems for which  $K$  is as small as possible. There exists a general trade-off between letting everybody use a system, which means an internal control mechanism and letting only a certain number of people use the system controlling it through queues. Queues are easy to operate and this analysis shows that the second solution will often be the best. In order to go beyond this elementary analysis, we would have to study in more detail the various controls used in the operation of various systems. This cannot be done on such a simplified and general level. When the

---

(3)L. D. Edie and R. S. Foote, Traffic Flow in Tunnels. Highway Research Board, Vol. 37, 1958.

(4)L. D. Edie and R. S. Foote, Effect of Shock Waves on Tunnel Traffic Flow. Highway Research Board, Vol. 39, 1960.

domain is more restricted, such as highway transportation, there exist studies of the control of the system, see, for example R. E. Fenton, "On the Flow Capacity of Automated Highways."<sup>(5)</sup> However, our conclusion is at a lower level, since the only point we wanted to make was that the interaction of the various elements of a system are likely to create conditions worse than the ones we describe in Chapter II. However, it argues in the same way, since it proves that several independent system with no interaction between the systems will operate better than a single system, the elements of which interfere with each other.

---

<sup>(5)</sup>J. G. Bender and R. E. Fenton, On the Flow Capacity of Automated Highways, Transportation Science, Vol. 2.

## Chapter V

Statistics, Consequences and Implications of the Analysis

At this point, we felt that trying to build more models to describe the operation of urban services and the way they approach congestion would not add anything to our argument and that it was more interesting to consider to what extent our conclusions apply to real life services. This step is undertaken in the first part of this chapter and naturally leads to the consequences and policy implications of our argument.

We define the utility of a service in terms of the probability of being served versus the probability of having to wait or being denied the provision of the service. Relative to this type of utility function our argument was that the optimum rate of utilization (maximum utility per unit of capacity) and the corresponding utility per unit of capacity were increasing for increasing capacities. We also showed that the peak of the utility curves around the optimum utilization tended to become sharper and sharper. Thus, let us consider for a given service a sample of facilities. We shall assume that the respective rates of utilization of these facilities will be distributed around the optimum rate of utilization and spread according to the sharpness of the maximum of the utility curves. In other words, we assume that any given service is operated at the optimum. This is a very strong assumption, the discussion of which would be a thesis by itself. However, there is ground to argue that oversized capacities will generate the demand which will fill them up, whereas when services are not available people learn to deal without them. The first case would describe certain highways. We know that highways typically generate

new construction and thereby the demand for highway construction. The second case would more typically fit public transportation. Therefore, if our conclusions are correct, we should find an average rate of utilization increasing with increasing capacities of the facilities and a standard deviation of this rate of utilization decreasing for increasing capacity of the facilities.

As a first example of service, we shall consider the provision of health care and, more particularly, one of the major elements, hospital care. There is a rising concern for the ever increasing price of hospital care<sup>(1)</sup>, and therefore there exists a large amount of data related to the problem. We used the official statistics of the department of Health Education and Welfare for the 201 Standard Statistical Metropolitan Area<sup>(2)</sup>. In each of these areas, we consider only general hospitals, and we define the capacity of the hospital system of one metropolitan area as being the total number of hospital beds available in the area. We call utilization ratio  $\rho$  of the system the utilization ratio of the beds, which is a statistical datum. We then define six groups according to the values of  $N$ , and for each group we compute the mean and the standard deviation of  $\rho$ . The results are summarized in Table 8 and in the corresponding figure.

---

(1)Report of The National Conference on Medical Costs. Washington, D.C., June 1967, U. S. Department of Health Education and Welfare.

(2)Hospitals. A County and metropolitan Area Data Book. Public Health Service Publication, U. S. Department of Health Education and Welfare.

