METHODS OF ALLOCATION

Ross Parish

Monash University

METHODS OF ALLOCATION

In this article I describe, classify, and - to a limited extent - analyse and evaluate in terms of economic criteria some methods of allocation that are commonly used in the distribution of goods by firms and bureaucracies. The coverage is not exhaustive. I distinguish between non-competitive allocation and large-numbers competitive allocation, and within the latter category, between price and non-price allocation. All competitive methods of allocation involve surplus seeking by potential recipients, and the non-price methods - which in my terminology comprise allocation by ordeal, and allocation by product degradation - give rise to surplus dissipation, i.e. to the phenomenon that is the concern of the burgeoning rent-seeking literature.

Not all of the classificatory distinctions are clear-cut: some categories shade into one another. It should also be borne in mind that in reality many, if not most, goods are allocated by a mixture of methods. Thus, with marketed goods and services, waiting lists and queuing often play a minor allocative role.

Allocation and Competition

Methods of allocation may be classified according to the extent to which the allocation is imposed from "above" (whether by law, authority, or custom), or determined by competition among potential consumers. Imposed allocations are apt to be very inefficient in the Pareto sense, and may also be costly to administer. Competition has the advantage of allowing consumers to select themselves according to the intensity of their preferences, but it can take forms that involve substantial real costs, and the resulting allocation may not be Pareto-efficient.

Competition may occur at different stages of the allocative process. First, it may be engendered by the process whereby the allocative rules are established or modified. Lobbying of legislators by groups favoring or opposing rent control is a case in point; court proceedings seeking to change the method of selecting among applicants for civil service jobs is another. Second, the allocative criteria may be such that the good goes to the "winners" of a competitive process; for example, to those willing to pay at least the market-clearing price, or to those receiving the highest scores in a competitive examination. Third, the process need not end with the primary allocation: incentives may exist to reallocate some of the good, and this normally takes place by some form of market competition.

Thus, under wartime rationing and price control, black markets developed wherein goods changed hands at prices above the legal maxima.

The classification of methods of allocation, set out below, is based on whether, and in what form, competition of the second type occurs, ie on the extent to which the primary allocation is imposed or competitively determined. Preallocation and reallocation competition are ignored (except in so far as they are "built-in" to the primary allocation method) not because they are unimportant, but because they are so prevalent as not to provide sharp classificatory distinctions. Also, the fact that they might be especially induced by some primary allocation methods, provides an additional reason for concentrating on the primary allocation. Thus, it seems reasonable to suppose that competitive methods of allocation may induce less rule-influencing activity than non-competitive methods, since competition within the rules provides an alternative means of obtaining the good for those who value it highly. Reallocation has the potential to occur whenever the primary allocation procedure fails to exhaust the possibilities for mutually-advantageous exchange of the good. Whether and to what extent it does occur depends upon the magnitude of the transaction costs involved, and these are affected by whether or not the transactions are legal, the nature of the good, and the outcome of the primary allocation. For example, if the good is dispersed geographically as a result of the primary allocation, the costs associated with reallocation might be higher than if the first allocation had not taken place.

Allocative Criteria

The allocation of goods normally involves sellers/allocators transferring units of the good to buyers/recipients who satisfy certain eligibility criteria. Whether, to what extent, and in what form potential consumers compete for the good depends on the nature of these criteria. If the recipients

are distinguished on the basis of innate characteristics, competition among them for possession of the eligibility requirement is not possible (except by way of fraud): potential recipients accept passively the allocation imposed by the rationing authority. Allocations based on innate characteristics of the recipients will be called ostensibly non-competitive--the qualifier "ostensibly" serving to remind us of the possible existence of competitive activities designed either to change the allocative rules or to reallocate the good.

If individuals can meet the eligibility criterion by purposeful behavior the basis exists for competition among potential recipients. Such criteria I will call behavioral, and methods of allocation based on them, competitive. Thus under price allocation, the goods go to those most able and willing to pay for them, while with queue rationing, the recipients are those most willing to devote time to standing in line.

In both price and queue rationing, the behavior whereby individuals qualify themselves as recipients is directed solely toward this specific end: the price one pays, or the time one spends in line, serves no other purpose. But frequently goods are allocated according to behavioral criteria where the behavior concerned is normally intended to serve some purpose quite distinct from meeting the eligibility requirement. Social services are often allocated on the basis of criteria of this type, such as income status, marital status, family responsibilities, etc. Such criteria can be regarded as innate in the immediate short run, but as more or less behavioral in the long run, since, after the method of allocation is announced, individuals have incentives to modify their behavior so as to meet the eligibility requirement. Thus, with mean-tested pensions, people have an incentive to keep their retirement incomes below the eligibility ceiling, either by saving less or by converting their wealth into forms that do not produce

income as defined by the pension administrators. I will call methods of allocation based on this type of criterion intermediate, since their outcome is intermediate between the ostensibly non-competitive and the competitive methods. The extent to which the initial, imposed, allocation is modified by competitive behavior depends on the magnitude of the incentives for, and costs of, complying with the criterion, for those who ordinarily would not comply.

I shall now consider in more detail, and with examples, each of these categories of allocation method, and distinguish some sub-categories within them.

