CHAPTER ONE

Introduction: The Idea of Local Justice

Jon Elster

LOCAL JUSTICE deals with decentralized in-kind allocation of scarce goods and necessary burdens. In the present volume we discuss the allocation of three goods—college admission, kidneys for transplantation, and immigration rights—and of one burden, viz., layoffs from work. In my earlier book, *Local Justice,* I also considered, albeit much more briefly, the following:

- military service in wartime
- demobilization from the army
- allocation of sperm for artificial insemination
- selection of adoptive parents
- award of child custody
- admission to kindergarten
- division of household work
- allocation of prison space
- rationing in wartime

The task of this introduction is to introduce the basic conceptual issues of local justice problems and to illustrate them with examples taken from the four case studies developed in subsequent chapters. Basically, therefore, what follows is a condensed and more narrowly focused version of Chapters Two through Five of *Local Justice.* More substantive comments on local justice in

America, as revealed by the case studies, are reserved for the conclusion.

The Magnitude of the Problem

To assess the importance of local justice issues in the life of the citizens, one would have to take account of the full range of such questions that can arise in the lifetime of an individual, from the nursery school to the nursing home. I believe we would find that local justice decisions are no less important than the market and redistributive state policies in shaping the life-chances of the individual. This will have to stand as conjecture. It is possible, however, to be more precise about the four issues discussed in the following chapters.

To begin with, importance can be assessed from two points of view. On one hand, we can ask how many individuals are affected by a given type of allocative justice. On the other hand, we can ask how strongly it affects each of those individuals. Immigration becomes very important on both dimensions: it affects huge numbers of would-be immigrants in other countries, for each of whom immigration to the United States is intensely desirable. From a different perspective, the volume and composition of immigration are also very important for those who are already U.S. citizens, notably in an era of kin-based immigration. In quantitative terms, the Immigration Act of 1990 targets about 700,000 immigrant admissions per year.

Going to college is an important determinant of life chances. Going to a particular college is less important. Admissions officers administer a good that is sought after by very many and that may be seen as very important or relatively unimportant, depending on which of these two perspectives is chosen. In quantitative terms, nearly 1½ million students move from high school to one of nearly 3,000 colleges every year. Of these, about half are 4-year colleges that exercise some selection among applicants.

Being laid off from a job can be a very bad thing; keeping your job can be very important. It is also a problem that affects a substantial number of individuals. From 1981 to 1985, 5.1 million workers with 3 years or more of tenure on the job were displaced,

2.8 million due to plant closings, 2.3 million due to layoffs from continuing operations.

Because of the fallback option of dialysis, getting a kidney transplant is not a life-and-death issue (unlike heart and liver transplants). A transplant does, however, represent a very substantial improvement in the quality of life. Currently, over 24,000 people are waiting for cadaveric kidneys, while the available supply of cadaveric kidneys per year has varied between 6,900 and about 7,800 in recent years. In addition, about 2,000 kidneys from living donors are made available each year.

Goods Versus Burdens

In a formal sense, the distinction between goods and burdens is inessential because exemption from a burden can always be conceptualized as a good. In a psychological sense, however, the distinction matters. To assume a burden constitutes a loss compared to the status quo, whereas getting a good represents a gain. Because people have different attitudes toward gains and losses,[1] we should expect to see different principles used in their allocation.[2] For the sake of convenience, this introduction refers only to "goods" as a shorthand for goods and burdens.

Occasionally, it may be hard to say whether a given object of allocation is a good or a bad. As Romm notes in Chapter Four, in many collective bargaining agreements, "senior workers have the right to waive their seniority and take layoff. Apparently, they often do so in the event of temporary layoffs." But in the cases discussed in the present volume, this ambiguity is marginal.

Some Characteristics of Goods

Four major dimensions of the goods allocated in local justice systems are rigidity of supply, divisibility, homogeneity, and scarcity.

Goods are rigid in supply if there are no reliable ways of increasing the available amounts at short notice. The supply of kidneys, for instance, fluctuates for a number of reasons. Also, the government might take legislative steps in order to affect the long-

term determinants of supply. But at any given time hospitals cannot promise a patient that a kidney will be forthcoming, even disregarding the question of goodness of match. The rigidity is overcome only in countries (such as Norway) that rely heavily on relatives as living donors. After kidney allocation, we can rank jobs, college places, and immigration in increasing order of flexibility. These are goods that are either divisible (jobs) or easily expandable at short notice, either at the margin (college places) or in large numbers (immigration places).

