OPTIMUM UTILIZATION OF THE TRANSPORTATION SYSTEM*

by Tjalling C. Koopmans

Professor of Economics, The University of Chicago, and Research Associate, Cowles Commission for Research in Economics

The purpose of this paper is to give an application of the theory of optimum allocation of resources to one particular industry. I shall, therefore, not speak on that theory in general. I shall use one of its basic propositions, which was very admirably put forth in the paper presented by M. Allais. This proposition says that a system of prices corresponding to marginal costs is necessary to guide the optimum allocation of resources in a productive system. If cost is minimized in each branch of production on the basis of such a system of prices, each unit of any (divisible) factor of production will be used in such a manner that its contribution to the satisfaction of ultimate consumers is highest.

For this proposition to be valid, it is not necessary that such prices are established in a market where exchange of goods takes place; they may also be accounting prices determined only for the purpose of guiding allocative decisions. I shall give examples of both kinds of prices.

It may be useful, indeed, to consider applications of this proposition to particular industries. The meaning of the marginal cost concept is not always obvious to the engineer, manager, or business economist. It is true that, where perfect competition exists, the mechanism of the market will bring about prices reflecting marginal cost. In a sphere like transportation, however, where perfect competition does not prevail throughout the industry, specific analysis is needed to bring out in quantitative terms what the marginal cost is in any particular case, and how it can be determined.

In order to simplify our problem, I shall consider a homogeneous transportation system, that is, a system in which there is only one type of moveable equipment. For instance, there is only one type of ships all of the same carrying capacity, speed, and other characteristics. Or, there is only one type of railroad cars, or highway trucks.

Let us first consider the case of a railroad connecting only two terminals, A and B, a case which has also been discussed by Pigou in his

* The text of this paper follows closely the stenographic transcript of the original verbal presentation. It will be reprinted in Cowles Commission Papers, New Series, No. 34. A monograph giving a more systematic exposition of the subject is in an early stage of preparation.

book, *The Economics of Welfare.*[1] Let us assume that there is a given demand for five trains each day to go loaded with goods from A to B; that there is a demand for only three trainloads daily to go from B to A. Let us express cost simply in terms of equipment tied up, *i.e.*, in train-days incurred daily. Then, if we wish to transport an additional trainload from A to B, that increase in demand will require an additional train to be run daily from A to B loaded with goods. The cost incurred directly by that movement is the sum of the times spent loading in A, moving to B, and discharging in B, by one train. But it will also be necessary to move the train back empty from B to A, because we assume no change in the requirement of three loaded trains daily in that direction. The marginal cost in this case, expressed in equipment time committed each day, corresponds therefore to the whole turn-around time of one train, loading, moving, discharging, moving back. On the other hand, the marginal cost of adding one trainload daily from B to A is given only by the time spent loading in B and discharging in A, because the time spent moving would have to be spent in any case, to approximately the same amount, as a result of the fact that otherwise that train would have to be moved empty. We thus find a sizable difference in marginal cost according to the direction of transportation. This was clearly recognized by Pigou. It is difficult to understand why he regarded this difference as of comparatively small importance.

Since most transportation systems connect many terminals, we shall now consider how the determination of marginal cost works out in a general network of routes. Let us assume, however, that the *program* of transportation is constant in time. Constant daily or monthly requirements for transportation from each terminal in the network to each other terminal are assumed to be given. Let us assume further that the *performance times* involved in the various tasks of loading, moving, discharging, are constants in time and in the sense that on each route they are independent of the number of trains or ships that carry out these tasks. This implies an assumption of absence of congestion.

We shall again assume that the cost of a program can be expressed in amount of equipment required, or, synonymously, in equipment time committed in each unit of time. This is not as unrealistic as it may seem. There have been situations where equipment time was the decisive element of cost. For instance, in the shipping problems of the two World Wars, the controlling bottleneck was the number of ships available. All other costs, like wages and fuel, even though important by themselves, were negligible compared with the opportunity cost of

[1] 1920. Ch. XV, 85, p. 266.

using a ship for one highly urgent purpose rather than for some other highly urgent purpose.