Table 8

Mean and standard deviation of the utilization ratio for the six groups of hospitals

$N$	< 400	400 1000	1000 2000	2000 5000	5000 10000	$\geq 10000$
$\bar{p}$	71.5%	76.3%	77.9%	78.88%	77.8%	79.6%
$\sigma_p$	7.55%	7.63%	6.36%	6.02%	5.55%	2.93%

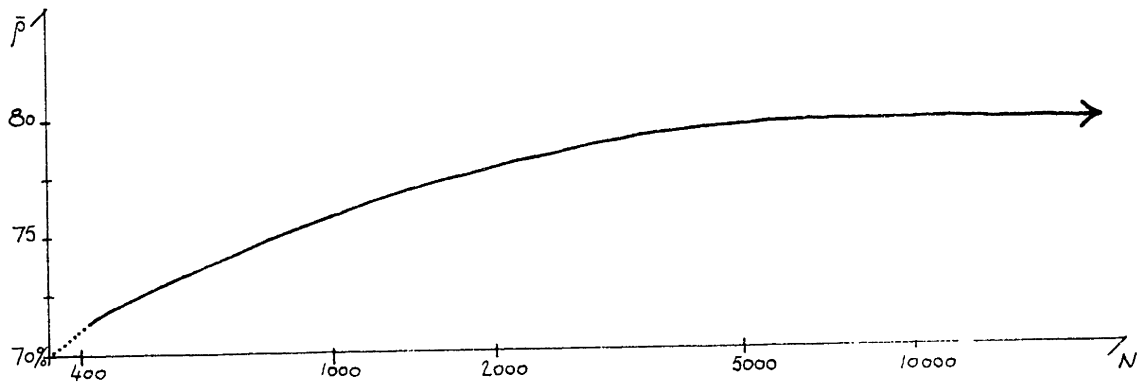


Figure 19

Mean utilization ratio of the hospital beds as a function of the capacity

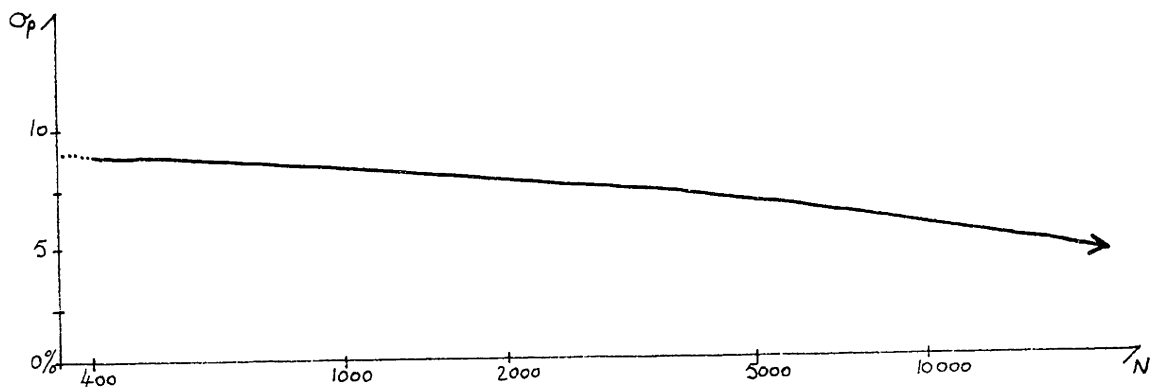


Figure 20

Standard deviation of the rate of utilization of the hospitals beds as a function of the capacity

We note that the average rate of utilization tends to level off at 80%. This is not very surprising, since there probably exists a physical limit to the rate of utilization of hospital beds. Notice that 80% is often listed as the "proper level" of utilization of hospital beds. Apart from this point, we unfortunately notice that our results are not statistically very significant (about .3 level). Since we have not taken into consideration any other factor than capacity, our cross tabulations are nevertheless worth some attention. Notice that the curves which we fitted to the statistical data are very similar to the theoretical curves of Chapter II.

The second example we considered is the solid waste disposal system. We concentrated our attention on one particular facility: the incinerators. Through the Department of Public Health of the Commonwealth of Massachusetts we had access to the data describing the incinerators in the state. There actually exists twenty of them. We defined the capacity  $N$  of an incinerator as being the rated capacity  $N$ , and the utilization ratio as being the ratio of the average daily input to the capacity of the incinerator. We then divided the facilities in three groups according to their capacities and for each group we computed the mean and the standard deviation of the utilization ratio. The results are summarized in Table 9.

$N$	0-150	150-300	300-500
$\bar{\rho}$	.47	.39	.33
$\sigma_{\rho}$	.17	.15	.09

Table 9

Capacities (tons/day), mean and standard deviation of the utilization ratio of the three groups of incinerators of the state of Massachusetts.



