CHAPTER 7

Rationing Services: Limitation of Access and Demand

Theoretically there is no limit to the demand for free public goods. Agencies that provide public goods must and will devise ways to ration them. To ration goods or services is to establish the level or proportions of their distribution. This may be done by fixing the amount or level of goods and services in relation to other goods and services. Or it may be done by allocating a fixed level or amount of goods and services among different classes of recipients. In other words, services may be rationed by varying the total amount available, or by varying the distribution of a fixed amount.

This usage is consistent with the familiar application of rationing in wartime. During World War II, for example, automobile tires were rationed by restricting their production for domestic purposes and limiting individual purchases, making them costly, and establishing priorities among users (doctors were privileged in this respect). This chapter considers rationing in street-level bureaucracies that has the effect of fixing (usually to reduce or limit) the level of services. The next chapter takes up rationing that differentiates among clients.

The rationing of the level of services starts when clients present themselves to the worker or agency or an encounter is commanded. Like factory workers confronted with production quotas, street-level bureaucrats attempt to organize their work to facilitate work tasks or liberate as much time as possible for their own purposes. This is evident even in those services areas in which workers have little control over work flow. For example, police often cannot control work flow because most police assignments are in response to

citizen initiated calls.[1] Dispatchers, however, make every effort to permit officers to finish one call before beginning another. Officers often take advantage of this practice by postponing reporting the completion of a call until after they have finished accumulated paperwork. In this way police officers regularize the work flow despite substantial irregularity in requests for assistance.

The way in which work comes to the agency significantly affects the efficiency and pleasantness with which it is accommodated. Official efforts to influence the flow of work vary greatly. They range from the mild advisory of the post office providing patrons with information concerning the times when delays are likely to be longest, to the extreme measures taken by a New York City welfare office that closed its doors at noon rather than admit a greater number of Medicaid applicants than could be processed by available personnel in an eight-hour day.[2]

Clearly there are costs to clients in seeking services. In both of the above examples agencies seek to inform clients of the costs and the problems they will encounter—in the first instance, if they seek assistance during days when post office patronage is heavy; in the second, if in ignorance of the situation they attempt to apply for Medicaid and cannot be accommodated because of the high intake demand relative to intake workers. In many instances even the failure to inform clients of likely costs in seeking service constitutes a consumer complaint.

The highest costs are borne by potential clients who are discouraged from or forbidden access to bureaucratic involvement. While exclusion from client status is usually accomplished on the basis of legal grounds, the population of the excluded or discouraged includes many whose exclusion is a matter of discretionary judgment. The ineligibility of tenants evicted from public housing, students expelled from school, or welfare claimants deemed uncooperative depends not on fixed criteria alone, but also on interactions with street-level bureaucrats.

The Costs of Service

To analyze individual influence it has sometimes proved useful to recognize the relationship between citizens' influence and their command of personal resources such as money, status, information, expertise, and capacity for work.[3] People who have these resources tend to be more powerful than

those who do not. When people have them they enhance personal influence. When workers for public agencies have them they may be used to direct or subordinate clients or discourage clients from further interactions with the agency.

MONETARY

Street-level bureaucracies can rarely assign monetary costs for services, since by definition public services are free. However, monetary costs *are* imposed in several instructive instances. In income-providing programs citizens' contributions to the income package may be manipulated as policy. Medicare patients may be asked to pay a higher deductible before insurance provisions become operable. Food-stamp recipients may be asked to pay more for their stamps. The effective taxation of earned income in welfare reduces the number of people in contact with this street-level bureaucracy. Clearly differences in monetary costs serve to ration street-level bureaucrats' services.

Programs sometimes force clients to incur monetary costs that discourage them from seeking service. Acquiring records from other agencies to establish eligibility or securing transcripts for appeals can be costly, particularly if travel is involved. Agencies that keep bankers' hours impose monetary costs on working people who cannot appear without losing wages. Appointments sometimes require parents to seek babysitters. Street-level bureaucracies that seek to minimize these penalties introduce evening office hours, or they provide child-care services.