I shall distinguish, for any program, the *direct cost* and the *indirect cost*. The direct cost of the program (or of any increment thereto) is the equipment tied up at any time in loading, loaded movements and discharging (or its increment). The indirect cost arises whenever there is a departure from perfect balance in the program. In general, certain terminals will receive more goods than they dispatch, and other terminals will be in the reverse situation. Generally, a continual movement of empty equipment is required from points of equipment surplus to points where there is a deficit. The amount of equipment inevitably tied up in empty movements is called the indirect cost of the program.

In a transportation system that is not too unbalanced, the direct cost is by far the more important element in total cost. But in the marginal cost of given increments to the program, the indirect cost is always important and deserves a good deal of study. It has a more complicated structure than the direct cost, and it enters into marginal cost in a more subtle way.

As an example for the discussion of this problem I have chosen the flows of dry cargo on the ocean shipping routes of the world in the year 1925. For the study of indirect cost, we need only consider the net shipping surplus of each port or area of limited size. We can roughly assume that the net dry-cargo shipping surplus of an area is proportional to the net excess of the weight of all goods (other than mineral oils) arriving in sea-borne trade over the weight of all such goods departing. In Table 1, such net receipts figures are computed for areas designated by "representative ports" and indicated by dotted lines on Figure 1. Let us simplify our problem by assuming that the figures of 1925 are constant flows applying through time for an indefinite period, without seasonal movement or other fluctuation or trend. Furthermore, let us calculate as if all traffic going to or from a particular area were going to or from its representative port. The representative ports which, by our assumption, have a net surplus of shipping are Lisbon, Athens, Yokohama. All other representative ports are shipping deficit ports.

Let us now for the purpose of argument (since no figures on war experience are available) assume that one particular organization is charged with carrying out a world dry-cargo transportation program corresponding to the actual cargo flows of 1925. How would that organization solve the problem of moving the empty[2] ships most economically

[2] Technically, ships here referred to as empty take in a certain amount of ballast for stability reasons. "Empty movements" are accordingly described in shipping parlance as "ballast traffic."

from where they become available to where they are needed? It seems appropriate to apply a procedure of trial and error whereby one draws

TABLE 1

Net receipts of dry cargo in overseas trade, 1925

Unit: Millions of metric tons per annum

(1)	(2)	(3)	(4)
Area represented by[1]	All cargoes other than mineral oils		
	Received	Dispatched	Net receipts
New York	23 ·5	32 ·7	–9 ·2
San Francisco	7 ·2	9 ·7	–2 ·5
St. Thomas	10 ·3	11 ·5	–1 ·2
Buenos Aires	7 ·0	9 ·6	–2 ·6
Antofagasta	1 ·4	4 ·6	–3 ·2
Rotterdam*	126 ·4	130 ·5	–4 ·1
Lisbon*	37 ·5	17 ·0	20 ·5
Athens*	28 ·3	14 ·4	13 ·9
Odessa	0 ·5	4 ·7	–4 ·2
Lagos	2 ·0	2 ·4	–0 ·4
Durban*	2 ·1	4 ·3	–2 ·2
Bombay	5 ·0	8 ·9	–3 ·9
Singapore	3 ·6	6 ·8	–3 ·2
Yokohama	9 ·2	3 ·0	6 ·2
Sydney	2 ·8	6 ·7	–3 ·9
Total	266 ·8	266 ·8	0 ·0

Source: *Der Guterverkehr der Weltschiffahrt*, Statistisches Reichsamt, Berlin, 1928.
[1] See Figure 1.
*The figures in columns (2) and (3) for this area contain an equal amount of traffic within the area, between smaller areas from which this area was composed.

tentative lines on the map that link up the surplus areas with the deficit areas, trying to lay out flows of empty ships along these lines in such a way that a minimum of shipping is at any time tied up in empty movements.