1. Ostensibly Non-Competitive Methods of Allocation

In defining allocations based on innate characteristics an non-competitive, the word "innate" is not to be taken too literally. Essentially what is meant is a characteristic that cannot be altered or acquired in the relevant time period. It thus includes characteristics acquired in the past (provided they were not acquired in anticipation of their becoming criteria for the particular allocation under discussion) such as educational qualification, prison records, etc., as well as such inherent traits as age, sex, race, etc. All of these characteristics are used as criteria in the allocation of jobs by employers. Age and sex are relevant criteria for the receipt of certain social services, and race is an important allocative criterion in some societies.

Heredity determines the succession of the crown in peaceful monarchies, of titles in aristocracies, and of land ownership under primogeniture. It is also an important criterion for determining citizenship, and eligibility for membership of certain clubs and societies (eg. Daughters of the American Revolution). Admission to some schools and colleges is easier for the

offspring of alumni. This is true not only of private institutions: in at least two Australian cities preferential admission to some state high schools is given to the sons and daughters of "old boys" and "old girls" of the school. (The hereditary principle came to be adopted in the following way. The schools concerned were originally academically-selective, admitting the better students from a wide geographical area. The changing tides of educational ideology deemed this arrangement to be too "elitist", and the schools were turned into non-selective schools serving their immediate neighborhoods. However, as a result of agitation from outraged alumni, the policy was modified as indicated above).

Allocations determined by some random process, such as the drawing of lots, or on the basis of "equal shares for all," both belong to the non-competitive category. They can be thought of as rewarding the "innate" characteristics, "luck" and "existence," respectively. The drawing of lots is not uncommonly used to determine eligibility for some good when the number of qualified applicants exceeds the number of items available. For example, in Australia in the post-war period soldier-settlement farms were allotted to qualified veterans by ballot. Undoubtedly the most significant recent use of random allocation was for the requisitioning of services rather than the rationing of a good: I refer to the "birthday ballot" method of determining the draft eligibility of young men. Rationing on an equal shares basis may in certain circumstances be widely perceived as being equitable--basic foods in wartime being a case in point. Nor is it uncommon as a rough but quick outcome of negotiation or adjudication of competing claims.[3]

With non-competitive allocation, the problem of equilibrating demand and supply is replaced by that of devising and adjusting the allocation rules so as to reconcile consumers' aggregate entitlements to the good with the

available supply. This task is easier, the more elastic the supply, and the more elastic the agency's budget: if both are perfectly elastic, all the adjustment can be made on the supply side. Otherwise, some adjustment to the allocation rules is required: the eligibility requirements tightened or relaxed, and/or the individual ration reduced or increased. Since neither planning nor subsequent adjustments are ever perfect, imposed allocation schemes seldom stick to their announced criteria, nor rely entirely on their changing criteria to allocate the good: supplementary methods of rationing, such as queuing and waiting, also come into play.

The administrative and enforcement costs of such schemes are often substantial, on account of the activities just mentioned and of the need to check on the eligibility of the recipients.

Since imposed allocations do not give consumers the opportunity to express the intensity of their preference for the good by self-selection, they are bound to be inefficient in the sense of leaving many opportunities for mutually-beneficial exchange unexploited. Considering first goods that each individual consumes only in unit quantity (eg. particular books, particular medical precedures) but which are valued differently by different individuals, the efficiency of the allocation depends on the association between individuals' eligibility and their willingness to pay for the good. If those meeting the eligibility criterion were each willing to pay a greater amount for the good than the greatest amount that any ineligible person was willing to pay, the allocation would be perfectly efficient. If the eligibility criterion was distributed essentially at random with respect to willingness to pay, those excluded from consumption would, on average, value the good as highly as those to whom it was allocated, and there would be a considerable loss of potential economic surplus.

For goods consumed in different quantity by different consumers, imposed allocation is inefficient not only in separating consumers from non-consumers, but also with respect to the quantities allotted to different consumers. Typically each consumer is allotted a standard ration, or one of a limited variety of rations, with the "more eligible" receiving larger rations than the "less eligible," Again there is a loss of surplus arising from the failure to equate different consumers' marginal evaluations of the good.

2. Competitive Methods of Allocation

With competitive methods of allocation, the recipients select themselves, by their willingness to make some sacrifice in exchange for the valued good. Hence, except for transactions based on mistaken perceptions, it can be assumed that each recipient values the good at least as highly as that which he sacrifices in order to obtain it.

The most common, and best understood, competitive method is, of course, allocation by price.

Of competitive, non-price, methods of allocation, one group includes queue rationing, rationing by waiting,[1] and some others to be described shortly. In these the good is allocated according to individuals' willingness to undertake time-consuming, inconvenient and/or disagreeable activities in association with the transaction. It therefore seems appropriate to refer to them as allocation by ordeal.

Queuing and waiting are often the more spectacular manifestations of a deterioration in the quality of service offered by sellers of a good when, for some reason, its price is held below the market-clearing level. Other aspects of service degradation include shorter business hours, lack of advertising and display, and similar phenomena. But just as service degradation

is a natural response of sellers to price control, so is a reduction in the quality of the product itself. Hence I distinguish allocation by product degradation as a third sub-category. It includes not only planned product degradation designed to evade price control--not a particularly significant phenomenon these days--but also the spontaneous degradation of the quality of service provided by common-access facilities subject to congestion or similar externalitites.

Price Allocation

For purposes of comparison with other methods to be described below, certain aspects of price allocation deserve brief mention:

1. The good is exchanged for money.

2. There is a tendency, the strength of which depends on various circumstances, for a single price (aside from transport and other service cost differentials) to prevail at any point of time in connected markets.