TIME

Just as available time is a resource for people in politics, it is also a unit of value that may be extracted from clients as a cost of service. Clients are typically required to wait for services; it is a sign of their dependence and relative powerlessness that the costs of matching servers with the served are borne almost entirely by clients. It is to maximize the efficiency of workers' time that queues are generally established. A primary reason that clinic-based practice is more efficient than home-based practice is simply that it is patients and not physicians who spend time traveling and waiting. Policemen also allocate time costs by stopping to question young people who, while not guilty of any crime, are judged to require reprimanding.[4]

Some teachers in some school systems make home visits to meet with parents, while others schedule parent-teacher conferences after school on specific days set aside for such purposes. (If there are two parents and one or both work, both are unlikely to be able to meet with the teacher.) These al-

ternative perspectives on parent-teacher conferences measure significant differences in the value placed on time of parents and teachers.

Time costs are often assessed by street-level bureaucrats as delay; they are often experienced by clients as waiting. Bureaucracies can reward clients by expediting service, punish them by delaying service. Court postponements can function in this way, as can an increase in the time between intake interviews and placement on the welfare rolls. Importantly, bureaucracies often have little interest in reducing delay, since more expeditious processing would simply strain available resources.

Assessed time costs may also be experienced as inconvenience, although they are levied as procedure. For example, when an agency refuses to receive complaints over the telephone and requires that they be written, it may cut off complaints lodged frivolously or on impulse, but also discourages complainants who would protest if it were easier.[5] Requirements to complete multiple forms and produce extensive documentation function similarly. It is possible to make an argument that since the real costs of delay and elaborate procedures are the activities foregone while waiting, that is, opportunity costs, it is justifiable that poor people wait longer than the more affluent, since the opportunities foregone are less valued by the society.[6] However, at the very least this elitist view is based on a calculus to the terms of which clients have not consented.

INFORMATION

Giving or withholding information is another way in which services may be rationed. Clients experience the giving or withholding of information in two ways. They experience the favoritism of street-level bureaucrats who provide some clients with privileged information, permitting them to manipulate the system better than others. And they experience it as confusing jargon, elaborate procedures, and arcane practices that act as barriers to understanding how to operate effectively within the system. The emblematic carrier of this characteristic is the court clerk who runs his words together in an undecipherable litany to the dominance of court procedures over citizens' rights.[7] At the bureaucratic level the giving and withholding of information is most obvious in examining how agencies manipulate their case loads by distributing or failing to distribute information about services.

Conventionally, analysts assess the demand for services by studying client rolls and visits. (Demands are statements directed toward public officials that some kind of action ought to be undertaken.)[8] If it is recognized that manifestations of client involvement may not fully reflect client interests,

analysts contrive ways to assess underlying needs, for example, through attitudinal and census surveys. From this assessment administrators and politicians make claims about appropriate levels of services.

However, if it is recognized that organizations normally ration services by manipulating the nature and quantity of the information made available about services, then it is easily seen that demand levels are themselves a function of public policy. Client rolls will be seen as a function of *clients' perception* of service availability and the costs of seeking services. Client demand will be expressed only to the extent that clients themselves are aware that they have a social condition that can, should, and will be ministered to by public agencies.

When New York City reduced acceptance rates for new welfare cases at seven centers by 17 percent it accomplished this feat by tightening the application process. This meant not only more careful scrutiny of applicants' claims, but also more documentation and inquisition was required, which contributed a separate measure of rationing.[9]

This perspective is illustrated by indices of need for legal assistance for domestic problems. When a sample of Detroit residents were asked if they required a lawyer for assistance with some domestic-relations matters, scarcely more than 1 percent answered affirmatively. It would have been difficult to predict from this survey that approximately 40 percent of the clients of legal aid and neighborhood law offices originally sought help with domestic problems.[10]

Needs become manifest when the institutions that might provide assistance send out signals that they stand ready to assist. The 40 percent of the clients who originally sought help with domestic matters might have been only a small portion of the population that could have benefitted from such assistance. Some who could have used such services may have been deterred from seeking them. Since legal services are vastly underfunded, even more dramatic demonstrations of need might have materialized if more lawyers had been available.