The lines on Figure 1 correspond to an optimal solution of that kind, if we can assume that time spent is proportional to navigational distance. The procedure of trial and error can be illustrated as follows: Each surplus area serves a number of deficit areas, and the type of experimental variation that one would explore is to shift a certain "marginal" deficit area from one surplus area to another, with compensation elsewhere. For instance, one might think of cutting the link from Lisbon to West Africa, substituting a compensating link from Lisbon to San Francisco; one might explore several other limited adjustments of that kind, calculating in each instance the (positive or negative) net saving of shipping so achieved, on the basis of the performance times involved in the alternative movements. In that way one would arrive at what may be called a "local" optimum, that is, a routing plan of empty ships that cannot be improved upon by adjustments of the type described. The question arises whether one cannot mislead oneself in that way. Is it

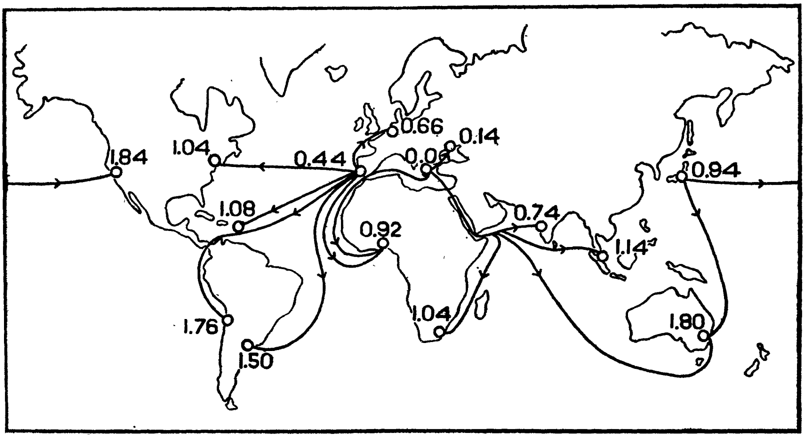

Figure 1. Optimal Routes of Empty Shipping Corresponding to World Dry-Cargo Flows in 1925.

The Figure shown with the representative port of each area represents the net shipping surplus of that area in millions of metric tons of dry-cargo capacity.

not possible that, by a very drastic rearrangement in the linking of surplus and deficit ports, another perhaps better optimum could be found which cannot be detected by any "small" rearrangement?

The question is answered by the *first theorem*: If, under the assumptions that have been stated, no improvement in the use of shipping is possible by small variations such as have been illustrated, then there is no—however thoroughgoing—rearrangement in the routing of empty ships that can achieve a greater economy of tonnage.

The reason for this statement is a mathematical one which can be only briefly suggested: The function we are minimizing, the total amount of shipping tied up in the various flows of empty shipping, is the sum of the monthly flows on all routes, each multiplied by the constant performance time involved in that movement. We are thus minimizing a linear function of the flows of empty ships under two types of restrictions. In a continuing program, the number of ships going into any area per unit of time, with or without cargo, must equal the number of ships going out. Therefore, there is a first set of restrictions in the form of linear equalities saying that the sum of all flows of empty ships out of any area less the sum of all such flows into that area is equal to the shipping surplus of that area, as prescribed by the program. This surplus may of course be negative. There is a second set of restrictions which says that a flow of empty ships cannot be negative. This is a linear inequality. We are thus minimizing a linear function subject to linear equalities and linear inequalities in the variables involved.

If we take the flows of empty shipping on all possible routes as the Cartesian coordinates of a point in an n-dimensional space, then the set of all points satisfying these two types of restrictions has the following property: If we select arbitrarily two points of this set, then all points located between those two points on the straight line connecting them will also belong to the set, *i.e.*, satisfy the restrictions stated. A point set with this property is called a convex set, and further analysis shows that the minimum value of a linear function on a convex point set is unique: Any local minimum is the absolute minimum.[3]

We now come to the second problem to be discussed: how to find estimates of marginal cost. The constant program for which an optimal routing plan of empty ships has been found is now subjected to variation, not in time, but as a matter of comparative statics. Besides the