3. There is also a tendency for the price (or price structure) to adjust to changes in supply and demand so as to clear the market.

4. These tendencies are the outcome of the self-interested activities of buyers, sellers, and middle-men.

5. Various institutions, conventions, and specialized occupations have evolved which facilitiate the process of price "discovery" and adjustment. These include centralization of transactions (eg. commodity-exchanges, shopping centers), disclosure of transaction prices (eg. auctions, trading by "open outcry," publication of market prices), product grading and standardization, brokers, dealers, and specialized traders of all kinds.

6. The stronger are the tendencies 2 and 3, the fewer the opportunitites for mutually-beneficial reallocation of the good.

Allocation by Ordeal

Rationing by ordeal resembles price rationing in that potential recipients have to choose between the good and what they are required to give up in order to obtain it. With queue rationing and rationing by waiting, the magnitude of the ordeal is determined by a process analogous to competitive bidding: a man who joins the queue earlier "bids higher" than one who joins it later, and is more likely to be served before the supply is exhausted. In some other cases, the nature and magnitude of the ordeal is determined by the supplier. An example is provided by a water-supply authority, in times of drought, banning the use of sprinklers on lawns and gardens, but permitting the use of hand-held hoses: to water one's lawn one has to stand, hose-in-hand, for the requisite period of time. Another example is the blocking of vehicular access to campsites in national parks so as to exclude those unwilling to undergo the ordeal of backpacking their equipment to the site. Thus different methods of price discovery, familiar in ordinary markets, are also to be found in "ordeal" markets.

However, rationing by ordeal differs from price rationing in several important respects, the most notable being that whereas the money price is at the same time a cost to the buyer and a benefit to the seller--and hence a transfer from the social point of view--the ordeal is a real cost, which, if (as in the cases mentioned so far) it confers no benefit on the supplier, is a price paid by the "buyer" but received by no one. In such cases rationing by ordeal involves trade-off by the recipient but no transfer to the supplier of what is traded off: it is not exchange. It is therefore customary to speak of the cost of the ordeal as pure waste, but one needs to be careful in interpreting this statement. It is not pure waste in the sense that it serves no purpose; it serves to allocate the good to those most

willing to suffer the ordeal. It is pure waste in the sense that if the same function could be performed by price allocation the costs of the ordeal could thereby be avoided.

Another difference from price rationing is in the strength of the market equilibrating forces. These derive, under price allocation, from the incentives motivating buyers, sellers, and intermediaries. But with rationing by ordeal, if the supplier gains nothing from the ordeal, he has no incentive to seek a higher "price", or, if his "price" is lower than others, to advertise the fact to potential recipients. Similarly there is no incentive for "price" arbitrage by middle-men. Thus the whole burden of market equilibration is thrown upon consumers, who still have an incentive to search for those sources of supply that involve the least painful ordeal. For example, the transport costs which determine the extent to which prices diverge over space would be enormously greater if, as under queue rationing, spatial equilibration was performed only by individual consumers, each subject to a low transaction limit, rather than, as under price rationing, by sellers or dealers transporting the good by the truckload. Ordeal comparisons and hence equilibrating actions by consumers may also be hampered by uncertainty as to the "price" prevailing at any time or place: we have all had the experience of discovering that the shortest queue at banks or supermarkets is not necessarily the fastest. (It must be admitted, however, that the same problem would arise in choosing which among several auction markets to patronize). In certain cases of "disorderly queuing"--overloaded telephone lines being an example--"price" uncertainty might be very high.

A third difference is that, even if the market-equilibrating tendency is strong enough to establish a uniform, market-clearing ordeal, the costs of undergoing the ordeal will vary from person to person. If the ordeal is

standing in line, people will differ in their subjective evaluations or the time given up and the inconvenience incurred. Hence the single ordeal "price" will translate into many different money-equivalent prices, and the resulting allocation will not be Pareto-efficient. The remaining opportunities for mutually-beneficial exchange may be realized by post-ordeal good-money exchanges, or through the contracting out of the ordeal by those for whom it is more costly to those for whom it is less costly. Though these arrangements may differ in form and legal description, they are indistinguishable in their economic effects, which are to make the allocation more Pareto-efficient, and to reduce the aggregate social costs of the ordeal.

A fourth respect in which allocation by ordeal often differs from price allocation is in the options confronting the would-be consumer. The ordeal is ordinarily part of the transaction and its costs a major part of the transaction costs. The ordeal is thus usually a "price" per transaction, so that the unit "price" of the good depends on the size of the transaction. The rational consumer will therefore minimize his ordeal by making each transaction as large as practicable, and, to prevent this sort of price evasion, transaction limits are usually imposed. Queue rationing is almost invariably accompanied by transaction limits of some sort, either artificial or natural (ie, derived from the nature of the good--for example haircuts, games of golf), rigid (e.g. no more than five dollars' worth of gasoline). or somewhat flexible (e.g. no more than one tankful of gasoline). Queue rationing thus confronts the consumer with the same sort of choice he has to make when a discriminating monopolist offers him "x units for y dollars," ie. an all-or-nothing choice in which he compares the cost of the ordeal with his total valuation of the number of units of the good comprising the maximum transaction. Depending on the transaction size and other circumstances (especially the consumers' opportunities for storing the good or

rearranging the timing of consumption), rationing by ordeal therefore may induce consumers to pay a higher price (in terms of the money-equivalent cost of the ordeal) than if the same quantity had been rationed by a single per unit price in the ordinary way, with the consumer being free to buy as little or as much as he choses.