Information about service is an aspect of service. Withholding information depresses service demands. For example, the campaign to reform welfare by dramatically increasing the welfare rolls was based on the view that a political movement could help overcome the stigma attached by potential recipients to welfare status. It could provide the information necessary to realize a substantial increase in the number of recipients.[11] The failure of public welfare agencies to make sure potential recipients receive the benefits to which they are entitled contrasts dramatically with the success of social security

and Veterans' Administration benefits. The difference is that the clients of these two income support programs—the elderly, and veterans—are not socially stigmatized.[12]

Client statistics may not indicate much about the objective needs of the client population but they reflect a great deal about the organizations that formally cater to those needs.[13] Thus growing demand for adult continuing education partly exists in the felt needs of the adult population, but the demand also is responsive to the publicity generated by colleges and universities and their desire to attract students and their tuition. The demand for emergency police services exists to an unknown degree, but the introduction of a 911 central telephone number and dispatch system makes it more likely that citizens believe the police will respond quickly. After the system is introduced the increase in 911 calls will be responsive to organizational factors such as publicity about the service and response time as well as more objective factors such as population growth and changes in the age distribution of the population.

Although the dominant tendency is for street-level bureaucracies to attempt to limit demand by imposing (mostly nonmonetary) costs for services, there are some times when they have a stake in increasing their clientele. They will do this through an analogous rationing process, now directed toward increasing utilization.

Agencies are likely to try to increase their clientele when they are newly established and have to prove their ability to put services into operation. Thus the tripling of service complaints when Boston introduced its Little City Halls program was particularly welcome by its sponsors.[14] Efforts to increase clienteles were generally noticeable when central funding sources launched many subordinate service agencies, which saw themselves competing for funds in the next fiscal cycle. Such agencies would "beat the bushes" for clients in order to demonstrate that they were worthy of future support. Community action agencies and neighborhood mental health centers have been cases in point.[15]

Established street-level bureaucracies may also attempt to increase their clientele if they perceive themselves under attack and calculate that demonstrations of significant service provision, or increases in clientele, might aid their cause. Relatedly, street-level bureaucracies may attempt to increase the number of clients when they are competing against other programs with similar objectives. Such agencies perceive that they are competing for the same client pool, and that only the more successful will survive in the next budget cycle.

This competition also is conducive to quasi-legitimate fraud directed to-

ward making the agencies look better. For example, when drug treatment centers were few they could afford to impose rigorous residential requirements, particularly since clients' commitment to their own rehabilitation was considered critical to therapy. When the number of such institutions increased in the early 1970s in response to available funding, and the population of drug users started to decline, to increase their clientele the centers began to relax their enrollment requirements (for example, by accepting clients who previously would have been judged too difficult to help). They also relaxed attendance requirements, so that a treatment bed might be occupied by someone who was not in fact a full-time resident of the center. Besides drug treatment centers, other organizations that have competed for larger shares of a fixed client pool include mental health centers funded in the same city, and academic departments competing for students within a university.

In theory this bureaucratic competition might provide precisely what bureaucracies importantly lack—a substitute for market place accountability. This, of course, is the idea behind educational vouchers. However, the healing effects of competition are too often mitigated by the residual bureaucratic aspects of the competing organizations. Faculty members in academic departments with declining enrollments are still protected by the tenure system, rewards for research (and bringing in research grants), and other factors that protect them from being assessed solely on criteria of service to students. Similarly, educational voucher experiments have foundered on teachers' tenure, union opposition, and parental inability to express preferences within the system for lack of information on the implications of the available choices.

PSYCHOLOGICAL

Bureaucratic rationing is also achieved by imposing psychological costs on clients. Some of these are implicit in the rationing mechanisms already mentioned. Waiting to receive services, particularly when clients conclude that the wait is inordinate and reflects lack of respect, contributes to diminishing client demands.[16] The administration of public welfare has been notorious for the psychological burdens clients have to bear. These include the degradation implicit in inquiries into sexual behavior, childbearing preferences, childrearing practices, friendship patterns, and persistent assumptions of fraud and dishonesty.[17] Nor have these practices been confined to the "unenlightened" 1950s, although some of the more barbaric features of welfare practice, such as the early dawn raids to catch the elusive "man-in-the-house," are no longer practiced.

To take a modest example, women applying for Aid to Families with Dependent Children at times are required to submit to an interview with lawyers, in which they must agree to assist the welfare department in prosecuting the father of their children. Apparently many women are unwilling to agree to this, since it would jeopardize the tenuous but at least partially satisfactory relationship that they may have with the childrens' father. They fear that the support they currently do receive and the positive benefits of good relations with them would be cut off by alienating the fathers, who may not be making substantial incomes anyway. Applicants are thus forced to lie or risk the loss of an important relationship. The interviews are conducted in a legalistic way with little sympathy for the position of the applicant. Many eligible potential clients do not complete the application process, because they prefer not to suffer these pressures and indignities.[18] Like so many monitoring precedures in welfare, it is unclear if monies recovered through these procedures equal the costs of engaging in them.