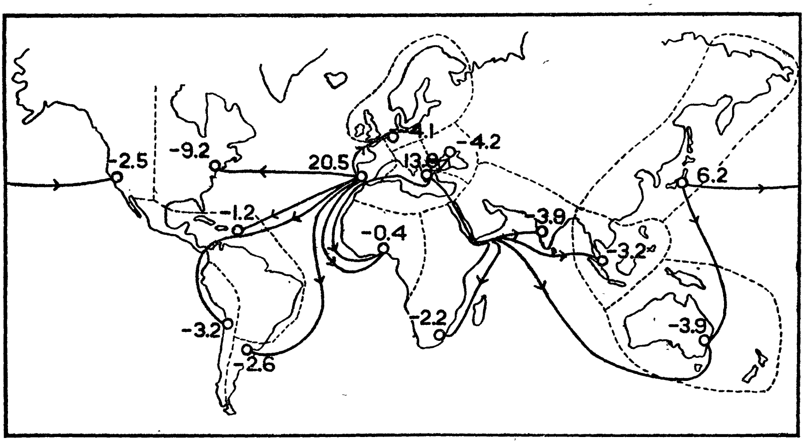

Figure 2. The Potential Function of Dry-Cargo Ships Corresponding to the Composition of Demand in 1925. Unit: Ship-Months.

constant program already considered, we consider another constant program which differs from the previous one only with respect to the amount of cargo to be shipped on just one route. This amount is

[3] It is possible that the minimum value is reached at different points simultaneously; instead of the one lowest point in a valley there is then a horizontal line constituting the lowest part of a valley, or even a low plain at the bottom of a valley, or its analogue extended into more dimensions.

increased by, say, one shipload a month. The calculation of marginal cost with respect to that change in the program can be performed with the help of a certain calculus illustrated by Figure 2.

For each port in which empty movements originate or terminate (or both) we define the value of a *potential function*, which is a valuation placed on the location of a ship in that port. This definition proceeds as follows: We assign an arbitrary value to the potential function in one arbitrary port, in our example the value zero in the port of Athens. From there we follow routes travelled by empty ships according to an optimal routing plan for the original (unchanged) program. In order to derive the potential in Bombay from the potential in Athens, we add the time involved in an empty movement from Athens to Bombay. We *add* because the movement from Athens to Bombay is in the direction of empty traffic. In the same way, this procedure defines the potentials in Odessa, Singapore, Sydney, Durban, and Lagos as certain positive figures. From any of these ports, we cannot go on along routes of empty shipping except by moving counter to the flow of such ships, as for instance along the route from Sydney to Yokohama. Therefore, in that case, we *subtract* the amount of time spent in the empty movement Yokohama-Sydney from the potential in Sydney in order to obtain the potential in Yokohama. In this way the potential is defined in any port, linked with Athens by the *graph* of optimal routes of empty shipping for the original program.[4]

I shall now formulate a rule for determining the marginal cost of a given change in the program. Let us take as an example the addition of one ship to the monthly loaded movement from San Francisco to Antofagasta. The marginal direct cost is simple — it is given by the time involved in loading, moving, and discharging, on that route. The marginal indirect cost, according to the *second theorem*, is equal to the loss in potential sustained by a ship while going from the port of departure to the port of destination. In our example, that loss is positive, because the potential at destination (1.76) is lower than at the port of departure (1.84). Therefore, the marginal indirect cost involved in this particular change in the program is $1.84 - 1.76 = 0.08$ ship-months, incurred monthly, or 0.08 of the continuous active availability of one ship.

[4] It can be shown that a closed circuit can be contained in the graph of optimal routes only if the performance times involved are such that the definition of the potential applied around the circuit does not lead to a contradiction. It is, however, possible for the optimal graph of empty traffic to break up into disconnected parts. In such special cases, differences in potential between ports on the same connected part are defined, but differences in potential between ports that are not connected by the graph are not defined.