This table immediately shows that the rates of utilization are not increasing for increasing capacities as we assumed, but in fact decreasing. We immediately inquired if there were a particular reason for this. It turned out that most incinerators in the state are in the process of being shut down because of the pollution they create. The Department of Public Health actually tries to prevent the communities from using their incinerators. However the maximum rate of utilization authorized by the Department and the effectiveness of the enforcement is variable according to the particular conditions of the area. It is usually not constraining for small incinerators but very constraining for large incinerators, which both create more pollution and are located in more polluted areas. Therefore, if we go back to the curves of Chapter II, we see that the curves corresponding to small capacities will not be changed by the introduction of the constraint. On the other hand, the larger facilities will be operated exactly at the maximum authorized rate.

Our statistical experiment should therefore show the same spread of the utilization ratio for small capacities but hardly any deviation from the mean for large capacities constrained by law. In fact, in the second case we are also measuring the variability of the constraint. Thus, the second measurement is a little more significant for our argument and indeed the results correspond to what we have just said: the larger the facilities, the less variable their rates of utilization.

The third example that we shall use is sanitary sewage. As opposed to solid waste disposal, the problem of sanitary sewage has now been considered for a long time, at least in terms of the amount of sanitary sewage per capita required in a city. We therefore do not need to perform

statistical measurements for this case and we can simply use the results of the various studies which estimate the required size of a sewage system as a function of the size of a city. The results are compiled in Figure 21. Notice that one of the sources is as old as 1918. If we assume that the capacity of the system has to be equal to the maximum flow and if we define the rate of utilization as being the ratio of the average to the maximum flow we immediately see that the rate of utilization approaches a fixed limit exponentially as the size of the system increases, and the difference between the maximum and the minimum flow decreases exponentially as the size of the system increases.

These three examples give some feeling for the relevance of our dissertation. We could certainly choose other services, and more particularly certain facilities providing the service; then, define the capacity of these facilities or the capacity of a system of facilities and the rate of utilization of these facilities and come up with more evidence to support our argument. For example, we could study airports, hotels, trains, fire stations, buses. In fact, for buses it would be more interesting to collect some evidence supporting the argument of Chapter III. On April the 14th 1972<sup>(3)</sup>, we measured the waiting times for buses in front of MIT. The experiment lasted for 150 minutes and we measured the waiting times of the passengers boarding more than 50 buses. We found an average waiting time

---

(3)Paul F. Bursaux, scheduling the buses at rush hour, MIT.

