

# Waiting, Exchange, and Power: The Distribution of Time in Social Systems<sup>1</sup>

Barry Schwartz  
*University of Chicago*

So far as it limits productive uses of time, waiting generates distinct social and personal costs. The purpose of this paper is to explore the way these costs are distributed throughout a social structure and to identify the principles to which this allocation gives expression. The main proposition of our analysis is that the distribution of waiting time coincides with the distribution of power. This proposition is based on the assumption that an individual's power reflects the scarcity of the goods or skills he possesses; accordingly, the relationship between a server and client may be characterized in terms of organized dependency, for which waiting (under certain conditions) provides an accurate index. However, if delay is related to the client's position in a power network, then he may show deference to a server by an expressed willingness to wait, or a server may confirm or enhance his own status by deliberately causing him to wait. Secondary interactional modes thus come to subserve a relationship originally grounded in a supply-demand structure. The broader implications of this correlation allow us to characterize stratification systems in terms of the apportionment of time as well as the distribution of other kinds of resources.

Delay and congestion are relevant to the analysis of social systems because they undermine the efficiency with which these systems conduct their business. Indeed, one Russian economist (Lieberman 1968–69) recently observed that because of its enormous cost in terms of more productive activities foregone, delay in waiting rooms and queues merit the status of a social problem (pp. 12–16). A gross estimate of the dimensions of this problem is furnished by Orlov, who reports that the Soviet population wastes about 30 billion hours a year waiting during their shopping tours alone. This is the equivalent of a year's work for no less than 15 million men (*New York Times*, May 13, 1969, p. 17). Another study shows that monthly queuing for the payment of rent and utilities wastes at least 20 million man-hours a year in Moscow alone (*New York Times*, June 25, 1972, p. 23). If figures like these were aggregated for the entire service sector of the labor force, social inefficiency occasioned by clients' waiting would stand out even more dramatically.

<sup>1</sup>This paper was supported by grant 1-5690-00-4335 from the Ford Foundation and by the Center for Health Administration Studies, University of Chicago. The writer wishes to acknowledge the very useful comments made on this paper by Peter Blau and Morris Janowitz.

The problem of delay may be more acute in some societies than in others; however, no modern society can claim immunity in this respect. Every social system must "decide" not only how much different members are to be given from a collective supply of goods and services; there must also be a decision as to the priority in which their needs are to be satisfied. Queuing for resources is in this sense a fundamental process of social organization, regardless of the specific level of its affluence. Indeed, though the amount of waiting time per unit consumption may be minimal in the richer, consumer-oriented societies, a higher volume of consumption leaves open the possibility that more time is lost in waiting under conditions of affluence than under conditions of scarcity.

On the other hand, it may be said that the social costs of waiting, no matter where they are incurred or what their absolute level may be, merely derive from the summation over an entire population of rather negligible individual losses. But this does not seem to be the case. As one American commentator (Bradford 1971) puts it: "None of us would think of throwing away the nickels and quarters and dimes that accumulate in our pockets. But almost all of us do throw away the small-change time—five minutes here, a quarter hour there—that accumulates in any ordinary day. I figure I probably threw away a full working day in the dentist's office this past year, flicking sightlessly through old magazines" (p. 82). Even in the more opulent of modern societies, then, waiting time creates significant deficits for the individual as well as the system. At issue, however, is (1) the way such cost is distributed throughout a social structure and (2) the principles which govern this distribution. These questions are the subject of the present inquiry.

We begin with the assumption that delay is immediately caused by the relations of supply and demand: when the number of arrivals in some time unit is less than the number an organization can accommodate, waiting time will be relatively brief; but if the arrival rate exceeds the service rate, a "bottleneck" is created and a longer waiting period results. Delay is in this sense occasioned by limitations of access to goods and services. However, this model does not explain socially patterned variations in waiting time. We must therefore explore the institutional constraints which sustain observable levels of scarcity and which organize the priorities granted to different groups of clients. These constraints are shown to be the expressions of existing power relations.

As we proceed, however, we discover that a purely structural model is tied to the very assumptions it seeks to extend, inasmuch as it takes objective scarcity as its point of departure. We then demonstrate how scarcities grounded in structured power relations may be deliberately magnified by the very people engaged in these relations. And so delay is found to be

partially independent of supply and demand, in whose relation this variable was originally thought to find its exclusive source.

The above argument is informed by the assumption that time is a generalized resource which may be expended productively or wastefully with respect to the acquisition of other, more concrete advantages. As such, our analysis will not only help clarify the way "productive time" and "idle time" are allocated in a social system; it will also show how this distribution affirms and even reinforces that system's power arrangements.

#### WAITING, SCARCITY, AND POWER

When economic exchange involves relatively massive demand for specialized services, disturbances at the level of synchronization of supply and demand results in congestion. Waiting thus finds its organizational precondition in the scarcities occasioned by an advanced division of labor.

Waiting is related to scarcity in two respects. When the demand for a good or service exceeds its absolute supply, people may queue up before it is actually made available in order to ensure they will be accommodated. Others will wait with no guarantee of being served.<sup>2</sup> This latter condition is most widespread and conspicuous in the Soviet Union, where, according to Orlov (*New York Times*, May 13, 1969, p. 17), an average shopper must often wait in long lines at three to five stores in order to buy the item he wants. The same kind of problem arises in more consumer-oriented countries during periods of peak demand, for example, intercity transportation during holidays, theaters and restaurants on weekends, bargain days in department stores, important sporting events, etc.

Regardless of the scarcity of a good, however, organizations tend to minimize the employment of servers; in doing so they minimize labor costs and enhance profits. Similarly, those who sell their skills tend to create queues so as to minimize their idle time. A second condition of waiting, then, is the ratio of supply of servers to demand for the services which they are prepared to offer. The fewer the servers in relation to the number of clients they must accommodate, the greater will be the average client's waiting time. Moreover, the greater the scarcity of a service, and the more inelastic the demand for that service, the less is a server compelled to reduce the waiting time of clients. Urgency of need thus minimizes the probability of "balking," that is, refusing to enter the queue, and "re-neging," or abandoning the queue after having entered.

Waiting is patterned by the distribution of power in a social system. This assertion hinges on the assumption that power is directly associated with an

<sup>2</sup> Others wait with no possibility of being served. For a very convincing explanation of this practice, see Mann and Taylor (1969).

individual's scarcity as a social resource and, thereby, with his value as a member of a social unit (see Blau 1964, p. 118). Accordingly, the person who desires a valued service generally cannot gain immediate access to its dispenser but must instead wait until others are accommodated.

However, it would probably be more precise to say that the capacity to make others wait is a property of roles and not their incumbents. The petty bureaucrat or cashier, for example, may himself possess little that is of value to others; however, he governs access to resources which are. As a result, he is able to keep people of great substance waiting for as long as he sees fit. Of course, when a server's power derives solely from his access to his employer's resources, that power can only be exercised over clients. In the absence of valued personal qualities his position in the organization itself will be a lowly one.

#### Waiting and Exchange

After a certain point, waiting becomes a source of irritation not only because it may in itself be wearisome, boring, and annoying, but also because it increases the investment a person must make in order to obtain a service, thereby increasing its cost and decreasing the profit to be derived from it. This loss to the waiter is related to the fact that time is a finite resource; its use in any particular way implies the renunciation of other rewards and opportunities. Put differently, in waiting, usable time becomes a resource that is typically nonusable. This transformation is mediated by the power relation between server and client: time, whose use is ordinarily governed only by the client—that is to say, expended for the sake of a benefit that he alone desires—is transformed during the waiting period into a resource that is governed only by the one whom the client attends.

However, the formal interactional properties of waiting are independent of vicissitudes in its personal cost. To be able to make a person wait is, above all, to possess the capacity to modify his conduct in a manner congruent with one's own interests. To be delayed is in this light to be dependent upon the disposition of the one whom one is waiting for. The latter, in turn, and by virtue of this dependency, finds himself further confirmed in his position of power. Looked at in a different way, it may be said that while having to wait may under certain conditions be negative and harmful to the interests of particular individuals, it often furthers the interests of those who keep them waiting. Waiting is therefore a negative condition only when we confine ourselves to the standpoint of the person who is delayed.