Psychological sanctions serve to reduce the demands from clients within the system as well as help to limit those who come into it. The defendant in a lower criminal court who asserts that he or she does not understand the charges will be silenced by the hostile response of the judge or clerk who unenthusiastically attempts to redress the complaint. Teachers, by varying their tone of voice, encourage or discourage pupils from asking questions. A lawyer in responding to clients can communicate the opinion that the inquiry is stupid and the client unworthy of a thoughtful response.

The importance of psychological interactions for rationing service is manifest in the extent to which clients will sometimes seek or approve of service simply because they like the way they are treated. Although they later find against them, sympathetic judges sometimes give thoughtful attention to defendants or complainants with weak cases simply in order to make them feel that they had their day in court. The reported gratitude of citizens who are treated in this way may indicate how little people have come to expect from government. It would seem that clients sometimes judge services positively if they are treated with respect regardless of the quality of services. In this connection a study of clients' evaluation of walk-in mental health clinics revealed that "clinic applicants are satisfied with almost any response [from staff] at first so long as the emotional atmosphere of the contact is comfortable."[19] While seekers of mental health services may be particularly sensitive to the quality of initial client-staff interactions there is every reason to think that these interactions form a substantial part of clients' initial evaluations of schools, courts, police, and other street-level services where there are no clearly defined service products to be obtained.

Rationing Services: Limitation of Access and Demand

Queuing

The most modest arrangements for client servicing impose costs on clients. This is evident in the way clients are arranged, or required to present themselves, for bureaucratic processing. Even the most ordinary queuing arrangements—those designed to provide service on a first-come, first-served basis in accordance with universalistic principles of client treatment—impose costs.[20]

Queues that depend upon first-come, first-served as their organizing principle elicit client cooperation because of their apparent fairness, but they may ration service by forcing clients to wait. When clients are forced to wait they are implicitly asked to accept the assumptions of rationing: that the costs they are bearing are necessary because the resources of the agency are fixed. They are also controlled by the social pressures exerted by others who wait. This is one of the functions of the line, waiting room, and other social structures that make it evident that others share the burden of waiting for service.

While resource limitations may be unalterable in the very short run, they are not necessarily immutable. They derive from allocation decisions that consider it acceptable to impose costs on waiting clients. Costs will not be imposed upon clients equally. Long lines processed on a first-come, first-served basis relatively benefit people who can afford to wait, people whose time is not particularly valuable to them, or people who do not have other obligations.

Poor people often suffer in such a system. Not only may clients who appear more affluent get served first because it is thought that the costs of waiting are higher for them,[21] but agencies often paternalistically develop policy as if the costs to the poor were nonexistent. A visit to the waiting room of a welfare office in any inner-city neighborhood is likely to convey the impression that the Welfare Department assumes recipients have nothing else to do with their time. Recipients learn the lesson of people who must seek service from a single source. Like the telephone company, the welfare department is able to pass on to the customer the costs of linking people with service. This system also benefits the average client to the disadvantage of people with extraordinary needs, since initially it has no mechanism for differentiating among clients. However, where the injury to people with extraordinary needs is likely to be severe, as in police work or medical emergencies, the ordering of services is often deliberately structured to search for and respond to this information.

An alternative to the first-come, first-served waiting room or line is the first-come, first-served queue by appointment. This system is also normatively acceptable and theoretically has the advantage of eliminating many of the costs of waiting time. In this queue the costs may appear to be reduced for the average client, but they may still be significant if appointments are crowded together to insure client overlap, as is typically done in health clinics and other medical settings. Crowding appointments may be done for the convenience of bureaucrats whose time is considered more valuable than that of clients, and who thus are guaranteed a flow of clients even if one misses an appointment. The costs of such a queue will also be borne by clients who seek service but cannot afford to wait for it, who are not disciplined enough to make and keep appointments, or who are not sure enough of the likely benefits of service to invest in seeking it. What appears to the street-level bureaucrat as a fair way to allocate time may be seen by the client in the light of past experiences of bureaucratic neglect and taken as a sign that the agency is unlikely to be responsive, or that the problem is unlikely to yield to assistance.