Goods can be divisible or indivisible. Because few goods are literally indivisible, indivisibility must be taken to mean that division would entail a drastic reduction or even destruction of the good's utility. To cut a child in two, as in Solomon's first judgment, would not leave anything of value. By contrast, dividing a loaf of bread in two parts entails an increase in total utility, assuming that bread has decreasing marginal utility for each consumer and that the utility functions of different consumers are roughly similar. An intermediate case is found in St. John 19:23–24: "Then the soldiers, when they had crucified Jesus, took his garments, and made four parts, to every soldier a part; and also his coat; now the coat was without seam, woven from the top throughout. They said therefore among themselves, Let us not rend it, but cast lots for it, whose it shall be." The coat would not have been rendered entirely worthless by being cut into four parts. If they had been sufficiently risk averse the soldiers might have preferred getting a quarter of the cloth with certainty over a 25 percent chance of getting the entire coat.[3]

In the local justice issues considered below, the only divisible good is that of work. In this case, work-sharing is a prominent alternative to layoffs and is currently considered in several European countries as a way of reducing double-digit unemployment figures. In general, such schemes may involve either losses or gains in productive efficiency. If there are losses, the employer may insist on lower wages. Yet even working at lower wages, employees may experience a utility gain (compared to a layoff lottery) because of the increasing marginal utility of money. And even if there is no such gain, they may prefer work-sharing because of risk aversion.

Next, the goods to be allocated can be homogeneous or hetero-

geneous. When nothing matters beyond the simple decision of Yes or No, as in immigration, the good is homogeneous; otherwise it is heterogeneous. Heterogeneity can matter in two ways. First, some units of the good may be of inherently higher quality, so as to be more intensely desired by all applicants. Second, some units of the goods may be superior in the eyes of some individuals, but inferior in the eyes of others. Although admission to college in itself is a homogeneous good, the coupling of admission with financial aid gives rise to the first kind of heterogeneity (unless the amount of aid is the same for all). Even with open admission (i.e., in the absence of scarcity), the allocation of financial aid would still create issues of distributive justice. Kidneys are also heterogeneous, because a given kidney may be more or less suitable for a given recipient. Getting a kidney is not all that matters: it is also important to get one with a good match. This illustrates the second kind of heterogeneity. Heterogeneity (especially of the second kind) is a reason against using an allocating mechanism that fails to differentiate among recipients, such as equal division, rotation, lotteries, or queuing. It may or may not be a decisive reason, depending on how and how much the heterogeneity matters and on the strength of other considerations.

Finally, the good can be scarce or abundant. By abundance, I mean that there is enough of the good to satiate everybody who wants or needs it. With indivisible goods, such as immigration, kidneys, or college places, abundance means that there are enough units to satisfy all applicants. With a divisible good, such as work, abundance means that everyone can work as many hours as he or she wants at the going rates. (Note that the allocation of overtime work can also be a matter of local justice.) In cases where there are both homogeneity and abundance, no issues of distributive justice can arise. Although the cases discussed below are all characterized by scarcity, with or without homogeneity, there are instances in which the distributive conflicts are due to heterogeneity under conditions of abundance. The allocation of dormitory places to students or of committee places to legislators are examples.[4] In some countries, parts of the system of higher education also function in this manner: although everyone is guaranteed a place at some university, not all universities are equally attractive or not equally attractive to all applicants. Although Conley's analysis be-

low of college admissions considers the process only from the point of view of the individual university, one might argue that in the whole system of selective colleges almost all applicants are eventually accepted somewhere (partly, however, because of self-selection). In the terminology to be explained below, although each college uses a selection procedure, the system as a whole approaches a placement procedure.