The one-sidedness of this statement may be balanced by two considerations which specify its applicability. First, the disadvantage of the waiter may be a detriment to the server as well. For, the benefits of waiting (such as respite from previous interaction and an opportunity to prepare for subsequent involvement) presumably shrink as the costs of activities foregone



increase; their intersection constitutes grounds for the waiter to renege from the waiting channel without being served. Correspondingly, the benefits a server receives by keeping a particular client waiting may initially be of greater value to him than time spent servicing that client; however, the declining value to the server of keeping that client waiting further and the rising cost entailed in delaying service also reach a point of intersection. This convergence constitutes grounds for offering service. The relationship between the first intersection and the second is crucial: if he is to stay in business, the server had better decide to serve before the client decides to leave. Although waiting represents an unfavorable exchange position for a client, the very principles which make it so subject the server to pressures which mitigate the extent of his delay. However, these pressures vary in terms of the scarcity and value of service, which suppress the probability of renegeing. This leads back to our initial point that his scarcity enhances the exchange position of a valued server who, while needing clients, needs no particular client; he can therefore take his time about serving any one of them.

#### The Stationary Server

The highly advantageous position of the server is intensified when viewed in terms of a stationary server/mobile client model. This arrangement not only affirms the power of the former (for the latter must expend resources to come to obtain the service he offers) but also works to his advantage in other ways. First, the stationary server has at hand sufficient opportunities for alternative involvement to offset the loss to which the tardiness of a client would otherwise subject him; second, he has the power to schedule and thereby control the sequence and pace of his activity.

The latter advantage is most conspicuously instanced by the widespread practice (particularly common among physicians with a large following) of overscheduling—setting up two or more appointments at very narrow intervals in order to ensure that possible delays on the part of clients, or a run of quick services, will not leave the server with idle time. Yet, even when confronted with an empty waiting room the server has at his disposal enough alternative involvement materials to minimize his loss. These may involve “secondary queues,” for example, paper work, checking of supplies, necessary calls to colleagues and clients, etc. On the other hand, the waiter is usually unable to transport enough supplies to keep himself maximally occupied, at least from a productive standpoint. Even the client who can bring his business to the waiting room in a briefcase may find himself unable to work comfortably in this strange and perhaps distracting setting. He is then cut off from queues which await *his* service.

Thus, by making the client wait the server may often impose a loss

without suffering one himself. On the other hand, by renegeing, the client fails to impose a substantial loss upon a server, who may continue to operate productively; yet, the client subjects himself to loss in terms of time already invested in waiting. Moreover, even when he is forced into idleness the server may charge his clients for the time lost. This is especially true among psychiatrists, whose rigid 30- or 50-minute treatment sessions prevent them from overscheduling. On the other hand, when the popular server is delayed and forces his client into excessive idleness he does not consider it necessary to make compensation.<sup>3</sup>

Just as the mobile client may find himself at the mercy of a stationary server, the mobile server can be used hard by a stationary client. This server may have to wait until the client is ready to be seen and may get caught up in other peoples' queues, for instance, traffic jams that occur between visits to clients. Hence the increasing reluctance on the part of professionals to leave their offices for fieldwork and house calls. This practice redounds to their moral as well as material benefit, on the basis of the "if you want an audience, you come to *me*" principle (see Spencer 1886, p. 105). This seems to be confirmed by the fact that professionals generally do go to clients whose status exceeds their own. The doctor or lawyer who refuses to conduct business in the homes of the ordinary will more often than not rush to the ailing or wailing Mr. Big, taking care not to keep him waiting.

On the other hand, there are some servers who by the very nature of their work are forced to be mobile. These include insurance agents, door-to-door salesmen, delivery men, messengers, repair men, subcontractors, and the like. Of these, it is perhaps the latter who embody the clearest exception to the tendency for server mobility to be a disadvantage (Glaser 1972, pp. 90–107). Because the popular building subcontractor can take on more work than he can finish directly, and then keep the customer hooked by beginning a job he will only finish at his own convenience, this type of server can make a virtue out of the absolute necessity of his mobility (but of course only while the market is in his favor).

<sup>3</sup> There are exceptions, however. Thus, "on some airlines, when there is a flight delay, the passengers . . . may get a meal, a long-distance call or even an overnight hotel room, all free of charge." But "with few exceptions, the availability of complimentary services is left to the initiative of those passengers who are knowledgeable enough to request them. . . . Passengers who are bashful or unaware of their rights may get nothing. . . . The Civil Aeronautics Board said four airlines indicate they will take the initiative in informing the passenger of such services. All other carriers will provide information as to services only upon the passenger's request" (*Chicago Daily News*, September 2–3, 1972, p. 27). The same is true with regard to the practice of "bumping." On some flights, airlines will sell more tickets than there are seats, expecting a certain percentage of those with reservations not to show up. If all do appear, the last ones to have made reservations are rejected and made to wait for the next flight. The CAB regulations entitle those so delayed to immediate compensation in the form of a fine equal to the price of the ticket (within a \$25–\$200 limit). Until recently, this right was honored only at the passenger's request (*Chicago Daily News*, October 10, 1972, p. 28).

### Stratification of Waiting

Typical relationships obtain between the individual's position within a social system and the extent to which he waits for and is waited for by other members of the system. In general, the more powerful and important a person is, the more others' access to him must be regulated. Thus, the least powerful may almost always be approached at will; the most powerful are seen only "by appointment." Moreover, because of heavy demands on their time, important people are most likely to violate the terms of appointments and keep their clients waiting. It is also true that the powerful tend not to ask for appointments with their own subordinates; rather, the lowly are summoned—which is grounds for them to cancel their own arrangements so as not to "keep the boss waiting."

The lowly must not only wait for their appointments with superiors; they may also be called upon to wait during the appointment itself. This may be confirmed in innumerable ways. For one, consider everyday life in bureaucracies. When, in their offices, superordinates find themselves in the company of a subordinate, they may interrupt the business at hand to, say, take a phone call, causing the inferior to wait until the conversation is finished. Such interruption may be extremely discomforting for the latter, who may wish not to be privy to the content of the conversation but, having no materials with which to express alternative involvement, must wait in this exposed state until his superior is ready to reengage him. The event becomes doubly disturbing when the superior is unable to recover from the distraction, loses his train of thought, and is unable to properly devote himself to the moment's business. Of course, the subordinate is demeaned not only by the objective features of this scene but also by his realization that for more important clients the superior would have placed an embargo on all incoming calls or visitors. He would have made others wait. The assumption that the client correctly makes is that his own worth is not sufficient to permit the superior to renounce other engagements; being unworthy of full engagement, he is seen, so to speak, between the superior's other appointments. In this way, the client is compelled to bear witness to the mortification of his own worthiness for proper social interaction.

While the derogatory implications for self are clear when the person must repeatedly step aside and wait until the superordinate decides that the granting of his time will not be excessively costly, debasement of self may be attenuated by the client's own consideration that his superior is, after all, in a position of responsibility and assailed by demands over which he may not exercise as much control as he would like. But even this comforting account may be unavailable when the server himself initiates the interruption. It is possible for him to make a call, for example, or to continue his work after the client enters, perhaps with the announcement that he will "be through in a minute."

It is especially mortifying when the superior initiates a wait when an engagement is in progress. Thus, a subordinate, while strolling along a corridor in conversation with his superior may find himself utterly alone when the latter encounters a colleague and breaks off the ongoing relationship in his favor. The subordinate (who may not do the same when encountering one of his peers) is compelled to defer by standing aside and waiting until the unanticipated conversation is finished. Nothing less is expected by his superior, who, finding himself gaining less from the engagement than his inferior, assumes the right to delay or interrupt it at will if more profitable opportunities should arise.

*The immunity of the privileged.*—The relationship between rank and accessibility implies that waiting is a process which mediates interchanges between those who stand on different sides of a social boundary. These divisions and the rules of access which correspond to them are found in organizations which are themselves bounded with respect to the outside world. This fact raises the problem of access when outsiders or clients (as well as insiders, that is, employees or co-workers) seek contact with persons situated at different points in a service hierarchy:

Low down on the scale are the men you can walk right up to. They are usually behind a counter waiting to serve you on the main floor, or at least on the lower floors. As you go up the bureaucracy you find people on the higher floors and in offices: first bull pens, then private offices, then private offices with secretaries—increasing with each step the inaccessibility and therefore the necessity for appointments and the opportunity to keep people waiting. Recently, for example, I had an experience with a credit card company. First, I went to the first floor where I gave my complaint to the girl at the desk. She couldn't help me and sent me to the eighth floor to talk to someone in a bullpen. He came out, after a suitable waiting time, to discuss my problem in the reception room. I thought that if I were to straighten this matter out I was going to have to find a vice-president in charge of something, who would keep me waiting the rest of the day. I didn't have time to wait so I took my chances with said clerk, who, of course, didn't come through. I'm still waiting for the time when I have an afternoon to waste to go back and find that vice-president to get my account straightened out.<sup>4</sup>

The above statement suggests that delaying a typical client may be a prerogative of important servers. However, we must also recognize that powerful clients are relatively immune from waiting. This remark accords with Tawney's (1931) emphasis on the asymmetry of power relations. "Power," he writes, "may be defined as the capacity of an individual, or group of individuals, to modify the conduct of other individuals or groups in the manner which he desires, *and to prevent his own conduct being modified in the manner in which he does not*" (p. 229; emphasis added).