For some clients the costs of waiting may be quite high. In one legal services program approximately 40 percent of eligible clients who received an appointment with a lawyer for the following week did not keep the appointment.[22] This may have been because the problem dissolved during the intervening time, or because merely talking to the intake worker provided a degree of comfort. However, it is equally likely that clients who did not keep their appointments could not keep them but were afraid to say so, were not organized enough to show up at the appointed time, or faced their legal problems without professional advice. Or it may have been that the applicants for assistance interpreted the demand to wait for appointments as a sign that legal services was not likely to be responsive and assumed that, like other public agencies, it would not in the end prove helpful.

In any event, the day a client appears to seek assistance may be the day when he or she is most open to help or the street-level bureaucrat is most likely to be able to intervene successfully. Catherine Kohler Reissman has written about mental health services in an analogous situation.

> It is obvious that the disequilibrium created by a crisis is a powerful therapeutic tool that is lost if the situation is allowed to degenerate, through postponement, into a chronic, long-term problem.[23]

Similar to the queue by appointment is the waiting list; clients are asked to wait for what is usually an undetermined amount of time until they can be accommodated. Although it appears to be straightforward on the surface, the

waiting-list system has several important latent functions. First, as we have seen in the case of Boston public housing, a waiting list tends to increase the discretion of street-level bureaucrats by providing opportunities to call clients from the waiting list out of turn, or to provide special information that will permit them to take advantage of ways to be treated with higher priority.[24] Waiting lists also permit agencies to give the appearance of service (after all, clients *are* on a waiting list) and to make a case for increased resources because of the backlog of demand.[25] The waiting list appears to record the names of potential clients who are seeking service but cannot be accommodated, although it is obvious to all that many names continue on the list only because the agency has not attempted to discover who is actively waiting and who has long since ceased to be interested.

Some social agencies act as if the waiting list usefully filters potential clients who are truly in need of service and strains out those whose needs are not substantial and who thus drop off. This system of rationing may also provide for a period of time in which spontaneous recuperation may occur, again reserving client spaces only for those who are needy.[26] However, it is uncertain whether continuation on the list is a sign of substantial need or precisely the opposite, a sign that the potential client is successful enough in managing the problem that he or she can wait patiently for services.

A queuing arrangement that maximizes the costs to citizens at the expense of a relatively small number of street-level bureaucrats is employed by lower courts, which typically require defendants to appear on a given day, but notify them only as to the hour they should appear. In a typical situation fifty to one hundred defendants, possibly with a friend or member of their family, must be ready for a hearing or arraignment, with substantial penalties if they do not appear precisely at the beginning of the session (when their names are first called). Here they must wait until the judge arrives, and then wait again while the judge gives priority to defendants in the lockup who may require attorneys, defendants whose attorneys plead that they have to be elsewhere, and defendants whose cases require the testimony of waiting police officers, who themselves are subject to other priorities. Only when these and other priorities are accommodated will the docket be called in alphabetical or some other order.

Defendants may be innocent but by virtue of being arrested are judged guilty enough to pay in time and uncertainty the price that the court exacts for scheduling cases for the primary convenience of the judge. Although practices vary from court to court it is typical that defendants will not be told even approximately when their cases will be called, so that they must wait in the courtroom, possibly for most of the day, until they receive a hearing.[27]

The defendant who has waited through such a day has been instructed in the costs of continued interaction with the court system and must consider whether exercising rights or even pleading innocent in a minor matter, although legally valid, is worth the time and irritation. Some court systems have recently recognized that similar problems, including frequent postponements, inhibit witnesses from appearing and testifying in trials. But the same analysis rarely focuses on defendants and their experiences in court.

This queue by roundup is also typical of jury impaneling, where citizens are called for a week of service and must sit in a jury room awaiting assignment, often for several days, perhaps never to be called. The system officially is justified by the fluctuating and relatively unpredictable demand for jurors, and again is premised on the high value placed on the court's time relative to citizens' time. To insure that there are always people ready to serve, more jurors are called than will be required. If the court could tolerate a postponement now and then for lack of available jurors, and if jurors were called to report serially during the week rather than all at once, less time would be wasted for prospective jurors. But such practices could only be adopted if the time of prospective jurors were accorded more value relative to judges' and lawyers' time than is currently the case.