With some types of ordeal, the cost is directly proportional to the quantity of the good received, so that the all-or-nothing choice aspect is not present. Thus with hand-held hoses, for a given hose diameter and water pressure, the time spent holding the hose is directly proportional to the amount of water used.

So far I have considered only ordeals that are purely allocative in function, and confer no benefit on, or serve no interest of, the supplier. In many cases, however, the supplier benefits, materially or otherwise, directly or indirectly, from the recipients' ordeals. An important class of indirectly beneficial ordeal includes those that serve as a screening device, selectively favoring characteristics believed to be complementary with the good or service being rationed, thus helping to ensure that it will be utilized effectively. The supplier may desire this outcome for altruistic or selfish reasons, or both. The most obvious example is the allocation of places in institutions of higher education on the basis of academic attainment as measured by competitive examinations. The better examinees are assumed to be more likely to make better use of the educational services provided by the institution; also, teaching them is likely to be a less onerous and more gratifying experience for members of the institution.

Babies for adoption, and places in some private schools, are often rationed, in part, by long waiting times. In the case of adoptions, the waiting time may serve to test the genuineness of the aspiring parents'

desire for a child--presumably an important precondition of their loving it and caring for it long-term--by giving those for whom the desire proved to be transient an opportunity to drop out. Schools may regard the willingness of parents to put their child's name on a waiting list years in advance as an earnest of their interest in the school, and in their children--both desirable traits from the school's point of view.

In some cases of rationing by ordeal, the benefit received by the supplier from the ordeal is much more direct than in the screening examples discussed above. A good example is the allocation of new licenses to engage in a regulated industry to employees with long and unblemished service in that industry. In Australia this method is commonly used to allocate new taxi license and new commerical fishing vessel licenses. Entry into both these occupations is restricted by regulatory agencies, licensees receive monopoly rents, and licenses consequently become quite valuable. By issuing new licenses the regulatory agencies have the power to bestow a free gift worth many thousands of dollars on fortunate individuals. But the method chosen ensures that recipients work for their "gifts", and, further more, that the product of their work is received by the existing licensees. The prospect of receiving a valuable asset after long service presumably attracts more career-oriented drivers and crewmen into the respective industries--thus reducing labor turnover--and induces them to accept lower wages and worse working conditions than they otherwise would. Recipients of new licenses thus pay for them by "good and faithful service," while existing licensees receive payment in the form of a less costly and less troublesome workforce. (Swan, 1979). The arrangement thus amounts to the selling of licenses for payment in kind. Its purpose is to render as nearly complete as possible the monopoly power of the cartel of existing licencees,

in a politically-palatable manner, the disguised payment in kind being far less likely to provoke adverse comment--indeed, probably being widely regarded as "fair"--than would the direct sale of licenses for money.

In rationing by ordeal, then, there is a continuum, from cases where the ordeal serves none but an allocative function, through instances where it confers some benefits on suppliers, to those where it amounts to a payment in kind to the supplier.

Allocation by Product Degradation

If the price of a good is controlled below the market-clearing level, producers frequently respond by reducing the quality and the cost of the good. (The opportunities for doing so depend on the nature of the good and whether the authorities tightly control its specifications as well as its price). This response tends in some cases to alleviate the shortage by causing both the supply and demand curves to fall: the former because of the lowered costs of producing the good, the latter because consumers are unwilling to pay as much for low as for high quality goods. Where the deterioration of the good or service is the result of reduced maintenance of productive capital, the response is commonly a way of evading restrictions on the withdrawal of capital from the price-controlled activity: reduced maintenance and, _in extremis_, abandonment of rent-controlled premises by landlords, is a contemporary example.

Product degradation often plays an important role in the rationing of certain free or subsidized services provided by governments, such as health and education services. The demand for publicly-provided services almost invariably exceeds the government's willingness to supply them, so that various means of rationing are introduced or arise spontaneously. With free

health services, rationing by queuing and by waiting is common. But the quality of the medical service itself also suffers, as doctors and hospital staff respond to the excess demand by giving shorter consultations, more perfunctory diagnoses, and shorter stays in hospitals, and as the government seeks to contain costs by restriciting the procedures that may be used and the drugs that may be prescribed. As the quality of the publicly-provided service falls, an increasing proportion of consumers find it unacceptable, and switch to private suppliers.

Perhaps the clearest examples of rationing by product degradation are provided by common-use facilities that are subject to congestion externalities or analogous phenomena. These include roads, parks, beaches, pools, etc; also utility services provided through interconnected networks. After a point, as more consumers use these facilities simultaneously, the quality of the service deteriorates. When access to a facility is free, degradation of service quality is the only form of rationing to which it is subject, ie. it is the only factor discouraging its use. Even when access is rationed in some way (eg. by admission charges) product degradation continues to have an allocative function. This is for two reasons. First, revenue maximization (or surplus maximization) would seldom dictate that the price be so high as to eliminate congestion entirely. Second, since the demand for the services of such facilities is usually subject to wide fluctuation, it is normally impracticable to ration optimally by the fine-tuning of price, so that the degree of congestion is allowed to fluctuate and play an active equilibrating role.