Selection, Admission, and Placement

Following Willem Hofstee,[5] I distinguish among selection, admission, and placement as procedures for allocating indivisible goods. A selection procedure compares individuals against each other, usually by producing a ranking list, and accepts them by starting at the top and going down the list until the good is exhausted. An admission procedure compares individuals against an absolute threshold and offers the good to all those and only those who exceed the threshold. A placement procedure regulates the allocation of nonscarce heterogeneous goods, ensuring that each individual ends up with some unit of the good.

Among the goods discussed here, college places, layoffs, and kidneys for transplantation are allocated by selection procedures. In the case of kidneys—a good with rigid supply—this could hardly be otherwise. Colleges, however, might choose to admit all those who fulfill certain criteria, which might be set so as to ensure an expected number of applicants that is compatible with the capacities and needs of the institution. Because universities can always accommodate marginal adjustments in teacher/student ratios and similar factors, small deviations from the expected number would not matter. Or one could have an admission system for one subset of applicants (e.g., in-state residents) and then use a selection system to fill up the remaining places. Because the pool of applicants is uncertain and comparisons among them are costly, such procedures might have some advantages from the institution's point of view. In the case of layoffs, however, the candidates for layoffs are known; moreover, the only way of establishing the desired cut-off point for (say) seniority is to establish each worker's seniority, which means that one might as well use a ranking list.

As Mackie explains in Chapter Five, the American immigration system has been and remains based partly on admission, partly on selection. On one hand, there is kin-based admission; on the other hand, there is selection by queuing, lotteries, and other procedures within overall quotas for different categories and regions. Because kin-based admission was adopted on the basis of its intrinsic justice rather than as a proxy for a numerical target, it has had a number of unforeseen consequences. In the terminology explained below, the use of admission here serves as a principle of first-order allocation and not—as is the case when admission is a proxy for a numerical target—as a second-order principle.

The four case studies have no examples of a full-blown placement system. The idea that college admission is an indirect placement system has already been noted. Mackie discusses proposals to organize refugee admission as a placement system, with several countries sharing the refugees among themselves.

First-Order, Second-Order, and Third-Order Actors

Following Guido Calabresi and Philip Bobbit,[6] I distinguish between the first-order decisions that determine the amount of the scarce good that is available for distribution, and the second-order decisions that determine how that amount is to be allocated. I deviate from their usage by relying on a more inclusive concept of first-order decisions. Whereas they refer to these as decisions by "society," meaning presumably the political system, I include any decisions, by any set of actors, that are *intentionally* made for the purpose of affecting the supply of the scarce good. (Thus, a change in the speed limit that reduces the number of car accidents and thereby the supply of organs for transplantation is not a first-order determination of kidney transplants.) With regard to immigration, Congress and the president are the only first-order actors. With regard to college admissions, the first-order actors are the board of trustees or regents of the university. With regard to layoffs, first-order decisions are made by top management. With regard to kidneys, the supply is affected by legislative and financial measures and by the decisions by individuals whether to let their organs, or those of relatives, be used for transplantation purposes.

Second-order decisions about allocation are usually made by a different set of actors. The only exception is immigration: here Congress sets the criteria for admission as well as number of admittees. (In fact, as noted above, Congress has less control over the total number of immigrants than over the categories of individuals to be admitted.) In the allocation of college places, admissions officers, constrained by general university policies and (for public schools) by the legislature, are the main second-order actors. In the allocation of kidneys, transplant centers, heavily constrained by the centralized UNOS rules (see Chapter Three), make the crucial second-order decisions. In the selection of workers for layoffs, lower-level managers and personnel managers interact to produce the final list, as described in Romm's interviews. Although in most cases these managers are not directly constrained by legislation, they act to minimize the risk of litigation based on various antidiscrimination laws.

We may also introduce the concept of third-order actors: applicants for—or potential recipients of—the scarce good. In the most extended sense, this could include almost everybody. All non-Americans are potential immigrants to the United States, and everybody can get end-stage renal disease (ESRD). A more tractable idea is to define third-order actors as those who, under some set of current circumstances, would apply for the good. This includes not only actual applicants, but also those who would apply if the institutional rules were changed. Thus, some people may not apply for elite colleges because they know they cannot get in, but they would apply if those colleges switched to a lottery system. As noted above, some people may be deterred from getting on the list for transplantation because of the costly post-operative medication. The vast appeal of immigration lotteries indicates that the number of potential immigrants to the United States in this sense is large indeed.