Why is this theorem valid? It can be briefly indicated. If such an addition to the program is made, the net monthly surplus of ships in San Francisco is reduced by one; likewise, the net surplus in Antofagasta is increased by one. The flow of empty ships from Yokohama across to San Francisco can therefore be reduced by one ship a month. But that upsets the balance in Yokohama, and it will be necessary to move one additional ship monthly from Yokohama to Sydney, and so on. This dispenses with the necessity of sending one ship monthly from Athens to Sydney, and so on. The sequence of adjustments is closed when it is found ultimately that the monthly arrivals in Antofagasta of empty ships from Lisbon are reduced by one. Now, the algebraic sum of the time-expenditures and the time-savings involved in such a sequence of adjustments is precisely equal to the difference in potential between the end (Antofagasta) and the beginning (San Francisco) of a chain of routes of empty shipping, determined by application of the definition of potential along the chain.

In a war economy in which shipping is the essential bottleneck, the usefulness of marginal cost estimates as described is obvious. Such estimates are needed to guide decisions of programming authorities, for instance, in balancing competing claims for shipping services, or in determining the best source of a raw material on shipping grounds. It may be added without proof that the estimates described are applicable to finite (as distinct from infinitesimal) changes in the program, which are not so large as to require a change in the optimal routes of empty traffic.

What relevance does the foregoing analysis have to peacetime transportation problems where there is a market instead of an allocating authority, and where equipment time is not the only relevant measure of cost? I believe that the main part of marginal cost will still be arrived at along the lines described. In the first place, the equipment time committed by a change in demand is again to be accounted for, in the present case on the basis of the market valuation of equipment time (the opportunity cost of the use of equipment). In the shipping market, this valuation is expressed by the time-charter rate of a ship; in rail transportation no market quotation is available, but proper accounting procedures will reveal the net rental value to a railroad of the use of a car or train. In addition, the cost of fuel consumed and of labor to go with the equipment will also be roughly proportional to the time spent moving. Hence the same analysis is still largely valid for a considerable part, I would say the main part, of marginal cost.

How has the shipping market done its job without resorting to anything like the analysis described? To answer this question, we can make use of a theorem which M. Allais has already pronounced:

A perfectly competitive market automatically brings about pricing according to marginal cost. Therefore, to the extent that the tramp shipping market has been competitive—and that is to a very large extent through a long period in its history—the individual comparisons of alternative voyages made by many shipowners acting independently have broadly given effect to the process of minimizing the amount of shipping involved in empty movements; or rather of maximizing the amount of transportation that is performed by a given amount of shipping, which is an equivalent formulation. The totality of these individual decisions has furthermore produced a set of interconnected freight rates on various routes, reflecting marginal cost.

There is a definite need for an explicit analysis of marginal cost in rail transportation, where there is nothing like a competitive comparison of alternative courses of action by individual train owners. In the United States, movements of trains are laid out and rates are set by a number of railroad managements acting under the supervision of a regulatory agency of the government. As a result, I would surmise, the railroad rates have no connection whatever with marginal costs. The cases are rare in which rates in different directions are different, and I do not know of cases where a railroad's rate system has been made dependent on the composition of traffic. We must realize the social cost involved in this disregard of marginal principles—cost in terms of the decrease of social benefit that we derive from our transportation system. If rates do not reflect marginal cost, they provide no inducement or guidance toward private or public decisions regarding industrial location that will improve the balance in the use of the transportation system. For instance, in the United States, processing industries are more concentrated in the Northeast quarter of its area. Therefore, there is a net flow of raw materials from South and West to East, which is a larger movement in terms of weight or bulk than the reverse net movement of manufactured goods from the Northeast to the South and West. We are, of course, all made to pay for the extensive movement of empty cars thus necessitated, but we are not made to pay in such a way as to set up an incentive to change the situation. A system of railroad rates corresponding to marginal costs would quote higher rates per carload of goods carried toward the Northeast, where the predominant movement goes, than it would quote for the reverse direction. Such rates would contain just the optimal inducement to move processing activities away from the Northeast.

I must make one other qualification here. For a rate system according to marginal cost as regards different routes to be beneficial, it would likewise have to be in accordance with marginal cost as between different commodities. The present rate system also does not satisfy this criterion.