## Quantity of Sanitary Sewage

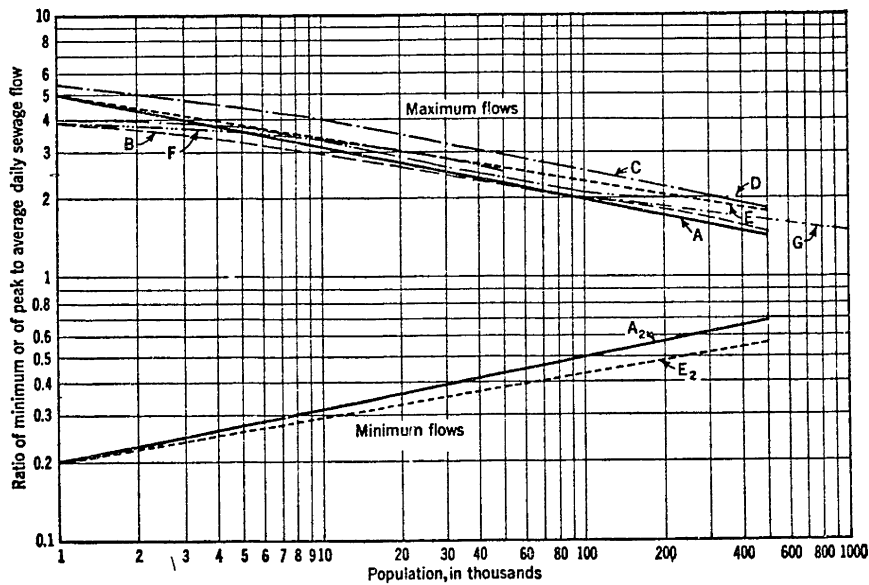


Figure 21

- Curve A source: Babbitt, H. E., "Sewerage and Sewage Treatment," 7th Ed., John Wiley & Sons, Inc., New York (1953).
- Curve  $A_2$  source: Babbitt, H. E., and Baumann, E. R., "Sewerage and Sewage Treatment," 8th Ed., John Wiley & Sons, Inc., New York (1958).
- Curve B source: Harman, W. G., "Forecasting Sewage at Toledo under Dry-Weather Conditions." Eng. News-Rec. 80, 1233 (1918).
- Curve C source: Youngstown, Ohio, report.
- Curve D source: Maryland State Department of Health curve prepared in 1914. In "Handbook of Applied Hydraulics." 2nd Ed., McGraw-Hill Book Co., New York (1952).
- Curve E source: Giffit, H. M. "Estimating Variations in Domestic Sewage Flows," Waterworks and Sewerage, 92, 175 (1945).
- Curve F source: "Manual of Military Construction." Corps of Engineers, United States Army, Washington, D.C.
- Curve G source: Fair, G. M., and Geyer, J. C., "Water Supply and Waste-Water Disposal." 1st Ed., John Wiley & Sons, Inc., New York (1954).

Ratio of extreme flows to average daily flow compiled from various sources.

Reproduced from page 33 of manual "Design and Construction of Storm Sewers" 1969 Edition with permission of the Water Pollution Control Federation and the American Society of Civil Engineers.

of 3.56 minutes. This is 42% more than the theoretical 2.5 minutes which would correspond to a constant headway between the buses. Theoretically this would lead to a probability of clumping of approximately 30%. We counted only 15% of clumped buses.

However, if we want to undertake the same type of measurement for every service, we very rapidly face several major difficulties. First, for certain services, the nature of the service itself is not clear. If we consider, for example, the case of police protection<sup>(4)</sup> we see that there are several components in this service ranging from traffic regulation to crime prevention. But, the nature of the service which is provided by one police force will be different from the service which is provided by another police force. Even though everybody agrees that a police force is necessary, the role and utilization of the police is by no means clear. In those conditions, it becomes almost impossible to define capacity and rate of utilization. This difficulty is not specific to the police case but would probably arise in the same way for any other services such as education or justice.

A second type of difficulty will arise from the specific type of model we have used to describe services. Many services are not provided through one or several isolated facilities, but merely through a network. This is the case of the telephone, which is a system of hierarchical networks. In such conditions, the problem of providing the service will either

---

(4)Richard C. Larson, Urban Police Patrol Analysis, MIT Press, Cambridge, 1972.

depend of the capacity of the facilities located at the nodes of the network (central exchange in the case of the telephone) or of the capacity of the links of the network (telephone lines). If the congested facility is the links of the network our analysis might not apply. We can easily define the capacity and the rate of utilization of one link of a network or, eventually the capacity and rate of utilization of a network between a sink and a source, but globally there is no such thing as a capacity and rate of utilization of a network. For example, we cannot define the capacity of a telephone system. Therefore, if such is the case the analysis we have undertaken will not be valid.

However, it is probably possible to solve the difficulty associated with the network problem by undertaking a different type of analysis which would hopefully lead to the same type of conclusions as ours.

For these reasons our analysis will not be as general as we would like it to be. Let us nevertheless consider its consequences and policy implications. Our argument was that percentage wise the difference between the optimum level of operation and the maximum level of operation of a service tended to become smaller and smaller as the size of the service was increased. But, we also proved that there were economies of scale associated with the provision of services. Thus, there will be an economic incentive to increase the size of services, and indeed this corresponds to what we observe. Therefore, if we do not consider local or isolated events which affect the rate of demand or the rate of service, and which are better taken care of by large services, but if we consider the secular trend in the rate of demand and in the rate of service, we see

that the condition of the large service is much more critical than the condition of the small service. Indeed what we are describing is more a long term problem than a short-term problem; or maybe should we say that it is not a local but a global problem.

Let us, for example, consider the minimal hypothesis that the demand for services increases as the Gross National Product which is about 6% per year. In fact, services are a superior good and the demand for services increases faster than the median income. Let us also consider a given facility and assume that we build new facilities so that their level of operation is optimum five years after construction is decided. If the service is large and its flexibility is below 6% (annual increase of demand) then if there is no new construction in the meantime, one year after being optimum the service will reach the limits of its possibilities. In other words, our facility is optimum after five years and useless after six years. On the other hand a small service will remain almost optimal for a long period of time. This proves that planning the size and capacities of services is a problem, the difficulty of which increases rapidly with the size of the service. For large services, we shall need very accurate predictions of the amount of service which will be required and of the conditions under which we shall be able to operate the service. More particularly, we shall not be able to wait until we observe decaying conditions in the operation of the service to start planning and building new facilities. Any mistake or underestimation will be likely to have disastrous consequences, and we have not taken into consideration the fact that large services require a much larger construction time than small ones.