Clients frequently may be quite willing to pay the costs of waiting. Clients undoubtedly understand that there are times when they will have to wait, unless bureaucracies hire enough staff to meet peak demand. And since demand in most street-level bureaucracies is to some degree unpredictable—even schools often have to hire new teachers or shuffle teacher assignments after school has started—it would be too costly to provide services so that waiting would never occur. Waiting becomes injurious and inappropriately costly only under certain conditions.

Waiting is inappropriate when it exceeds the time generally expected for a service. A person may not resent a two-hour wait in an emergency room to receive a tetanus shot if it is clear that patients with more serious claims are being served first. But the same amount of time spent waiting in line simply to hand in forms to renew a driver's license may be exceedingly irritating. Waiting may also be resented when it involves the violation of an implicit agreement. Waiting is regarded as inappropriate when clients have made an appointment, except when the appointment is considered only an approximation of the time of service (as in the case of office visits to doctors).

Still another situation in which clients resent the costs of waiting arises when they wait unfairly. Thus if a favored client gains access to service more easily than others it will be resented by those who are not favored. Sometimes unfairness in waiting time may be so slight as to go unnoticed by

clients. A study of black patients in Chicago hospital emergency rooms revealed that compared to whites waiting time was a little more than three minutes, incurred primarily by claimants with nonemergency conditions who sought help when the emergency room was relatively busy. But this cost is not actually trivial. It is worth noting that a modest three minutes or so, *for the 1,105 blacks in the sample alone,* would add up to a full working day for 2,619 people on a yearly basis,[28] a measure of one of the costs of institutional racism for the blacks of Cook County, Illinois.

Routines and Rationing

The existential problem for street-level bureaucrats is that with any single client they probably could interact flexibly and responsively. But if they did this with too many clients their capacity to respond flexibly would disappear. One might think of each client as, in a sense, seeking to be the one or among the few for whom an exception is made, a favor done, an indiscretion overlooked, a regulation ignored.

This dilemma of street-level bureaucrats is illustrated well by the legal services program. Individually, each attorney is obliged by professional norms to pursue fully the legal recourses available to clients. For impoverished clients this presumably means that attorneys should act on clients' behalf irrespective of cost. Only if this assumption is correct could the provision of legal services begin to redress the balance of power in the legal system, which every observer concedes favors those who command legal resources. But if all clients' legal needs were fully pursued there would be no time for additional clients. The dilemma is exquisite. To limit lawyers' advocacy is to deny poor people equal access to the law. To permit unbounded advocacy is to limit the number of poor people who can have such access. Only a reconstitution of the legal system could overcome the dilemma within the current patterns of inequality: either a radical departure in the amount of subsidies for legal assistance for the poor or a radical simplification of legal procedures.

When confronted with the dilemma of serving more clients or maintaining high quality service, most public managers will experience great pressures to choose in favor of greater numbers at the expense of quality. Their inability to measure and demonstrate the value of a service, when combined with high demand and budgetary concerns, will tend to impose a logic of increas-

ing the quantity of services at the expense of the degree of attention workers can give to individual clients. Street-level bureaucrats, however, may devise ways to sabotage management efforts to reduce interactions with clients. The costs of achieving compliance in the face of workers' resistance may sometimes be more than managers want to pay. An example of such worker resistance is related by Robert Perlman in his study of the Roxbury Multi-Service Center.

> Confronted with the complexity and number of demands being made on them, staff members resorted to shielding themselves from the mounting pressures. They extended interviews to postpone or avoid taking the next client. They scheduled home visits in order to avoid intake duty.[29]

Whether street-level bureaucrats oppose efforts to limit their interaction with clients, or whether they accept and encourage such efforts as a way of salvaging an unattractive or deteriorating work situation, is perhaps the critical question on which the quality of public service ultimately depends. Although street-level bureaucrats may sometimes struggle to maintain their ability to treat clients individually, the pressures more often operate in the opposite direction. Street-level practice often reduces the demand for services through rationing. The familiar complaints of encountering "red tape," "being given the run-around," and "talking to a brick wall" are reminders that clients recognize the extent to which bureaucratic unresponsiveness penalizes them.