Commodities for the transportation of which the demand is inelastic are charged higher. It is uncertain whether the introduction of directional rates of the type that I have discussed, without at the same time abandoning discrimination between commodities, would lead to a better allocation of resources than the present rate system. It would certainly not lead to the optimum allocation.

It is, of course, well known that a system of pricing at marginal cost will imply operation at a deficit whenever and wherever the density of traffic is distinctly less than the capacity of the road. Other provocative features of marginal cost pricing are rates depending on the composition of demand by routes, possibly seasonal rates, possibly also contracts based on future rates, announced by the management of the railroad system and at any time subject to revision for contracts still to be concluded. It will be necessary to strike a balance between the cost to enterprise of uncertainty regarding future rate levels, the cost to railroads of announcing and applying changes in the rate structure, and the desirability of closely reflecting in rates the ever present fluctuations in the composition of demand. Further development of the foregoing analysis in a dynamic direction as well as factual study of fluctuations in demand are required before an approximately optimal railroad rate system can be formulated.

In conclusion, I wish to emphasize that a theory of optimal transportation rates, of which the present analysis is a small beginning, would provide an indispensable groundwork for any theory of the optimum geographical distribution of industry.

Résumé

Dans cette communication les principes de "l'économie du bien-être" sont appliqués à tout système de transport où les marchandises sont transportées à l'aide de matériel mobile (par vaisseau, wagon, camion, avion, etc.)

Considérons par exemple l'allocation d'une masse de transports maritimes, soit par une grande entreprise, soit par une autorité telle qu'il en a existé pendant les deux guerres mondiales. Dans un cas statique simplifié le programme consiste en une matrice A dans laquelle l'élément a_{ij} indique le nombre constant de vaisseaux uniformes requis chaque mois pour chargement au port "i" à destination du port "j". Le

coût total, c.à.d., le nombre de vaisseaux actifs qui est nécessaire pour l'exécution du programme se subdivise en *coût direct*—le nombre moyen en train d'être chargés, de naviguer avec cargaison, et d'être déchargés—et *coût indirect*—le nombre naviguant à vide vers un port de chargement. Le coût direct est une fonction linéaire des éléments de la matrice du programme, et des durées (supposées constantes) des mouvements ou opérations.

Le premier problème est la réduction au minimum du coût indirect, par un arrangement adéquat de trafic à vide. Une méthode tentative de solution est justifiée par le THEOREME 1. *Un arrangement de trafic à vide qui ne permet pas d'économies par un changement menu quelconque dans l'assignation de routes, ne permet pas d'économies par un changement intégral quelconque des routes assignées.*

Le coût marginal de chaque constituant a_{ij} du programme est la dérivée du coût total par rapport à a_{ij}. Cette dérivée est établie au moyen du: THEOREME 2. *Il est possible de définer un potentiel, c.à.d., une fonction p_i dans chaque port "i" touché par le trafic a vide, telle que le coût indirect marginal du constituant a_{ij} soit égal à la perte de potentiel $p_i - p_j$ subie par un vaisseu en effectuant un voyage pour constituant du programme. Le potentiel s'accroit le long des routes de trafic a vide, dans le sens de ce trafic, d'une quantité égale à la durée du voyage.*

Dans une économie de guerre, les autorites chargés du programme peuvent se servir du calcul du coût marginal décrit ci-dessus, p.e., pour déterminer la source la plus appropriée, du point de vue du transport, d'une matière première quelconque. Dans une économie qui s'assigne l'utilisation optimale des ressources, des taux de fret égaux aux coûts marginaux doivent guider les décisions de transport et de location industrielle. Le marché du *tramp shipping*, où la concurrence a été presque parfaite pendant une longue période, a connu un système de taux de fret approximativement égal à tout instant aux coûts marginaux anticipés. La manque de concurrence des chemins de fer a facilité le développement de systèmes de taux dégagées du coût marginal, et aussi entraîne des pratiques non-économiques dans le transport et la localisation industrielle.

Mr. Koopmans' paper was discussed by Messrs. Maurice Fréchet, M. Allais, and the speaker.