Potential consumers' willingness to pay for access to a common-use facility may be represnted by a family of demand curves--such as those labeled 0, 1, 2, and 3 in Figure 1--where successively lower curves pertain

to successively higher degrees of congestion (lower levels of service quality). The degree of congestion is assumed to be a simple technical function of the number of consumers using the facility simultaneously: This relationship is shown by the curve AB, which is drawn on the assumption that the facility can accomodate a number (OC) of users without any deterioration in service quality, but that further increments in the number of users causes quality to fall at an increasing rate.

If the price of entry were progressively lowered from a high level to zero, the curve HJ, showing the number of users and the marginal consumer's willingness to pay, would be traced out. Until the number of consumers reaches OC, HJ follows the demand curve 0; thereafter, as congestion sets in and gets worse, HJ moves through successively lower demand curves as dictated by the technical relationship, AB. At a zero price the number of consumers would be OJ, which represents the satiation point on a demand curve intermediate between those labeled 2 and 3.

In a behavioral sense, HJ is <u>the</u> demand curve: it shows the relationship between price and quantity demanded. However, it differs from the usual demand curve in that it incorporates the effects of service quality deterioration, and, as a consequence, the area beneath it does not (even approximately) measure consumer surplus, the usual measure of which is given by the area beneath the appropriate member of the family of underlying demand curves. (Thus at zero price, the surplus--the hatched area in the figure--is that associated with the curve intercepting the quantity axis at J.) HJ is better described as the locus of points consistent with both conditions of demand and conditions of supply--where "supply" is understood to mean the numbers--service quality relationship. Each point on HJ is a potential equilibrium point in that, for the given price and quality of service it represents, the

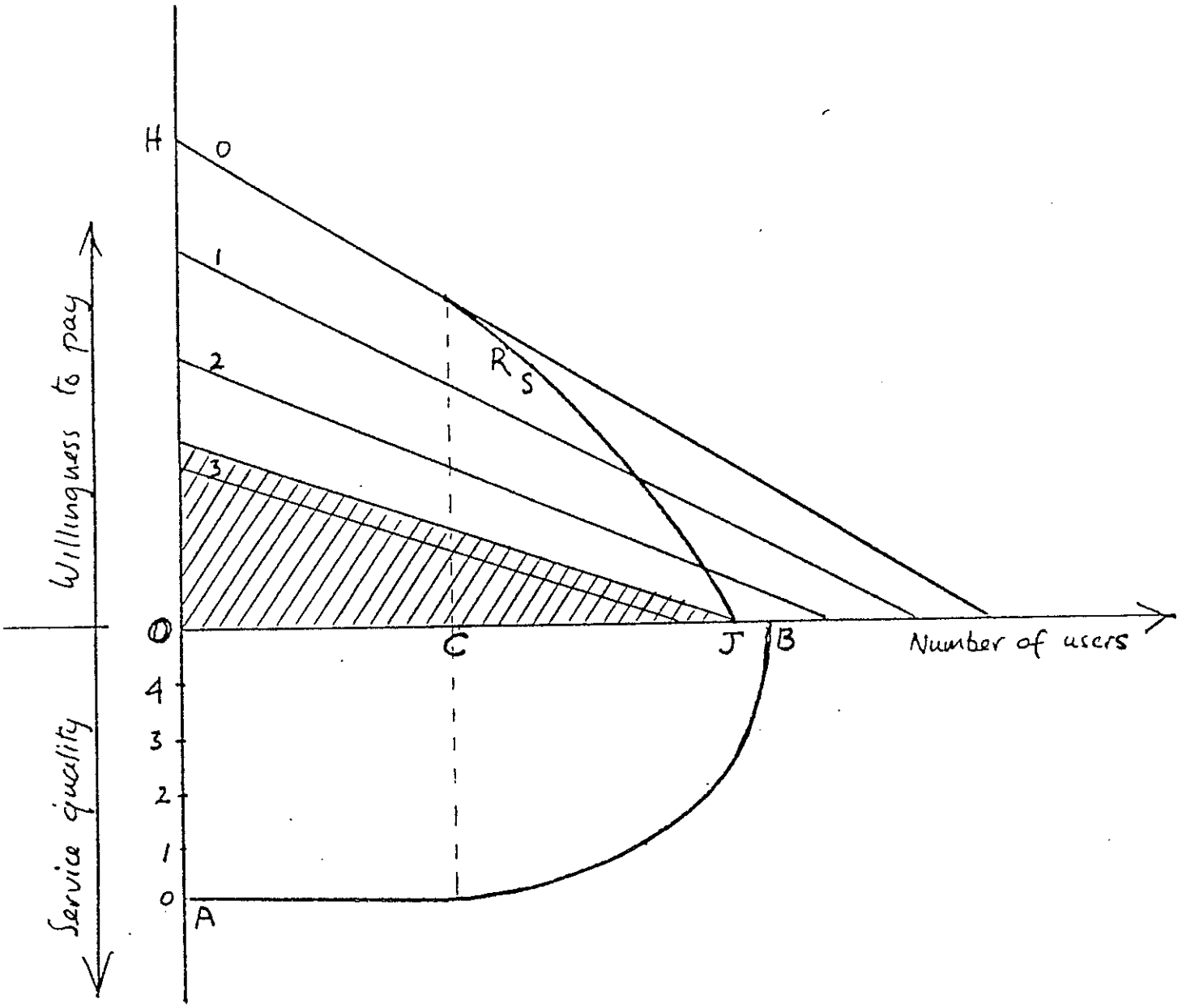
H
0
1
2
3
R_S
Willingness to pay
O
C
J
B
Number of users
4
3
2
1
0
A
Service quality

Figure 1

number of consumers who wish to use the facility simultaneously equals the maximum number that the facility can accomodate without the service quality falling below the given level.