Routinization rations services in at least two ways. First, set procedures designed to insure regularity, accountability, and fairness also protect workers from client demands for responsiveness. They insulate workers from having to deal with the human dimensions of presenting situations. They do this partly by creating procedures to which workers defer, happily or unhappily. Lawyers and judges, for example, generally accept court procedures that insulate them from erratic client demands. Police officials resist instituting (or more properly, reinstituting) a beat system because they are apprehensive that officers would become too involved with neighborhood residents, and thus perhaps engage in biased behavior. For similar reasons they often oppose assigning officers to the areas in which they reside, and they advocate reasonably frequent changes in assignment.

Social workers may be unhappy with the requirement to process endless paperwork rather than spend time providing client services. But whether happy or unhappy with job routines the fact remains that they serve to limit client demands on the system. The righteous objections of critics that routine procedures detract from primary obligations to serve clients are of little

account, since in an important sense it is not useful for the bureaucracies to be more responsive and to secure more clients.

Second, routines provide a legitimate excuse for not dealing flexibly, since fairness in a limited sense demands equal treatment. Unresponsiveness and inflexibility reinforce common beliefs already present that bureaucracy is part of the problem rather than the solution, and they further reduce clients' claims for service or assertions of need.

When routines lead to predictability they may promote a degree of client confidence. As a public defender lecturing his peers on increasing client trust advised: "It's better to tell a client you will see him in two weeks and then show up, than to reassure him by saying, 'I'll stop by tomorrow,' and never show."[30]

But agency practices do not always lead to predictability. When they lead to delay, confusion, and uncertainty they assign considerable costs to clients. At times routines established to protect clients are distorted to minimize contact or services. For example, to insure responsiveness housing inspectors may be required to make more than one effort to contact complainants. However, inspectors may become adept at telephoning complainants when they are unlikely to be home or fail to keep appointments punctually. In Boston this practice "enhanced the prospects of no one being home when the inspector arrived—a practice which when repeated thrice, enabled cases to be dropped."[31]

The significance of practices that subvert predictability, antagonize or neglect clients, or sow confusion and uncertainty is that they are generally *functional* for the agency. They limit client demands and the number of clients in a context where the agency has no dearth of responsibilities and would not in any way be harmed as an agency if clients became disaffected, passive, or refused to articulate demands. Any reduction in client demand is only absorbed by other clients who come forward, or by a marginal and insignificant increase in the capacity of street-level bureaucrats to be responsive to the clients who continue to press.

It is for this reason that we conclude that stated intentions of street-level bureaucracies to become more client-oriented, to receive more citizen input, and to encourage clients to speak out are often questionable, no matter how sincere the administrators who articulate these fine goals. It is dysfunctional to most street-level bureaucracies to become more responsive. Increases in client demands at one point will only lead to mechanisms to ration services further at another point, assuming sources remain unchanged.

The logical but absurd extension of the relationship between demand and services is exemplified by the apocryphal library that reduced costs by closing down. Yet it is a real problem that increased patronage of libraries, museums, zoos, and other agencies providing free goods increases their uncompensated costs when they succeed in becoming more attractive.

Undoubtedly there are dimensions of bureaucratic practice in which increased responsiveness does not add to workers' tasks. Addressing clients politely rather than rudely or indifferently is an area in which greater responsiveness is not necessarily burdensome to the work load. Furthermore, reorganization may result in increasing the responsive capacity of workers. However, most increases in responsiveness—doing more for clients, or even listening to them more—place additional burdens on street-level bureaucrats, who will subvert such developments in the likely absence of any strong rewards or sanctions for going along with them.

There are times when bureaucratic rationing is not simply implicit; limiting clientele or reducing services is the agency's stated policy. In response to reduced budgets or other developments that make client-worker ratios conspicuously high, agencies will reduce the scope of service in several characteristic ways. In reducing services explicitly they will continue to honor the formal norm of universalistic service patterns.

Street-level bureaucracies may reduce services geographically. They may formally narrow the catchment area from which clients are drawn or reduce the number of neighborhoods served by a program. Alternatively, because reductions in service are unpopular, street-level bureaucracies may prefer to reduce the number of centers, effectively cutting services to some areas without formally changing anyone's eligibility. When the borough of Manhattan, for example, consolidated its municipal court system, eliminating district courts in Harlem, it did not formally change access to the court, but informally it substantially increased the costs of using the court system to Upper Manhattan residents.