<sup>4</sup> Personal communication from Florence Levinsohn.



## Waiting, Exchange, and Power

The relative immunity from waiting which the powerful enjoy is guaranteed because they have the resources to refuse to wait; that is, because they can often afford to go elsewhere for faster service or cause others, such as servants or employees, to wait in their places. Thus, while the relationship between privilege and the necessity of waiting cannot be generalized in any deterministic way, there appears nevertheless to be a relationship between the two, with the least-privileged clients compelled to do the most waiting. This general statement is consistent with Mann's (1969) more specific observations regarding the stratification of waiting in lined queues:

The relationship between cultural equality and public orderliness is attenuated in the area of queueing because waiting in line is not a habit of all social classes in Western society. It is reasonable to suppose that if Mrs. Gottrocks joined a theater or a football line in the United States, Australia, or England, she would not be treated differently than anyone else, but it would be a rare event for someone of Mrs. Gottrock's status to use a line. Ordinarily, in both class-conscious and relatively class-free societies, the privileged class circumvent the line altogether and get their tickets through agents or other contacts.<sup>5</sup> Our point, then, is that queueing is confined largely to the less-privileged groups in society. [P. 353]

The privileged also wait less because they are least likely to tolerate its costs; they are more inclined to renege from as well as balk at entering congested waiting channels. On the other hand, the less advantaged may wait longer not only because of their lack of resources but also because their willingness to wait exceeds the readiness of those in higher strata. While they might have something else to do besides sitting and waiting, they might not have anything better to do. As a result, the least advantaged may pay less in profitable alternatives foregone and therefore suffer less than even those whose objective wait is shorter.

This relationship may be informed by another consideration, for which health-care delivery systems provide an example. Because of their scarcity, those who are able to pay for medical services are often forced to wait well beyond the time a server agreed to provide them. Yet there is some limit to the server's inconsiderateness, for, in principle at least, the client may decide that he has waited long enough and go elsewhere. On the other hand, those who are unable to pay for medical care may spend the better part of the day in outpatient waiting rooms, for consideration of the value of clients' time is far less imperative when these clients cannot take their business to someone else. In Britain's government-run maternity hospitals, for example, "a major complaint was that women dependent on the health service are treated offhandedly in hospitals and frequently have to wait

<sup>5</sup> Other "contacts" include the radio, over which Saturday and Sunday morning waiting times at many metropolitan golf courses are broadcast. This service, which saves many players many long delays, is performed almost exclusively for the middle and upper-middle classes.



more than an hour for checkups at antenatal clinics. Women who paid up to \$700 for private treatment were dealt with speedily and efficiently" (*Chicago Tribune*, June 12, 1971, p. 10). Thus, while long, agonizing waiting periods may be avoided only if one is willing to settle for more expensive service, the poor may avoid waiting only if they are willing to settle for no service at all. (The frequency with which they do select this option is, of course, unknown—as is the consequence of the selection.)

The above principle may be further illustrated in other, altogether different connections. It is noticeable, for one example, that in the "best" of urban department stores a customer is met by a salesperson as soon as he enters; the customer makes a selection under his guidance and makes payment to him. In establishments which are a grade below the best, customers may have difficulty finding someone to serve them during busy periods but, when they do, are accompanied by him, that is, "waited on," until the transaction is consummated by payment. The lowest-grade stores, however, provide few servers; as a result, customers must for the most part wait on themselves, then line up behind others at a cashier counter in order to make payment.

The above patterns are to be observed within as well as among organizations. In the typical department store, customers surveying high-priced goods like furniture and appliances will typically be approached immediately by a salesperson. Those in the process of selecting a handkerchief or pair of socks will not be so quickly attended and, when they finally are, will be dealt with more quickly. Likewise, clients who show interest in very expensive jewelry will be served at once and at length; those who are fascinated with costume jewelry will wait.

In general, it may be said that establishments which cater to a relatively wealthy clientele must serve them quickly (if the clients desire) not only because of the objective or assumed value of clients' time but also because they have the means to take their business elsewhere if it is not respected. Commercial places which service the less wealthy are less constrained in this respect because they tend to deal with a larger and/or less independent clientele. Within organizations, clients who promise to bring the most profit to a server enjoy a competitive advantage; they wait the least, to the disadvantage of their lesser endowed brethren who can find no one to honor the value of their time.<sup>6</sup>

<sup>6</sup> Even when circumstances make it necessary for the resourceful to wait, they suffer less than their inferiors. As a general rule, the wealthier the clientele, the more adequate the waiting accommodations. Thus, persons who can afford bail can await their trial (or, far more frequently, attorneys' bargaining on their behalf) in the free community. The poor must wait in jail. The same is true of facilities. In airports, for example, those who can afford it may simultaneously avoid contamination by the masses and engross themselves in a variety of activities, including fabulous eating and drinking, in "VIP lounges." The term "lounge" instead of the vulgar "waiting area" or "gate" is

Waiting and the Monopolization of Services

The above rule, however, rests on the assumption that faster alternative services are available to those who want and can pay for them. In fact, the availability of such alternatives is itself variable. Waiting is therefore affected not only by clients' resources and consequent ability to go elsewhere for service but also by the opportunity to do so.

It follows that establishments with many competitors are most likely to be concerned about the amount of time they keep clients waiting. Chicago Loop banks are among such organizations. In the words of one banking consultant, "The industry is too competitive to allow a dozen people waiting in line when they could just as easily take their business across the street where there is a teller at every window, a customer at every teller and waiting time is less than one minute" (*Chicago Tribune*, September 28, 1971, p. 7). However, organizations with few or no competitors are less obliged to reduce the waiting time of clients. (This condition makes waiting a national pastime in the Soviet Union, where most services are rendered by government-run establishments that are not subject to market forces.)

The enormous amounts of waiting time expended in dealings with public people-serving bureaucracies is directly related to monopolization of the various services which they offer or impose. Monopolization accords governmental units the power to maximize their efficiency of operation by minimizing service costs and, in so doing, maximizing client waiting. This "optimum solution" is exemplified by bureaus which distribute welfare benefits to long lines of disadvantaged people:

The number of Medicaid and public assistance applicants and recipients has become so great that [New York's] Department of Social Services is literally shutting its doors in their faces.

Many of the 45 social service centers close their doors early—12, 1 or 2 o'clock—rather than admit persons the workers realistically know cannot be seen that particular day.

The Medicaid office advises applicants to line up outside the doors before dawn. "You'd better get down here around 6:30 or 7 o'clock," said a person answering the telephone at the Medicaid office. . . . "We can only see 200 persons a day. If you want to be in the first 200 you better get here then—with your application filled out." The Medicaid office does not open until 8:30 A.M. . . .

Last week the department announced it had saved \$39 million by employing fewer case workers. [*New York Times*, November 21, 1971, p. 58]

However, the relatively wealthy as well as the poor are put to incon-

---

also applied to facilities set aside for those who travel a specified number of miles with (and pay a substantial sum of money to) a particular airline. In this as in many other settings, waiting locales for the poor and less rich lack the elaborate involvement supplies, pleasant decor, and other physical and psychological comforts that diminish the pain of waiting among those who are better off.

venience by having to wait in person for licenses, permits, visas, tickets, information and the like. Dealings with government-sponsored transportation facilities can also be cited as an example:

Before Amtrak took over, I would have had to call the Illinois Central to go to Miami. If I wanted to go to New York, I'd call the Penn Central. To go west, the Santa Fe. But now, under the streamlined, tax-supported Amtrak, one number, one central office, makes the reservations. They have computers and other modern devices the old system didn't have.