Which of these possible quantity-quality equilibria will be chosen depends upon the price (money or equivalent) that is charged. If this is zero, we have exclusion solely by service degradation, and the equilibrium is at point J where the number of users is maximized but the level of service is minimized. Pricing so as to maximize revenue, or total surplus, would involve operating in the vicinity of points R and S, respectively, with fewer consumers and higher quality of service. In either case the total surplus generated would be considerably greater than obtained under rationing by service degradation.

Rationing by product or service degradation in the common-access facility case has obvious affinities with rationing by ordeal. (One is alerted to this by the fact that queuing and the manifestations of congestion are similar in a purely descriptive sense.) In each a considerable part of the potential economic surplus is dissipated. Also, since people differ in their sensitivities to ordeals, or to losses of service quality, the resulting allocations will be Pareto-inefficient. However, there is one important difference, justifying their separate classification: this is, that although it may be possible for a consumer to obtain the good by contracting out the ordeal to someone else, this is not possible in the case of product degradation. In the one case the ordeal is associated with the transaction, and may be dissociated from consumption of the good; in the other, undergoing ordeal-like experiences is inextricably bound up with the activity of of consumption.

3. Intermediate Methods of Allocation

This category covers allocations based on criteria--such as income, marital status, place of residence, quantity transacted last year, etc.--that are "innate" in the short run but behavioral in the long run. Some individuals would possess the relevant characteristics irrespective of whether it made them eligible to receive the good, while others acquire it solely to become eligible. As previously mentioned, many social services are allocated according to this type of criterion, and program costs are often underestimated because the elasticity of supply of the eligibility requirement with respect to the benefit it confers is also underestimated. The allocation of production, marketing, and import quotas under supply-control and demand-control programs also often belong to this category. Quotas are usually allotted according to each business unit's share of the relevant market in some base period. Such an allocation, if unexpected, is non-competitive. But if quotas are periodically imposed and removed according to market conditions, their imposition comes to be anticipated, and those affected expand their activities so as to acquire higher "bases." When ordinary consumer goods are rationed, the eligibility criteria may also be of this type. An example is the rationing of gasoline on the basis of either car ownership or possession of a driver's license. Allocation on the first basis would be expected to result in the registration of cars that would otherwise be scrapped, and some reallocation of the stock of cars, while rationing on the second basis would induce many non-drivers to get driver's licenses.

This method of distribution initially imposes an allocation upon potential consumers. As with non-competitive allocations generally, the extent to which the rationed good (R) is misallocated depends on the association between possession of the qualifying characteristic (Q) and valuation

of the ration. The initial allocation is subsequently modified, as potential R consumers qualify themselves to receive a ration by acquiring Q. As the number of eligible recipients increases, the supply of R has to be increased, the eligibility requirement tightened, or the individual ration reduced. Expansion of Q production and reallocation of R is ultimately checked by declining valuation of R and Q, by rising costs of Q, and (possibly) by reductions in the ration or tougher eligibility requirements. The final outcome is characterized by the following:

Excess production of Q: The rationing scheme has the effect of subsidizing Q. Unless the subsidy (R ration) corrects an existing distortion, the additional production of Q is worth less than its costs of production, resulting in a welfare loss.

Misallocation of both R and Q: Instead of each good being consumed by those that value it most highly, the two are tied in consumption and both are consumed by those for whom the value of the Q+R bundle exceeds the costs of acquiring Q. Thus a low valuation of R (or Q) does not prevent an individual from obtaining it provided his valuation of Q (or R) is sufficiently high. The misallocation will be less, the higher the rank correlation between consumers' valuations of the R ration, and the surplus (positive or negative) that they derive from the qualifying quantity of Q. The discreet sizes of the ration and the qualifying requirements are further sources of misallocation of the two goods.

Competition for R involves real costs. These are incurred by those who acquire Q, or retain Q, or acquire more Q, so as to be eligible to receive R, and in each case consist of the excess of the cost over the valuation of the quantity of Q involved. These deficits on Q offset in part the surpluses on R. Acquisition of the eligibility-conferring good or status is, for those who would not otherwise possess it, a species of ordeal.

The competitive process also engenders transfers--of R from initially selected to newly-qualified consumers; and (if the price of Q is bid up) of surplus from Q consumers to Q producers.

It is reasonable to assume that the competitive reallocation of R raises its aggregate value. It gives those who value R highly the opportunity to obtain a ration, and disfavors those whose willingness to pay for R is low; however, both of these effects are attenuated, to a greater or less extent, by the incidence of offsetting valuations of Q. Reduction in the size of the ration and the concurrent wider distribution of R also tends to raise R's average value, since marginal units given up by original recipients become intramarginal units for the new recipient. But even if the final allocation of R is more efficient than the initial allocation, this is insufficient to ensure that the whole process is advantageous: in addition, the benefits of the reallocation of R have to exceed the costs associated with the excess production and misallocation of Q.

Summary and Comparison of Methods

The classification scheme developed above is set out in the accompanying table.

In comparing the various methods of allocation, the following points deserve to be emphasized:

1. Only price is capable, unaided, of <u>both</u> allocating a good among potential demanders <u>and</u> equilibrating demand and supply. However, non-price manifestations of competition among demanders, such as queues and waiting lists, can and do play a supplementary and short-run coordinating role. Also, in those (no doubt rare) cases where rationing by ordeal amounts to a form of payment in kind, the ordeal is, in priciple, capable of equilibrating the market.