There are a number of interaction effects among the three levels of decision. A change in the second-order rules may affect first-order decisions. Thus, as Dennis reports, for a while it was widely (but falsely) believed that allowing organs to be transplanted into foreign citizens would reduce the willingness of U.S. citizens to donate. More emphasis on seniority and less on ability in layoff procedures may cause a reduction in profits and thus a

need to lay off more workers. Conversely, a change in the total amount to be allocated may affect the rules of allocation, typically (I conjecture) such that reductions lead to more emphasis on efficiency while increases give more space for equity. New second-order rules can induce new third-order decisions, as applicants modify their behavior so as to become entitled to the scarce good. The use of seniority as a layoff principle induces people to stay in the firm, just as skill-based immigration selection induces potential immigrants to acquire scarce skills, and merit-based admission criteria induce students to work hard in high school. Conversely, unforeseen and undesirable third-order adaptation may induce changes in the second-order rules, to close loopholes. If would-be applicants to a public college begin moving to the state in order to benefit from the less stringent admission criteria or lower tuition rates for residents, the college may respond by imposing a one-year residency condition.

The Place of Money in Local Justice Systems

In *Local Justice* I defined this concept so as to exclude the allocation of money. Although many decentralized allocation processes do concern money, they are usually linked up with the global systems of transfers, subsidies, pensions, and the like. (Here and later *global* refers to the society-wide context, not to the international community.) The progressive income tax, for instance, undoes part of the inequality that is created in the labor market. By contrast, in-kind allocation is rarely subject to mechanisms of this kind. The losers are not compensated. Layoffs are an obvious exception to this statement because of the existence of unemployment benefits. Note, however, that this compensation occurs in the context of allocating a burden rather than a good. Because of the salience of the status quo, there will always be pressure to minimize negative deviations from the reference point.

I should add that in the previous paragraph I conflated two issues that really ought to be kept separate: the place of money in the allocation process and the existence of compensation for the losers. One could well imagine that those who fail in one arena received in-kind compensation in the form of goods allocated in

another arena. Those who are laid off from work might get priority in the college admissions process. Although this mechanism is not in fact observed, a similar one is: Drafted war veterans—the losers in the selection for the burden of doing military service—often get priority for jobs, notably government jobs, and for college and medical treatment. (However, volunteers and career soldiers also enjoy these advantages.)

In addition to money being used to compensate the losers, it can be offered to the winners to enable them to accept the scarce good they have been offered. Colleges do not decide only whom to admit and whom to refuse; they also decide on financial aid to those who are admitted. A similar complementarity between a scarce good and financial aid exists in the arena of kidney transplantation. As Dennis observes in Chapter Three, because of the limited Medicare coverage of post-transplantation medical expenses, "prospective transplant recipients who lack private health insurance might be deterred or even screened out." This observation indicates that money, in addition to being an object of allocation, also can serve as a mechanism of allocation. If the scarce good is useless without money, then refusal to give financial aid effectively amounts to a screening process. As Conley writes in her chapter below, when Yale University wanted to keep Jews out, "financial aid was limited for Jews by limiting financial aid to poor students in general."

Money can also be used as an allocation mechanism in a much more direct way—by allowing the better-off to buy the scarce good for themselves. While common in many other times, places, and arenas, in the United States today this practice is marginal in the four arenas we are considering. Nearly all U.S. universities use "need-blind" admission, in the sense that no student is turned away on the basis of inability to meet the costs of higher education. However, some financially strapped colleges such as Brown and Smith have recently abandoned this policy. Under the current American immigration scheme, 10,000 visas a year (less than 2 percent of the total) are available for those who will invest a million dollars or more in job-creating enterprise employing at least 10 U.S. workers. The National Organ Transplantation Act of 1984 prohibits the purchase of organs for transplantation. Although one might imagine workers who are candidates for layoffs paying

the firm money to keep their jobs, the idea is obviously impracticable and would certainly be illegal. Although one does observe workers taking a wage cut to avoid layoffs, this does not take the form of competition among individual employees.