Further research could be undertaken in two directions. First we have taken only into consideration very briefly the possibilities of increasing the service rates and capacities in order to face increasing demands. We have considered these as given and then computed the value of certain parameters (maximum, optimum) as a function of these capacities and service rates. It probably would be interesting to consider in a general way the relationship between the evolution of the demand and the evolution of the service. Maybe we could consider this as a control problem and find an optimum evolution of urban services.

A second area which could be investigated is the possibility of changing the mode of provision of urban services. We have argued that increasing the size of service systems would provide economies of scale but would be dangerous. Therefore, the question is whether there exists a way to provide services which would enable us to retain the economies of scale without giving up flexibility. Disaggregating the provision of services to several small independent units and thereby giving up the economies of scale might not be the only solution to the problem we have raised. The solution might be in controls, or in different types of queues, or in different types of services. If such is the case, it is urgent to know what these services should be. What about giving up the provision of services altogether?

The picture however is not so dark as we have made it. Teilhard de Chardin assumes that saturation is a precondition for the evolution of systems. The high densities associated with saturation increase the probability of transition to a higher level of organization which is, in a

certain sense, equated to intelligence. The typical example is the one of oversaturated solutions which, beyond a certain threshold, turn instantly to a more organized state (i.e., solid state). We will not go as far as Teilhard and discuss life and intelligence as higher degrees of organization of systems previously without life or intelligence, but we must notice that the idea of organization and levels of organization is most important in the problem we have discussed. Whether we have reached the necessary level of congestion for a possible transition to a higher level of organization is obviously the fundamental question.



Appendix to Chapter V

We suggest a method for studying the behavior of the system if we increase the service rate as fast as the demand. The system studied is the same as the system of the second part of Chapter II (Multi-Channel-Single Infinite Queue System). We shall use the same notations as in Chapter II.

We assume that the service rate is increased as fast as the demand rate. In other words we always increase the number of channels fast enough to maintain a constant utilization ratio  $\rho = \frac{\lambda}{M\mu}$ . We therefore want to compare the M channel system and the M+1 channel system at the same utilization ratio  $\rho$ . Let us therefore compare the probability of queuing in the M channel system and in the M+1 channel system for a constant utilization ratio.

$$Q_M = \frac{e_M(\rho M)}{D_{M-1}(\rho M)}$$

$$Q_{M+1} = \frac{e_{M+1}(\rho(M+1))}{D_M(\rho(M+1))}$$

$$\frac{Q_{M+1}}{Q_M} = \frac{e_{M+1}(\rho(M+1))}{e_M(\rho M)} \frac{D_{M-1}(\rho M)}{D_M(\rho(M+1))}$$

$$\frac{e_{M+1}(\rho(M+1))}{e_M(\rho M)} = \rho e^{-\rho} \left(\frac{M+1}{M}\right)^M$$

$$\frac{D_{M-1}(\rho M)}{D_M(\rho(M+1))} = \frac{M+1}{M} \frac{\int_{\rho M}^{\infty} E_{M-1}(y) dy}{\int_{\rho(M+1)}^{\infty} E_M(y) dy}$$

$$\int_{\rho M}^{\infty} E_{M-1}(y) dy = \frac{M}{M+1} \int_{\rho(M+1)}^{\infty} E_{M-1}(y) dy$$

$$\frac{D_{M-1}(\rho M)}{D_M(\rho(M+1))} = \frac{\int_{\rho(M+1)}^{\infty} E_{M-1}(y) dy}{\int_{\rho(M+1)}^{\infty} E_M(y) dy}$$

$$E_{M-1}(y) = E_M(y) - e_M(y)$$

$$\frac{D_{M-1}(\rho M)}{D_M(\rho(M+1))} = 1 - \frac{\int_{\rho(M+1)}^{\infty} e_M(y) dy}{\int_{\rho(M+1)}^{\infty} E_M(y) dy}$$

$$\frac{D_{M-1}(\rho M)}{D_M(\rho(M+1))} = 1 - \frac{E_M(\rho(M+1))}{\sum_{n=0}^M E_M(\rho(M+1))}$$

$$\frac{Q_{M+1}(\rho(M+1))}{Q_M(\rho M)} = \rho e^{-\rho} \left( \frac{M+1}{M} \right)^M \left[ 1 - \frac{E_M(\rho(M+1))}{\sum_{n=0}^M E_M(\rho(M+1))} \right]$$

$\left(\frac{M+1}{M}\right)^M$  is increasing with  $M$  and tends exponentially towards  $e$  when  $M$  tends towards infinity.

$$1 - \frac{E_M(\rho(M+1))}{\sum_{n=0}^M E_M(\rho(M+1))} \quad \text{is increasing with } M \text{ and tends even faster}$$

than exponentially towards 1 when  $M$  tends towards infinity.