Classification of Allocation Methods

Method of Allocation	Allocative Criterion	Examples
1. Ostensibly Non-Competitive	"Innate"	
	Age, Sex, Race	Affirmative Action programs
	Heredity	Citizenship, admission to schools
	Random Selection	Military conscription
	Equal shares	Food in wartime
2. Competitive	"Behavioral"	
a) Price	willingness to pay	Numerous
b) Ordeal	willingness to undertake ordeal	
(i) No benefit to supplier:		
Queuing		Numerous
Waiting		Public housing, justice
Other		hand-held hoses
(ii) Beneficial to supplier:		
Screening		Matriculation examinations
Payment in kind		Occupational licenses to long-service employees.
c) Product Degradation	willingness to accept low quality	Evasion of price control.
Ordinary goods		Nation Health Service, Public education
Publicly-provided services		Roads, beaches, utility networks
Common-access facilities		
3. Intermediate	"Innate" in short run, "Behavioral" in long run.	
	Income, marital status	Social services
	Occupational status	Special shops for high officials in Soviet Union
	Historical market shares	Marketing quotas, import quotas.

2. Price rationing is capable, in principle, of bringing about a Pareto-efficient distribution of the good. Also, to a greater extent than any other method, it incorporates an incentive structure conducive to efficient allocation.
3. Rationing by ordeal allows intensity of individual preferences to be expressed, but only through participation in a specific ordeal, the irksomeness of which will vary from person to person. Hence it cannot achieve an efficient allocation unless the ordeal is contracted out to specialists at a common money price, ie. unless it is supplemented by the price mechanism. An additional source of inefficiency is the weakness of the incentives operating to uphold the "law of one price" in ordeal markets.
4. Similar remarks apply to rationing by product degradation, except that the possibility of contracting out the competitive activity does not exist.
5. Non-competitive allocation is not based on individual preferences and hence cannot achieve a Pareto-efficient distribution. Many allocations are imposed precisely because a different outcome is desired by the allocating authority. In other cases the motivation is not so much to secure a particular allocation as to avoid the income-distributional consequences of price rationing. (The distributional aims can be accomplished but allocational inefficiency reduced by taking measures to facilitate the subsequent voluntary reallocation of the good--such as using transferable coupons as the rationing instrument.)
6. Under most forms of non-price rationing (and under price rationing too, in certain monopoly situations) allocative inefficiencies result from the discrete size of individual rations. With competitive non-price methods, discreteness takes the form of an imposed transaction limit,

whereby consumers may be subjected to all-or-nothing choices. A similar reduction in the range of choice occurs under rationing by product degradation, though here it applies to qualities rather than quantities.

7. Administrative and enforcement costs are likely to be high for non-competitive allocation schemes. However, as a partial offset, consumer's decision-making costs will be low, since, apart from identifying themselves and perhaps proving their eligibility, their role is a passive one.

8. Price allocation can also be described as allocation by exchange, and it is only with price rationing that the potentialities of exchange are realizable in full. Some forms of ordeal represent payments in kind to the supplier, and hence in these cases rationing by ordeal amounts to barter exchange. There is also an element of exchange present when the ordeal acts as a screening device. Apart from these cases, competitive non-price methods of allocation do not involve exchange, even though they do involve trade-off: that which is traded-off for the good is simply wasted. This dissipation of economic surplus via the competitive process, whether it be by ordeal, product degradation, or the acquisition of an eligibility-conferring good or status in the major cost of these methods, but they do give rise to misallocation as well.

In brief, non-competitive methods of allocation are Pareto-inefficient and may also be costly to administer and enforce, while competitive non-price methods may allocate more efficiently but waste real resources in the competitive process. Some idea of the relative importance of these costs of non-price allocation (excluding administrative and enforcement costs) can be gained by comparing, with the aid of simplifying assumptions, the loss of