At 10 minutes after noon, I dialed the new Amtrak reservation number. The line was busy, so I hung up and waited a few minutes and dialed again. It was still busy. Five minutes later, I tried again. It was busy. By 1 o'clock I had tried 10 times, and had heard only busy signals.

Enough was enough. I phoned the Amtrak executive office, to ask what was wrong with their reservation number. A woman there put me on hold. I was on hold for seven minutes. Then when she finally took me off hold, she switched me to somebody's office, and a secretary laughed and said: "Oh, yes, our lines are very busy."

At 2 P.M. it finally happened. Instead of getting a busy signal, it rang. It actually rang. . . . It rang. And it rang. And it rang. For eight minutes it rang. . . . So I hung up, got another cup of coffee and tried again. That was a mistake, because I heard another busy signal.

Then at 2:47 it happened. It rang. And somebody answered. I listened closely to make sure it wasn't a recorded message. No, it was really somebody alive. After that it was easy. In about eight or nine minutes the reservations were made.

The clock said 3 P.M. So I have to congratulate Amtrak. It took me only two hours and 50 minutes to complete a telephone call and make reservations. It would have probably taken me at least 10 minutes more than that to take a cab to O'Hare, board a plane, fly to Miami, and get off the plane. [*Chicago Daily News*, June 9, 1972, p. 3]

This instance is an especially informative one, for it demonstrates that the amount of time clients of an organization are called upon to wait is in large measure determined by the broader competitive structure in which that organization is situated. Longitudinal and cross-sectional means are brought to bear in this assessment. By reference to the temporal barrier to access to rail service after centralization and monopolization, relative ease of access before the transformation is implicitly affirmed. And after documenting the lengthy waiting time required in a noncompetitive service market, we find explicit reference to the ready availability of service offered in highly competitive ones (airlines, in this case). In this double sense, the institutional grounding of waiting time is a conclusion warranted by the facts.

We now turn to public services which by their very nature admit of no alternatives and which at the same time are so organized as to constitute the most radical instance of the principle we are now discussing.

*A day in court.*—Discrepancy between demand for and supply of "authori-

tative judgment” is perhaps the most notorious source of waiting for both rich and poor. In fact, those who look forward to their “day in court,” whether civil, criminal, or juvenile, very often find themselves spending their day in the courthouse corridor (many courts do not provide waiting rooms). In some courts, in fact, all parties whose cases are scheduled to be heard on a particular day are instructed to be present at its beginning when the judge arrives.<sup>7</sup> This is a most pronounced manifestation of what we earlier referred to as “overscheduling,” which in this case ensures that the judge (whose bench is separated from his office or working area) will not be left with idle time that cannot be put to productive use—a consideration which may help us understand the seemingly irrational practice of assembling together at the beginning of the day those who are to be served during its course. While this tactic guarantees that the judge’s valuable time will not be wasted, it also ensures that most parties will be kept waiting for a substantial period of time; some, all day long. Indeed, because they have no means to retaliate against the judge’s own tardiness or excessive lunch breaks, some individuals may not be served at all and must return on the next day to wait further. Clients’ attorneys, incidentally, keep them company during much of this time—a service for which the former pay dearly.

All of this is not to say that the organization of justice profits. It must, on the contrary, pay a very high price for support of its *prima donnas*. As one juvenile-court officer puts it: “[W]aiting to be called into court . . . is the most serious problem. Just from an internal point of view this means that a probation counselor usually accomplishes nothing in the hour or more he often has to wait to get his case into court. Usually during this waiting period he sees no people, does no counselling, can’t do dictation or other ‘desk-work’—his wait is complete, unproductive waste. These same problems apply to other professional people: caseworkers from the Department of Social Services, school principals, lawyers, etc.” (Fairfax County [Virginia] Juvenile and Domestic Relations Court, Memorandum, 1971, p. 1). While attorneys<sup>8</sup> and other professionals are fortunate enough to claim a

<sup>7</sup> A functional equivalent is found in the Soviet Union. “Aleksandr Y. Kabalkin and Vadim M. Khinchuk . . . describe what they termed ‘classic cases’ in everyday life in the Soviet Union, in which customers wait for the television repairman or for a messenger delivering a train or plane ticket that had been ordered by phone. To the question ‘About what time can I expect you?’ the stereotyped reply is, ‘It can be any time during the day.’ And people have to excuse themselves from work and wait—there is no other way out” (*New York Times*, November 7, 1971, p. 5).

<sup>8</sup> It may not be assumed that all lawyers earn while they wait. For example, the *New York Times* (August 25, 1971, p. 24) recently reported: “A lawyer who specializes in prosecuting landlords’ claims against tenants asked permission in Bronx Supreme Court yesterday to bring his cases there rather than in Civil Court because . . . he spent much time ‘just sitting and waiting.’ And consequently, he said, he was suffering ‘financial loss’ and felt he could not continue working in Civil Court.”



fee for doing nothing in a professional way, others are often denied this luxury. Authorities who are mindful of civil security, for example, wisely find it more expedient to dismiss cases (particularly such misdemeanants as traffic violators) for lack of witnesses and evidence than to tie up a large sector of the police force for the better part of the day in a crowded corridor. In this particular sense, the police are too important—their time too valuable—to be kept waiting. On the other hand, it may be claimed that by tying up defendants all day long in these same corridors justice may be served—provided, of course, that the defendants are in fact guilty as charged. However, the situation is quite different in felony cases, where casual dismissals are less probable. Under these circumstances police wait as long as defendants. In the Chicago Gun Court, for example, “40 or 45 police are waiting to testify at 9:30 A.M., when court begins. Cases are not scheduled for specific times, so most of them wait and wait. One recent day 31 were still waiting around at 1 P.M. The next day 20 were there at 1 P.M. And 23 the following day.” The same conditions prevail at the Narcotics Court where police waiting time “translates on an annual basis to 13,000 police days lost and \$700,000 in expenses” (*Chicago Daily News*, August 21, 1973, p. 14).

Two observations emerge from and transcend the particular content of what has just been said. First, the assertion that clients may pay a high price, in terms of time, in their dealings with public bureaucracies means that a societal cost, expressed in terms of aggregate client time diverted from more productive activities, must be written into the usually implicit but sometimes explicit “optimum solution formulae” by which particular “public service” organizations maximize their own efficiency. Because of this factor, the real cost of governmental services is not to be obviated by budgetary considerations alone.

Second, minimization of a powerful server’s idle time may subtract from the productivity of the organization as well as its clients. This observation, which is merely grotesquely evident in court settings, reflects the general principle that increments in efficiency in one part of a social organization often entail malfunction in other sectors. Accordingly, just as high concentration of power in an organization may lend itself to societal inefficiency, indexed by more productive client-time foregone, so concentration of power and honor in an elevated server may render organizations ineffective by maximizing idle time of subordinated servers. The more general import of this statement is that it amends the overly simplistic scarcity theory of waiting, which fixates our attention upon server shortage as a condition of client delay. The present statement shows that the organization of services, as well as their volume, provides occasion for waiting.

An additional point is that some persons and groups are relatively exempt from waiting. If we turn our attention once more to the courtroom,



we find that the powerful are most likely to enjoy such advantage. In making up the docket, for example, resources are taken into account. Defendants who are represented by an attorney are very often scheduled before those who are not (in Chicago traffic courts, at least). And cases involving important and powerful contestants, witnesses, and/or lawyers may be scheduled at their convenience and not be delayed for long periods of time. Similarly, attorneys who enjoy favor with the court clerk are also able to avoid long waits because they are allowed to schedule their case early.<sup>9</sup> Thus, while waiting time may be maximized by persons or in organizations which enjoy full or near monopoly on the services they offer, the relationship between the power and waiting time of their clients is probably attenuated rather than negated. For, while the powerful may lack the opportunity to take their business elsewhere, they nevertheless possess the resources to ensure that their needs will be accommodated before the needs of those with fewer means.

*The resource-availability theory.*—In summary, the relationship between servers' and clients' power in relation to waiting is asymmetrical. On the one hand, servers' holding power is contingent on clients' inability to frequent more distant and/or expensive servers; on the other, client autonomy requires the presence of alternative services. Despite their covariation, though, resources and alternatives seem to affect waiting time independently of one another. The resourceful wait less within both monopolistic and competitive organizations; regardless of clients' resources, however, waiting time tends to be longer in monopolistic settings. This "resource-availability theory" of waiting may also help explain the varying "optimum solutions" adopted by diverse organizations seeking the most profitable balance between losses due to keeping clients waiting and the expense of additional servers. While the theory predictably suggests that the balance arrived at reflects the relative power of organizations or individual servers, it also holds such resolutions to be "zero sum" in nature. This is most evident in monopolistic, "public service" bureaucracies or among charismatic officials who maximize the efficiency of their operation at great organizational and social cost, expressed in terms of productive time lost through waiting. The optimization of unit interests is thus often brought about at the expense of system interests.