Services can be limited in terms of clients' personal characteristics. Formally, agencies can change income eligibility levels. Informally, they may limit service by failing to print posters in Spanish or by placing notices in old-age and nursing homes rather than in public housing in order to attract primarily an elderly population.

Street-level bureaucracies also can formally or informally ration services by refusing to take certain kinds of cases. The decriminalization of drunkenness, for example, formally exonerates policemen from dealing with alcoholics (although public disapproval still places pressure on the police to do something about drunks). Informally departments can limit the clientele if

officers choose to ignore public drunkenness, or they can reduce its place in departmental priorities.

Even when limiting services is not explicitly the function of rationing practices, service limitation often is not an unintended consequence of bureaucratic organization. Street-level bureaucrats and agency managers are often quite aware of the rationing implications of decisions about shorter office hours, consolidation of services, more or fewer intake workers, or the availability of information. Consider, for example, the efforts of the Budget Bureau of New York City in 1969 to decrease welfare expenditures. In a document remarkable for our purposes the Bureau suggested several ways to save close to $100 million.[32] In addition to reducing allowance levels, which would supply the bulk of the savings, the bureau recommended four administrative changes. Each would explicitly ration services in some way. A new intake procedure was proposed that would require applicants to be actively seeking jobs prior to the intake interview. This would force people to accept low-wage work, and, it was hoped, "more aggressive utilization of existing leverage over the employables would . . . have a deterrent effect on applications for welfare."[33] The authors recognized that for this innovation to be effective a substantially greater capacity of public employment agencies would be required, but there was no discussion of the costs of achieving this increase.

More frequent recertifications would be conducted to induce recipients who were on the rolls but no longer eligible because of changed circumstances to initiate case closings. (More than half of all case closings were then initiated by clients.) This reform would reduce the time between changes in clients' circumstances and the next reporting period.

Closing seven outreach centers would save some of the costs of running the centers, but more importantly, "larger savings are anticipated from secondary effects. . . . The most important of these is the opportunity *to build up and maintain the maximum legal backlog* between intake and eligibility increasing average backlog from two weeks to a full month."[34] Among other secondary benefits of center closings, the authors of the recommendations expected that "the relative inconvenience to the client of self-maintenance on emergency grants (for which application is normally made at the center more than once a week) may have some deterrent effect on [those] marginally eligible for welfare."[35]

Finally, stronger management audits would introduce greater uniformity in the system and provide better checks on welfare employees, who are portrayed in the document as more interested in enrolling clients than in controlling welfare costs.

Of equal interest are the strategies considered but not recommended. These included reducing intake hours, drastically closing intake centers, and requiring clients to provide increased documentation of birth, wages, rent payments, and other details of eligibility. While these provisions were rejected because they might result in unmanageable backlogs and infringe on clients' legal right to a response to their application within a month, the memo clearly recognizes that these measures would deter application rates by increasing the costs of applying to clients.

Provisions of this memo have been described at some length not because they are themselves remarkable but because they illustrate awareness at the agency planning level of the implications of rationing to limit client demand. It is naive to accept the rhetoric of public officials that their actions have the incidental effect of limiting or discouraging client demands. Rather, the opposite assumption is more useful analytically and more accurate empirically; namely, that public employees and higher officials are aware of the implications of actions taken that effectively increase or decrease client demand. They may deny such intentions publicly, of course, since their jobs require obeisance to norms of public service. They may not favor such policies personally, and they may regret that funding limitations preclude being able to serve more clients. Nonetheless, it is appropriate to assume that public agencies are responsible for the rationing implications of their actions.

In 1976 New York City introduced administrative controls that were credited with reducing the acceptance rate for new welfare applicants by half and terminating 18,000 cases a month. But this was accomplished because eligibles were being turned away "by very negative administration of work and parent-support rules," and because half of those terminated failed to show up for recertification, to respond to mailed questionnaires, or to verify school attendance. Their ineligibility was strictly a matter of difficulty or reluctance to pay the costs of remaining on the rolls until forced to do so. Meanwhile, according to an administrator, welfare centers are "overcrowded," "noisy," and "dirty." "Some clients wait four to five hours for service and too often are required to make more than one visit to the center to complete their business. In addition, they don't know the names of people who are serving them."[36] In these and other ways *eligible* clients are asked to pay the costs of seeking relief.