Therefore  $\frac{Q_{M+1}}{Q_M}$  is increasing with  $M$  and tends faster than exponentially towards  $\rho e^{-\rho} e = \rho e^{1-\rho}$ .

We can derive two conclusions from these few pages.

First: If the utility is defined in terms of probability of queuing it is always less interesting to add an additional channel to a large service than to a small one.

Second: If the rate of utilization is close to one, the influence of an additional channel tends to be very small.

Notice too that an alternate approach to this problem of increasing service rates would be to modify our model of Chapter V.

## BIBLIOGRAPHY

I General References

- Michel Crozier, *La Societe Bloquee* aux Editions du Seuil, Paris, 1969.
- Francoise Choay, *Connexions, Immobilieres Constructions*, Paris, 1971.
- Francoise Choay, *Le Sens de la Ville* aux Editions du Seuil, Paris, 1972.
- Jay W. Forrester, *World Dynamics*, MIT Press, Cambridge, Massachusetts.
- Le Corbusier, *La Charte d'Athenes*, Plon, Paris, 1943.
- *Des Villes pour les Hommes, Economie et Urbanisme*, Bineau Caillot, Paris.
- *Sept ans de Vie de la Region Parisienne et de son District*, Cahiers de l'IAURP, Number 16, July, 1969, Paris.
- N. Wiener, *Cybernetics*, Cambridge Technology Press, Cambridge, Massachusetts, 1948.
- Lefebvre Henri, *La Revolution Urbaine*, Gallimard, Paris, 1970.
- Levy Strauss Claude, *Tristes Tropiques*, Plon, Paris, 1955.
- Franz Kafka, *The Trial*.
- Jencks C. and Riesman D., *The Academic Revolution*, Anchor Books, Double Day & Co., New York, 1969.
- Kerr Clark, *The Uses of the University*, Harper and Row, New York, 1963.
- Callahan Raymond E., *Education and the Cult of Efficiency*, University of Chicago Press, Chicago, 1962.

## II. Economics

- Ganz, Alexander, Emerging Patterns of Urban Growth and Travel, HRR, Number 229, 1968.
- Quandt, Richard E. (editor), The Demand for Travel Theory and Measurement, D. C., Heath, 1970
- Meyer, Kain and Wohl, The Urban Transportation Problem, Cambridge, Harvard University Press, 1965.
- Rothenberg, Jerome and Matthew Edel, Readings in Urban Economics, Mac Millan, New York, 1972.
- Rothenberg, Jerome, The Economics of Congestion an Integrated View, AER, May, 1970.
- Ridker, Ronald G., Strategies for Measuring the Cost of Air Pollution, in the Economics of Air Pollution, W.W. Norton & Co., New York, 1966.
- Kneese, Allen V., Water Pollution Economics and Research Needs, Washington, Resources for the Future, 1967.
- Gibbons, J. W. and A. Procter, Economic Cost of Traffic Congestion, Highway Research Board Bulletin 86, 1954.
- Vickrey, William S., Pricing in Urban and Suburban Transportation, AER, May, 1963.
- Maurice, Netter, Critique de la Tarification au Cout Marginal Social in La Vie Urbaine, Number 2, Paris, 1970.
- Alain, Juillet, Sur la Rente Fonciere Urbaine in la Vie Urbaine, Number 2, Paris, 1970.
- Mills, Edwin S., Economics Incentive in Air Pollution in Harold Wolozin (editor) The Economics of Air Pollution, W. W. Norton & Co., New York, 1966.
- Perloff, Harvey S. and Lowdon Wingo (ed.), Issues in Urban Economics, Resources for the Future, The John Hopkins Press, 1968.