However, the resource-availability theory must be qualified by recognizing existing limitations on monopolies' capacity to restrict service. For, despite their freedom from competition, monopolies must get their work

<sup>9</sup> This is to say that, as a scarce commodity, time or priority of service routinely becomes the object of struggle. Recognizing this, a court intake officer writes in a memo to his supervisor: "Intake counselors should assume more control over the setting of cases on the docket, with a proportionate decrease in the control now exercised by clerks" (Fairfax County [Virginia] Juvenile and Domestic Relations Court, Memorandum, 1971, p. 1).

done, lest their managing personnel be subjected to pressures from those who have the power to exert them. While monopolistic enterprises may not be responsive to their clients' evaluations, they are often subject to officials, such as elected representatives, who are not free to ignore these clients. If this be so, then the greater the number of people who are dependent upon a monopolized service, the more effective will be the pressure for its public regulation. Thus, while the telephone company is a monopoly, it does not enjoy unlimited freedom to reduce service and delay callers. The same can be said of the postal service (which is in fact sometimes overstaffed because it is a nest for political patronage) and other public utilities. Waiting time, then, is affected by political as well as economic constraints.

#### SOCIAL PSYCHOLOGICAL ASPECTS OF DELAY

On the psychological level, what has been said may be recapitulated in the following terms. The person who is delayed is not merely in a condition of objective dependence and subordination; because his only duty is to attend the call of a server, the waiter feels dependent and subordinate. To be kept waiting—especially to be kept waiting an unusually long while—is to be the subject of an assertion that one's own time (and, therefore, one's social worth) is less valuable than the time and worth of the one who imposes the wait. This is why objections to being kept waiting sometimes take the form, "Why should I wait for him? My time is as valuable as his!" The actually inferior feeling that often gives rise to such a protest is especially common in such places as crowded waiting rooms, wherein each client, confronting multiple reflections of himself, is more pointedly made aware of his suppliant status and of how utterly insignificant he is as compared with the person for whom he waits. Of course, waiting does not create the sense of subordination but only accentuates an initial inferiority, which is often presupposed by the fact that one is waiting in the first place. It needs to be said that this same sentiment has its parallel on the other side of the relationship, for the server calls out in himself the responses that he elicits in the ones he keeps waiting, which enables him not only to be conscious of his own power—to see himself from the point of view of his clients—but also to feel within himself the independent power that he extracts from those who wait for him.

#### Waiting as a Determinant of the Value of Service

The above statement is one elaboration on a theme to which we have tried to adhere throughout this report, namely, that waiting presupposes and occurs within an established context of power relations and is to be understood in terms of these relations. Power, to repeat, entails among other

things the capacity to provide scarce services which people must wait to receive. The significance of the service for the individual and the social power of the dispenser therefore hinges on its desirability.

“The only relevant question apart from the direct enjoyment of things for their qualities,” wrote Simmel (1971), “is the question of the way to them. As soon as this way is a long and difficult one, involving sacrifice in patience, disappointment, toil, inconvenience, feats of self-denial, etc., we call the object scarce. One can express this directly: things are not difficult to obtain because they are scarce, but they are scarce because they are difficult to obtain” (p. 68). Accordingly, if we regard waiting for a scarce service as an investment or sacrifice in return for a gain, we may measure part of the value of the gain by assessing the degree of sacrifice occasioned on its behalf. In Simmel’s (1970) words, “Valuation arises from the fact that something must be paid for things: the patience of waiting . . . the renunciation of things otherwise desirable” (p. 23).

The subjective value of the gain is therefore given not only by the objective value of the service but also by the amount of time invested in its attainment. This being the case, one may wait for another not only because he is a source of value; the other’s service becomes valuable (and he becomes powerful) precisely because he is waited for. While analytically distinct, the two parts of this phenomenon are empirically inseparable. What is more important is that they are functionally inseparable; this is to say that waiting subserves the distribution of power that it presupposes. It does so in two interrelated senses: a common willingness to wait for a service sustains its objective scarcity which, in turn, transforms itself (as we have seen) into a subjective value. This principle is particularly clear in its negative aspect: in the observation that services to which we have immediate access—which we can acquire without waiting—are of relatively little value to us. It is known, for example, that in seeking professional help from a person with whom he is unacquainted, the client does not always rejoice at being granted an immediate appointment, nor at finding an empty waiting room when he arrives for this appointment. Such ease of access may speak unfavorably of the server’s scarcity as a social or economic resource; it may disconfirm the worth of his service. In contrast, those who confront obstacles to service tend to have more confidence in its value, once it is acquired.

The above principle holds within but not outside of specific time limits. Beyond their upper boundary, a desired and valued service may be considered unattainable, and an otherwise willing client might just give up and renege. Those who choose to wait it out beyond that limit may, in so doing, find their estimation of the service to be actually lowered. This is because persons tend not only to place a higher value on services for which they must wait; they also demand more in proportion as they wait. After a

certain point, the latter tendency may outweigh the former, raising expectations to such a level as to render their satisfaction impossible. The reward then cannot possibly be worth waiting for, let alone enhanced by waiting for it.

It follows that if services acquirable without waiting are of little value to us, those who wait to service us may be attributed a negative value; these servers become our subordinates. Simmel put this more generally by saying: "We perceive the specific value of something obtained without difficulty as a gift of fortune only on the grounds of the significance which things have for us that are hard to come by and measured by sacrifice. It is the same value, but with the negative sign" (1971, p. 54). The original meaning of the term "waiter" accords with this formula. The waiter is, in this earlier sense, one who stands by, alert to the call and ready to respond to the demand of a superordinate. What the waiter waits for, then, is a command; he is, as the French expression makes clear, an *attendant*: one who caters to the whims of the ascendant. This earlier, courtly reference to waiting as a form of subordination is found even today (as a "survival," so to speak) in southeastern parts of the United States where we observe the very common substitution of "waiting *on*" (someone or other) in place of "waiting *for*." The mere transformation of the linguistic meaning of waiting, from a readiness to serve to a readiness to be served, has therefore not fully negated its essential sociological property: to wait on others and to be kept waiting exhibit the common element of subordination.

We have digressed in order to demonstrate the inverse case of a principle to which we now return, namely, that waiting is not simply a barrier to service but is rather the very condition of its subjective value. This idea must be addressed in further detail because it appears to contradict our earlier assertion that waiting is inimical to profit in social exchange. It now seems that the reduction of waiting time would not necessarily increase profit for a client (in an exchange with a server) because the value of that which is attended is itself dependent, at least in part, upon the very length of attendance. But from this an absurd hypothesis is deduced: that persons faced with the alternative of, say, a long and a short queue will join the longer one in order to enhance the value of what they will receive at the end of it (much as an individual might extend his feet from under a blanket on a cold night in order to enjoy the warmth that its withdrawal will provide). Absurd as it appears, there is some truth in this; but only in the following, limited sense: that those who wait the longest tend to value what they receive the most. But this only means that the subjective value of the service, that is, its value for the waiter, is positively modified in the very act of waiting, even though waiting itself is not desired, or, more precisely, simply because it is not desired. Therefore, the contradiction between

this principle and the earlier one, which finds waiting to subtract from the profitability of an exchange with a server, is obviated by the term subjective value, to which the objective observer would be quite indifferent.

### Making Others Wait

That waiting (within the limits referred to) will render a service more valuable, independently of its objective worth, seems to be an inherent feature of the psychology of social exchange. This property is perhaps made most intelligible by the principle of cognitive balance, which, according to Alexander and Simpson (1964, pp. 182–92), tends to equilibrate psychological investment and profit (for a more general statement, see Festinger [1957]). However, Simmel's was the first systematic discussion of this principle. His treatment is summarized in the observation that "even if [objects or services] possess no intrinsic . . . interest, a substitute for this is furnished by the mere difficulty of acquiring them: they are worth as much as they cost. It then comes to appear that they cost what they are worth."

We may turn to an important implication of this principle. Because the worth of a person is not independent of the amount of time others must wait for him, that person can maintain and dramatize his worth by purposely causing another to wait.