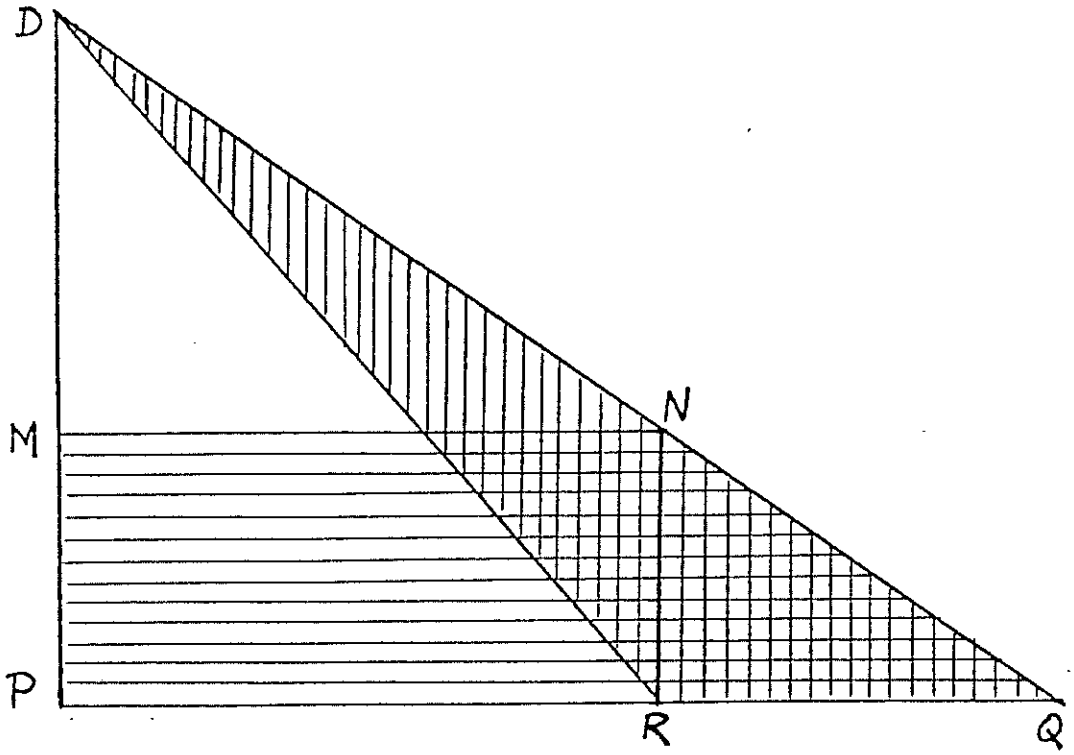

Figure 2

economic surplus associated with a reduction in the supply of a good, under various rationing methods. The losses are shown as hatched areas in Figure 2 and are explained below. The analysis is based on the following assumptions:

(i) DQ is the demand curve for a good that is demanded in unit quantity per time period by each potential consumer, with different consumers valuing the good differently.

(ii) Q is the quantity supplied initially and P is the initial price.

(iii) The supply--which is absolutely inelastic--is subsequently reduced to R and the good is rationed, (a) by allowing price to rise to its new equilibrium level, M, or, with price fixed at P, (b) by random allocation, or (c) by queue rationing.

(iv) All consumers value time equally in terms of money.

(v) Time spent in the queue is the only cost of queuing.

(vi) Consumers are sufficiently knowledgeable and experienced for an equilibrium queuing time to be established.

The loss of surplus under price rationing is the "welfare triangle" NQR. Consumers also lose the "revenue rectangle" MNRP, but this is a gain to sellers.

The expectation is that, under random allocation, those excluded would value the good as highly, on average, as those selected to receive it. Since each consumer consumes one unit, the total surplus would be reduced in the same proportion as the supply was reduced. The new surplus would therefore be DRP, and the loss of surplus the vertically-hatched triangle DQR.

On the given assumption, an equilibrium queuing-time price, equivalent to the equilibrium money price M, would be established. The loss of surplus under queue rationing would therefore be the horizontally-hatched area, ie.

the welfare triangle NQR PLUS the area MNRP, which is no longer an amount transferred to sellers, but a measure of the aggregate cost of the time spent queuing.

The loss of surplus is substantially greater with non-price rationing than with price rationing. Also, as compared with the surplus lost under price rationing, the additional loss under queue rationing is twice as great as the additional loss under random allocation. (The area of the rectangle MNRP is twice the area of the triangle DNR.) While this result is specific to linear demand curves--with convex curves, the comparative advantage of random allocation is reduced, and if they are sufficiently convex, negated[2]--it is nevertheless of interest in showing how wasteful competitive activities of a rent-seeking nature can be, even if they do result in a better (or, as in this case, a perfect) allocation.

Note that the loss of surplus with product degradation would be the same as with queue rationing if all consumers valued service quality identically. The members of the family of demand curves would then be parallel to one another. Imagine an initial equilibrium at quantity Q in Figure 2, followed by a reduction in the capacity of the facility and concomitant contraction of the technical relationship AB such that a new equilibrium was established at R, the end-point of a demand curve parallel to DQ. The loss of surplus would then be the area between DQ and the lower demand curve= which area, because of the parallelism assumption, would be equal to the area MNQT.

It is also worth noting that if there was a positive correlation across consumers between valuation of time and valuation of the good, the loss of surplus under queue rationing would be greater than as shown in Figure 2. (And similarly for the product degradation case, if the size of the premium for service quality was positively correlated with willingness

to pay for the service). Also, as Barzel has pointed out (Barzel, 1974), if the valuation of time rose faster (across consumers) than the valuation of the good, consumers with the highest valuation of the good would be excluded by queue rationing.

Footnotes

1. Queue rationing, and rationing by waiting, are quite distinct. In the former, the cost of the ordeal consists of the value of other activities given up while standing in line, plus or minus any inconvenience or pleasure derived from the queuing experience. The costs of the latter include the postponement of consumption, the additional foresight required in planning consumption activities, and, in the case of location-specific goods--such as public housing--the opportunities foregone by being "locked-in" to a particular location.

It is unfortunate that--presumably because "queue" and "queuing" are not American vernacular--the standard analysis of queue rationing (Barzel, 1974) should be entitled "A Theory of Rationing by Waiting."

2. Essentially this is because, with convex curves, more of the potential surplus derives from the higher reaches of the demand curve, where it is vulnerable to random allocation, but relatively untouched by queue rationing.

References

Barzel, Yoram, "A Theory of Rationing by Waiting," Journal of Law and Economics, XVII(1) (April, 1974), pp. 73-95.

Swan, Peter, On Buying a Job: The Regulation of Taxi Cabs in Canberra, Centre for Policy Studies, Policy Monograph No. 1 (Turramurra, 1979)