### III. Mathematical Tools and Models

- Freeman, Harold, Introduction to Statistical Inference, Addison Wesley, Reading, Massachusetts, 1964.
- Drake, Alvin W., Fundamentals of Applied Probability Theory, McGraw Hill, New York, 1967.
- Larson, Richard C., Urban Police Patrol Analysis, MIT Press, Cambridge, Massachusetts, 1972.
- Morse, Philip M., Queues Inventories and Maintenance, ORSA, Wiley, New York, 1958.
- Chandrasekar, S., Stochastics Problems in Physics and Astronomy in Selected Papers on Noise and Stochastic Processes, Nelson Wax (Ed.) Dover Publications, New York, 1954.
- Bourbaki, Annales de Bourbaki, Ecole Normale Superieure, Paris, 1947.
- Choquet, Cours d'Analyse de l'Ecole Polytechnique, Paris, 1966.
- Mandell, Belmore and John C. Malone, Pathology of the Travelling Salesman, Operations Research, Number 19.
- Raiffa, Howard, Applied Statistical Decision Theory, Boston Division of Research Graduate School of Business Administration, Harvard University, 1961.
- Frank H. and I. Frisch, Communication, Transmission and Transportation Networks, Addison Wesley, Reading, Massachusetts, 1971.
- Little, J. D. C., K. G. Murty, D. W. S. Weeley, An Algorithm for the Travelling Salesman, Operations Research, Number 11, Baltimore.
- Jillion, Beardwood, Halton and Hammersley, The Shortest Path Through Many Points, Proceedings of the Cambridge Philosophical Society, Cambridge, England, 1958.
- Varaya, P. P., Notes on Optimization, Van Nostrand Reinhold, New York, 1972.

#### IV. Traffic Control and Transportation

- Wohl, Martin, Simulation in Traffic Engineering, MIT, Cambridge, Massachusetts, Joint Highway Research Project 1960.
- Use of Poisson Distribution in Highway Traffic, Eno Foundation for Traffic Control, Saugatuck, Connecticut, 1955.
- Schuh, A., Probability Theory Applied to the Distribution of Traffic on Two Lanes Highway, Eno Foundation for Highway Traffic Control, Saugatuck, Connecticut, 1955.
- Edie, L. D. and R. S. Foote, Traffic Flow in Tunnels Highway, Highway Research Board, Vol. 37, 1958.
- Edie, L. D. and R. S. Foote, Effect of Shock Waves on Tunnel Traffic Flow, Highway Research Board, Vol. 39, 1960.
- Bender, J. G. and R. E. Fenton, On the Flow Capacity of Automated Highways, Transportation Science, ORSA, Vol. 4, Number 1, Baltimore, 1970.
- Yaffe, H. J., S. B. Thesis, MIT, 1970 (A Queuing Model for Car Passing).
- Munjal, P. K. and L. A. Pipes, Propagation of on Ramp Density Waves on Uniform Unidirectional Multilane Freeways, Transportation Science, ORSA, Vol. 5, Number 4, Baltimore, 1971.
- Bourne, John, Towards the Quantification of Transit Scheduling Procedures, Working Paper, MIT.
- Pour les Autobus dans Paris, Paris Project Number 8, APUR, Paris, 1972.
- Choix entre Transports Publics et Transports Individuels en Region Parisienne, Cahiers de l'IAURP, Number 26, Paris, 1971.

V. Medical Care and Public Health

- Report of the National Conference on Medical Costs, Washington, D.C., June, 1967, U. S. Department of Health Education and Welfare
- Hospitals, A County and Metropolitan Area Data Book Public Health Service Publication 1971, U. S. Department of Health Education and Welfare.
- Feldstein, Martin, Economic Analysis for Health Service Efficiency, Markhan Publishing Co., Chicago, 1968.
- Health Resources Statistics 1969, Public Health Service Publication, U. S. Department of Health Education and Welfare.
- American Hospital Association, Annual Report 1972, Public Health Service Publication, U. S. Department of Health Education and Welfare.
- Feldstein, Martin, The Rising Cost of Hospital Care, Information Resources Press, Washington, D.C., 1971.
- Fuchs, Victor S., Essay on the Economics of Health and Medical Care, Columbia University Press, New York, 1967.
- Statistics of the Department of Public Health of the Commonwealth of Massachusetts, Boston, 1972.
- Bithell, J. F., A Class of Discrete Time Models for the Study of Hospital Admission Systems, ORSA, Vol. 17, 1969.
- Design and Construction of Storm Sewers, 1969, American Society of Civil Engineers.