Of course, the imposition of a waiting period does not in itself make a person or his services valuable; it can only magnify existing positive evaluations or transform neutral feelings into positive ones. If these initial feelings are not favorable, or at least neutral, the waiting caused by a server may lower clients' estimations of his worth. Instead of a sought-after and important man, the server becomes an incompetent who cannot perform his job properly; thus is his initial inferiority confirmed. (This is why subordinates who know where they stand do not like to keep their superiors waiting.) Generally, the dramatization of ascendancy by keeping another waiting will do a server the most good when his social rank exceeds that of his client or when the difference between their ranks is ambiguous. In the latter case, ascendancy accrues to him who can best dramatize it; in the former, ascendancy may be dramatized by him to whom it already accrues.

Thus, just as authority is affirmed by the placement of social distance between super and subordinate, so temporal distance subserves the ascendancy of the person who imposes it. More precisely, the restriction of access to oneself by forcing another to "cool his heels" is instrumental to the cultivation of social distance. The importance of this point resides in its inconsistency with the assumption that waiting is primarily dependent upon the supply of servers and demand for their services. The kind of waiting



to which we now call attention is "ritual waiting," imposed without reference to scarcity of server time.

Now, ritual waiting is a form of mystification.

### Waiting and Mystification

Causing another to wait is a form of "mystification" (see Goffman 1959, pp. 67-70) because self-imposed restriction on accessibility underscores a server's scarcity and social value, thereby promoting awe among those who wait for him. Notwithstanding our own attempt at elaboration and extension, however, this line of thought does not take us far enough. For, if the reverence in which a server is held is to be profoundly felt, it must rest not only upon the essentially negative capacity to regulate access to himself; the server must also display the more positive ability to satisfy needs or alleviate tension within the person waiting for him. In holding himself apart, then, the charismatic server must also "do something" for the client. Furthermore, if whatever is done is to dazzle the client, its efficacy must apparently derive from the very person of the server, independently of the particular substantive benefits he is capable of providing. The latter, it might seem to the client, flow from the status of the server, but not from his specific individuality.

This consideration enables us to see in the ability to make others wait an ideal resource for mystification. For, when, after waiting some time, a client's turn is finally called, the summons itself fulfills a need which, having been generated by the distress of waiting for a service, can have nothing to do with the need for the service itself. It may be argued that the distinction between these two sources of tension is merely of analytic worth; that, empirically, they merge insofar as the waiting period may exacerbate anxiety over the condition that requires servicing. But this objection only confirms the fact that by simply making himself available the server can display a remarkable personal capacity to alleviate suffering. Because he is so intensely waited for, his very appearance makes us feel better. Hence the impression of an inherent power to relieve stress. (In this regard, see Bettelheim [1960], p. 87.) Because he is explicitly defined as the one to wait for (with all the messianic implications of such definition), the tension attending the wait can be relieved in no way other than through his appearance.

Lest it appear that the delay creates rather than enhances attraction, we should stress that the above statements are preconditioned by a performer's initial appeal. In this connection, an additional qualification must be made. Causing a delay will not only fail to enhance the status of an unattractive server; it will also fail to elevate the server who cannot conceal from a

client the fact that he is deliberately making him wait. For, if the object of imposing a delay is to give the impression of important business when none really exists, then the initial sense of awe must turn into infuriation when the mystery of power is seen through.

*Servers who do not serve.*—In suggesting that a server may dramatize the scarcity and value of his skills by making others wait for him we imply that he eventually must appear and provide his services, otherwise he could not possibly profit from their increased value. This implication is certainly valid in connection with most server-client relationships; but it does not flow from the most pronounced form of the histrionics of scarcity: when, in the face of the most intense anticipation, the server never appears! The sense of awe thereby occasioned, moreover, is perhaps most poignant when the server himself is unknown to his attendants, for what then emerges is the unadulterated sense of anticipation itself, uncontaminated by any personal reference. Such is the case of *Godot* (Beckett 1954), whose efficacy lies in no concrete, substantive achievement but in the pure fact that he is waited for.

*Delay and the maintenance of status boundaries.*—This radical case points up the two contradictory tendencies that are common to the standpoints of all servers: (1) the desire to enter at once into relations with others for “instrumental” reasons and (2) the impulse to hold oneself apart from them for “expressive” ones. This dilemma has a structural as well as a psychological referent. We know that the maintenance and purposes of social organizations require social contact not only among constituents of a single stratum but also between members of different (higher and lower) strata. This prerequisite poses a problem because interpersonal contacts between strata tend in diverse ways to undermine the distance and erode the barriers that distinguish them. Dedifferentiating tendencies of this sort could only redound to the disadvantage of the superior, who profits both materially and morally in proportion to the decisiveness of the separation. The delaying ritual of waiting helps resolve this dilemma. Although the status gap must be bridged by social contact, the contact itself can be depersonalized and formalized; it can be made “by appointment only.” This practice follows status lines in a very clear-cut way. While the factory worker, for instance, may approach his peers or even his foreman without appointment, he cannot do so if he is to meet with an executive. The subordinate must be delayed before he is allowed to make a cross-stratal contact. Such inconvenience nicely preserves the sense in which the superior is symbolically inaccessible to those beneath him. While an interactional breach of status boundaries may occur, it can be ritualized in a way which makes it appear that it does not.

The Imposition of Waiting as an Aggressive Act

If the temporal aspect of relationships between those occupying different social positions may be stated in terms of who waits for whom, then we would expect to find a reversal of the waiting-delaying pattern when persons "switch" positions. Furthermore, this reversal may be accentuated through retaliation by the one who suffered under the initial arrangement. A former president furnishes us with an example:

Ken Hechler, who was director of research at the White House from 1948 to 1952, recalled the day Mr. Truman kept Winthrop Aldrich, president of the Chase Manhattan Bank, waiting outside the White House office for 30 minutes. Hechler quoted Mr. Truman as saying:

"When I was a United States senator and headed the war investigation committee, I had to go to New York to see this fella Aldrich. Even though I had an appointment he had me cool my heels for an hour and a half. So just relax. He's got a little while to go yet." [*Chicago Daily News*, December 27, 1972, p. 4]

Punitive sanctioning through the imposition of waiting is met in its most extreme forms when a person is not only kept waiting but is also kept ignorant as to how long he must wait, or even of what he is waiting for. One manifestation of the latter form is depicted by Solzhenitsyn (1968a):

Having met the man (or telephoned him, or even specially summoned him), he might say: "Please step into my office tomorrow morning at ten." "Can't I drop in now?" the individual would be sure to ask, since he would be eager to know what he was being summoned for and get it over with. "No, not now," Rusanov would gently, but strictly admonish. He would not say that he was busy at the moment or had to go to a conference. He would on no account offer a clear, simple reason, something that could reassure the man being summoned (for that was the crux of this device). He would pronounce the words "not now" in a tone allowing many interpretations—not all of them favorable. "About what?" the employee might ask, out of boldness or inexperience. "You'll find out tomorrow," Pavel Nikolaevich would answer in a velvet voice, bypassing the tactless question. But what a long time it is until tomorrow. [P. 222]

The underlying technique for the aggressive use of delay involves the withdrawal or withholding of one's presence with a view to forcing another into an interactionally precarious state wherein he might confront, recognize, and flounder in his own vulnerability or unworthiness.<sup>10</sup> By such means, the superordinate not only affirms his ascendancy but does so at

<sup>10</sup> Of course, the impulse of stationary servers to make others wait for reasons that are independent of the scarcity of time is paralleled by the tactic, used by mobile servers, of keeping them waiting for these same reasons. Thus, a person may simultaneously exhibit contempt for a gathering and underscore his own presence (Parkinson 1962, pp. 73–74) by purposely arriving late. This measure is particularly effective when the proceedings require his presence.

the direct expense of his inferior's dignity. Russian bureaucrats are masters at invoking this routine in their dealings with waiting clients:

Casting a disapproving eye at the janitor's wet overshoes, and looking at him severely, Shikin let him stand there while he sat down in an arm-chair and silently looked over various papers. From time to time, as if he was astonished by what he was reading . . . , he looked up at him in amazement, as one might look at a man-eating beast that has finally been caged. All this was done according to the system and was meant to have an annihilating effect on the prisoner's psyche. A half-hour passed in the locked office in inviolate silence. The lunch bell rang out clearly. Spiridon hoped to receive his letter from home, but Shikin did not even hear the bell; he riffled silently through thick files, he took something out of a box and put it in another box, he leafed, frowning, through various papers and again glanced up briefly in surprise at the dispirited, guilty Spiridon.

All the water from Spiridon's overshoes had dripped on the rubber runner, and they had dried when Shikin finally spoke: "All right, move closer!" [Solzhenitsyn 1968*b*, pp. 482-83]

This kind of strategy can only be employed by superordinates who have power over a client in the first place. The effect on the client is to further subordinate him, regardless of a server's initial attractiveness or a client's realization that the delay has been deliberately imposed. Furthermore, this practice leaves the client in a psychologically as well as a ritually unsatisfactory state. The two presumably act back on each other in a mutually subversive way, for by causing his client to become tense or nervous the server undermines the self-confidence necessary for him to maintain proper composure. This tendency, incidentally, is routinely applied by skillful police interrogators who deliberately ignore a suspect waiting to be questioned, assuming that a long, uncertain wait will "rattle him" sufficiently to disorganize the kinds of defenses he could use to protect himself (Arthur and Caputo 1959, p. 31).

### Ritual Waiting and Autonomy

We have tried to show that while servers may cause others to wait in order to devote their attention to other necessary matters, they may also make people wait for the pure joy of dramatizing their capacity to do so. Such elation, we saw, is understandable, for by effecting a wait the server demonstrates that his presence is not subject to the disposition or whim of another and that access to him is a privilege not to be taken lightly. And, if access is a privilege, then one may sanction another by deliberately holding oneself apart from him. But we must now make explicit a point that was only implied in our previous discussions: that the imposition of waiting expresses and sustains the autonomy as well as the superiority of the self.

While the imposition of delay allows a superordinate to give expression

to his authority, waiting may also be imposed in protest against that authority. The latter achievement is valued, naturally, among those of despised status and low rank. Because they lack the wherewithal to do so in most of their other relations, the powerless, in their capacity as servers, delight in keeping their superiors waiting. The deliberately sluggish movements of many store clerks, telephone operators, cashiers, toll collectors, and the like, testify to the ability of the lowly as well as the lofty to dramatize their autonomy. This accords with Meerloo's (1966) assertion that "the strategy of delay is an ambivalent attack on those who command us" (p. 249). This kind of aggression is perhaps most pronounced under sociologically ambivalent conditions: as the legitimacy of the existing distribution of status honor ceases to be taken for granted, prescribed deference patterns give way to institutionalized rudeness, which may be expressed by appearing late for appointments with a superordinate as well as by dillydallying while he waits for his needs to be serviced.

It goes without saying that members of the dominant class are not above such invidious intention. Often having the means to do so, they merely execute it in a diametrically opposite manner: by compulsively refusing to wait. These are the people who are targets of advertising campaigns through which establishments of various sorts announce that their customers do not wait as long as those who shop in lower-priced competitor stores. General store chains (specializing in groceries) have run such ads in the recent past. That time may be a marketable commodity is also confirmed by the frequency with which we observe "No Waiting" signs in the front windows of barber shops. Similarly, those who object to waiting as a matter of principle find satisfaction in "instant-on" television sets, Polaroid photographs, etc. For many persons, the higher cost of using such services is offset by the personal sense of self-worth and autonomy thereby affirmed. By paying a higher price, the individual may back up his claim that he is "not the kind of person who will be kept waiting." It needs to be stressed that the inflated price he pays is instrumental to this act of self-affirmation: the value of his time and, therefore, his self, is enhanced precisely because another value is sacrificed on its behalf; the individual thus convinces himself, and perhaps others, that he easily pays a cash price for the opportunity to dispose of it as he wishes. He presumably has "better things to do with his time" than to expend it behind a queue of others (whose time and selves—because they are willing to wait—may not be as valuable as his own).

### Ceremonial Waiting

Because unwillingness to wait embodies a rejection of both the auspices under which it is demanded and the inferior self that awaits the incumbent



of the waiting role, it may be said that readiness to wait symbolizes a measure of deference toward the authority who imposes it. Those who are kept waiting beyond the appointed time by very high political or professional figures, for instance, may not exhibit indignancy or sullenness at being delayed; on the contrary, the client must exhibit gratitude that an audience is granted at all. Thus, the waiting period that is taken in stride by the client of an internationally applauded brain specialist would give rise to seething if inflicted by the neighborhood dentist.

This variation in waiter irritation is governed by a general rule: the more pronounced the honor of the server, the longer we are expected to willingly wait for him. One of the clearest instances of this rule is found in those colleges which have "more or less unofficially standardized periods that students are to await a tardy teacher, and in some instances the period is graded according to the teacher's rank" (Moore 1963, p. 53).

If readiness to wait with good grace conveys an individual's deference to a person more elevated than himself, we should not be surprised to find an inferior waiting for the very purpose of expressing deference. This form is perhaps most conspicuous in its collective expression. On a very cold day in 1963, for example, almost 250,000 people waited up to 10 hours outside the Capitol Rotunda, where President Kennedy lay in state. Such a massive collective deference gesture can be made intelligible by reference to a simple principle. Given the charisma of its object, an event may presumably be so awe inspiring as to render banal and irrelevant—even profane—whatever one might oneself do. One consequently measures up to the occasion by doing nothing at all. Moreover, because suspension of activity in deference to another entails forfeiture of alternative activities and associated rewards, deferential waiting comes sharply into view as a functional equivalent to sacrifice. When in addition the renunciatory deferential tribute is rendered in proximity to a sacred center, its personal meaning is naturally intensified and focused. As one member of the long queue leading to the Capitol Rotunda put it: "We were going to watch it on television in our room at the 'Y.' But the more we watched the more we felt we just had to be here ourselves. It's so awful we felt we had to do something—something" (*New York Times*, November 25, 1963, p. 5). To wait deferentially at a sacred center is thus to be "where the action is" or, more precisely, where the in-action is.

*The respect pattern.*—The above is simply an extraordinary expression of the mundane tendency for persons to subject themselves to a wait as a sign of deference for those with whom they have an engagement. Hall (1959) refers to this as "the respect pattern" (p. 18) which inclines persons to arrive a little early for meetings and rendezvous so as not to subject another or others to such inconvenience and abasement as has been herein described. Self-imposed waiting is governed by the same rule that regulates

the impatience of those on whom waiting is imposed by another: the higher the rank of that other, the more imperative an unambiguous demonstration of the respect pattern becomes. For example, White House etiquette (as enunciated by Emily Post [1965]) dictates that: "When you are invited to The White House, you must arrive several minutes, at least, before the hour specified. No more unforgiveable breach of etiquette can be made than not to be standing in the drawing room when the President makes his entry" (p. 48). One of the most radical modes of this kind of ceremonial waiting is found in the Ethiopian practice of "Studying the Gate."<sup>11</sup> This involves a procedure followed by those who desire an audience with the emperor, for whose sake callers arrive several hours before their appointment and wait patiently outside the door leading to his chamber. Thus situated, visitors exhibit their respect, subjects their devotion, to him.

Individuals may express deference not only by arriving early and waiting for the appearance of a distinguished person, they may also wait for the departure of that person before leaving themselves. For some occasions this possibility becomes an imperative. Thus, according to Fenwick (1948), "The two cardinal points of White House Etiquette are that no guest is late and that no guest leaves before the President and his wife have gone upstairs" (p. 469). This rule shows that just as waiting may ritually precede access to another it may also precede his departure. A most radical example of the latter is the phenomenon of the death vigil, wherein a group awaits news of the passing of a prominent member and disperses when it is received. This form stands as a functional parallel to waiting for an honored person to depart before leaving oneself.

When juxtaposed with our initial remarks, such considerations as these (as well as others introduced in this paper) admit of a typological possibility that deserves more singular attention than it has up to now received.

*Instrumental waiting and ceremonial waiting.*—It is possible to distribute empirical instances across a continuum limited at one end by purely instrumental waiting, necessitated by the server's interactional inaccessibility due to real demands on his time and energy. Between the poles of this continuum we find cases which present the difficulty of ascertaining to what extent the wait may, on the one hand, be occasioned by the server's objective scarcity and, on the other, by the demand for temporal tribute implied in his refusal to open himself up for interaction at the first available instant.<sup>12</sup> Perhaps most cases would fall into this middle category, for, as Shils (1970) suggests, "Deference actions are not . . . always massive actions of much duration.

<sup>11</sup> This practice was explained to me by Donald N. Levine.

<sup>12</sup> What has been said in reference to delay may also apply when clients are seen immediately. In this case, too, it is often difficult to tell whether a server wishes to ritually acknowledge a client's worth or whether that client is seen at once because there are no other demands on the server's time.

They occur moreover mainly at the margin of other types of action. . . . Between beginning and end, deference actions are performed in fusion with non-deferential actions" (p. 433). However, we have observed deferential waiting in far less attenuated, far purer (indeed, ceremonial) forms. These constitute the limit of the other end of the continuum to which we are referring.

As an extreme case, ceremonial waiting sets in relief the devaluating aspects of waiting in more ordinary contexts. Precisely because it exaggerates their degradational implications, ceremonial waiting permits us to analyze these less radical forms in terms of their ritual "distancing" or "boundary maintenance" functions, through which superordinates may dramatize and so confirm their position in the social structure. However, we must not forget that the superordinate may be challenged by the very same means through which he confirms himself. We have ourselves observed compulsive refusals to wait. In view of this, we must concede that ritualized status-elevational possibilities—or, at least, reaffirmational ones—exist on both sides of the server-waiter relationship.

#### SUMMARY

Delay is a relevant sociological datum because it is general throughout society, is a measure of access to goods and services, and indexes the efficiency of the organizations which distribute them. Above all, delay entails two kinds of very conspicuous costs. Having nothing to do with waiting as such but rather with the losses occasioned by it, value foregone through idleness is an extrinsic disadvantage. On the other hand, the degradational implications of being kept idle are intrinsic to waiting and can arise in no way other than through involuntary delay. The purpose of this paper was to explore the way these costs are distributed throughout the social structure and to identify the principles to which this allocation gives expression.

We have introduced the category of power, as exercised in server-client relationships, as the ultimate determinant of delay, the main assertion being that the distribution of waiting time coincides with the distribution of power. This proposition turns on the assumption that power is related to the scarcity of the goods and skills that an individual server possesses. Accordingly, the relationship between servers and clients in respect to waiting is an instance of an "organized dependency relationship" (Stinchcombe 1970): servers' holding power is contingent upon clients not being able to frequent less accessible and/or more expensive servers, while client autonomy requires their availability. Delay is therefore longest when the client is more dependent on the relationship than the server; it is minimized, however, when the server is the overcommitted member of an asymmetrical

relationship.<sup>13</sup> Personal and structural factors thus stand as intersecting contingencies: resourceful persons wait less within both competitive and monopolistic markets, while delay will be more pronounced in the latter regardless of personal power.

If waiting is related to a person's position in a power network, then a server may confirm or enhance his status by deliberately making another wait for him. In a more general sense, this is to say that the management of availability itself, regardless of the purpose for which an individual makes himself available, carries with it distinct psychological implications. Because a person's access to others indexes his scarcity as a social object, that person's social worth may only be realized by demonstrated inaccessibility. Openness to social relations may therefore be restricted not only to regulate interactional demands but also to enhance the self that one brings to an interaction. Because it is independent of the objective scarcity of servers and their resources, this type of delay was subsumed under the category of "ritual waiting." This form finds expression in positive as well as negative respects: just as a server may deliberately limit access to himself, so a client may wait when it is circumstantially unnecessary in order to exhibit deference to a server. The initial relationship between waiting and power thus gives rise to processes which strengthen it. That is to say, secondary dramaturgical modes have come to subserve a fact that was originally grounded in an objective supply-demand structure.

The broader implication of this essay is that it finds in time itself a generalized resource whose distribution affects life chances with regard to the attainment of other, more specific kinds of rewards. This is true in a number of respects. Time, like money, is valuable because it is necessary for the achievement of productive purposes; ends cannot be reached unless an appropriate amount of it is "spent" or "invested" on their behalf. On the other hand, the power that a time surplus makes possible may be protected and/or expanded by depriving others of their time. By creating queues to reduce idle periods, for example, a server exploits clients by converting their time to his own use. A server does the same by "overcharging" in the sense of deliberately causing a particular client to wait longer than necessary.

The monetary analogies we have used are not without some justification. Just as money possesses no substantive value independent of its use as a means of exchange, time can only be of value if put to substantive use in an exchange relationship. Both time and money may be regarded as generalized means because of the infinity of possibilities for their utilization: both are possessed in finite quantities; both may be counted, saved, spent, lost, wasted, or invested. And, just as the budget (which, for Weber [1964],

<sup>13</sup> The subsumption of the server-client relationship under the concept of differential commitment was suggested to me by Philip Blumstein.



is the highest form of economic rationality) “states systematically in what way the means which are expected to be used within the unit for an accounting period . . . can be covered by the anticipated income” (p. 187), so the time schedule—which may be the highest form of interactional rationality—states in an identical way how the time required for the performance of numerous activities can be covered by its anticipated availability. Accordingly, while the powerful can allocate monetary means to their own desired ends by controlling the budget, they also regulate the distribution of time—rewarding themselves, depriving others—through their control of the schedule. What is at stake in the first instance is the amount of resources to which different parts of a system are entitled; in the second, it is the priority of their entitlements. Far from being a coincidental by-product of power, then, control of time comes into view as one of its essential properties.

## REFERENCES

- Alexander, C. Norman, and Richard Simpson. 1964. “Balance Theory and Distributive Justice.” *Sociological Inquiry* 34 (Spring): 182–92.
- Arthur, R., and R. Caputo. 1959. *Interrogation for Investigators*. New York: Copp.
- Beckett, Thomas. 1954. *Waiting for Godot*. New York: Grove.
- Bettelheim, Bruno. 1960. *The Informed Heart*. Glencoe, Ill.: Free Press.
- Blau, Peter. 1964. *Exchange and Power in Social Life*. New York: Wiley.
- Bradford, Jean. 1971. “Getting the Most out of Odd Moments.” *Reader’s Digest* (June), pp. 82–84.
- Fenwick, Millicent. 1968. *Vogue’s Book of Etiquette*. New York: Simon & Schuster.
- Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Evanston, Ill.: Row, Peterson.
- Glaser, Barney. 1972. *The Patsy and the Subcontractor*. Mill Valley, Calif.: Sociology Press.
- Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. Garden City, N.Y.: Doubleday.
- Hall, Edward. 1959. *The Silent Language*. Greenwich, Conn.: Fawcett.
- Lieberman, E. G. 1968–69. “The Queue: Anamnesis, Diagnosis, Therapy.” *Soviet Review* 9 (Winter): 12–16.
- Mann, Leon. 1969. “Queue Culture: The Waiting Line as a Social System.” *American Journal of Sociology* 75 (November): 340–54.
- Mann, Leon, and K. F. Taylor. 1969. “Queue Counting: The Effect of Motives upon Estimates of Numbers in Waiting Lines.” *Journal of Personality and Social Psychology* 12 (June): 95–103.
- Meerloo, Joost. 1966. “The Time Sense in Psychiatry.” In *The Voices of Time*, edited by J. T. Fraser. New York: Braziller.
- Moore, Wilbert E. 1963. *Man, Time and Society*. New York: Wiley.
- Parkinson, C. Northcote. 1962. *Parkinson’s Law*. Boston: Houghton-Mifflin.
- Post, Emily. 1965. *Emily Post’s Etiquette: The Blue Book of Social Usage*. New York: Funk & Wagnalls.
- Shils, Edward. 1970. “Deference.” In *The Logic of Social Hierarchies*, edited by Edward Laumann et al. Chicago: Markham.
- Simmel, Georg. (1907) 1970. “Exchange.” Unabridged translation by Donald N. Levine.
- . 1971. “Exchange.” In *Georg Simmel: On Individuality and Social Forms*, edited by Donald N. Levine. Chicago: University of Chicago Press.



## American Journal of Sociology

Solzhenitsyn, Aleksandr. 1968*a*. *The Cancer Ward*. New York: Dial.

———. 1968*b*. *The First Circle*. New York: Harper & Row.

Spencer, Herbert. 1886. *The Principles of Sociology*. Vol. 2. New York: Appleton.

Stinchcombe, Arthur. 1970. "Organized Dependency Relations and Social Stratification."

In *The Logic of Social Hierarchies*, edited by Edward Laumann et al. Chicago: Markham.

Tawney, R. H. 1931. *Equality*. London: Allen & Unwin.

Weber, Max. (1925) 1964. *The Theory of Social and Economic Organization*. Glencoe, Ill.: Free Press.