Finally, income and wealth can be used as status criteria (see below). The "God committee" that allocated dialysis in Seattle (see Dennis's Chapter Three) used income and net worth as indicators of social usefulness. In this case, it was not a question of *using* one's wealth to get the scarce good: rather the mere *possession* of wealth showed that one was worthy of getting it.

Criteria and Mechanisms of Allocation

With the exception of immigration policy, none of the arenas we are considering has a uniform allocation procedure. Private colleges essentially do what they want. With regard to public colleges, admission procedures may be uniform within a state, but can differ strongly across states. Private enterprises use widely differing layoff procedures. Although unionized firms are more likely than the nonunionized to use seniority as a main criterion, the way in which it is implemented and combined with other criteria is far from standardized. More surprisingly, perhaps, the UNOS rules for kidney allocation turn out to have a large scope for local variations (contingent on UNOS approval).

This being said, the elementary building blocks that are used in a given arena are pretty much the same. These pure principles—to be discussed in a moment—are virtually never used by themselves to the exclusion of others. Rather, we find that in the allocation of a given good a number of different criteria and mechanisms are used, in succession or in combination. Mixed principles is the virtually exceptionless rule in local justice. In cases such as immigration policy, the complexity of the mix can become Byzantine in the extreme, as the reader will glean from Mackie's Chapter Five. The other arenas do not, however, lag far behind. Local justice is like a coastline: When we move closer up, the larger-scale variations reproduce themselves in ever-finer grain.

I distinguish among four main categories of pure principles: egalitarian criteria; queuing mechanisms; status-based criteria;

and a trio of criteria based on the individual's situation and performance in, respectively, the past, the present, and the future.

Egalitarian criteria. Equal division is often seen as a baseline, the procedure to be used in the absence of reasons to the contrary. (See the final section of the Conclusion, however, for an argument that in matters of local justice *efficiency* is more frequently used as a baseline.) However, the principle of physically dividing the scarce good so as to give equal amounts to everybody is inapplicable in the arenas of immigration, transplantation, and college admission. Jobs can be divided, however, as witnessed by the use of work-sharing as an alternative to layoffs. Romm's case studies do not find much evidence of work-sharing in practice, however.

When physical division will not work, there is always the alternative of dividing the good probabilistically—through an equal-chance lottery.[7] Although this mechanism has occasionally been proposed for layoffs, kidneys, and college places, it has never to my knowledge been adopted in practice in any of these arenas. Many admission officers will agree that, for a subset of the applicants, the decision to accept some and reject others is for all practical purposes a random event, but they shy back from using a formal lottery. The explanation may lie in Conley's observation that the admission officers to whom she spoke "did not want to alienate guidance counselors or applicants by arbitrary or nonsensical patterns of acceptances and rejections," such as might have been produced by a lottery device. The one arena in which lotteries are extensively used is immigration. Although the visas are actually awarded by a procedure of first-come, first-served by order of mail receipt, under the circumstances (almost a million applications for a few thousand positions) it mimics the features of a fair lottery.

Queuing. I use this word as a general term for a number of different mechanisms that are all based on the sheer passage of time. First, there is queuing proper—standing, sitting, or lying in line for the scarce good. This procedure is not used in any of the cases studied here. Second, there is allocation according to one's place on a waiting list, a procedure that is a partial determinant of the allocation of kidneys and of immigration places. Third, there is

allocation according to seniority, a procedure that is extensively observed in layoff situations. Finally, there is allocation according to age. This criterion can also be seen as a status-based principle. Note that college admission is the only arena in which no form of queuing is used.

When a queuing mechanism is used as the only principle of allocation, only the ordinal aspect of the ranking matters. As mentioned above, however, no actual processes are based on a single principle. When combined with other criteria, one can also take account of the cardinal aspect of the ranking. Romm finds that in some collective bargaining arrangements there is a quality-seniority tradeoff, in which length of seniority and not only the mere fact of seniority matters. In the early forms of the UNOS allocation system, candidates for transplantation received points for their ordinal place on the waiting list, a system that is capable of producing anomalies.[8] More recent forms, based on the number of months on the waiting list, are not vulnerable to this problem.

Status-based criteria. Status properties belong to the individual by virtue of biological, behavioral, or institutional determinations. (The phrase is vague and perhaps inadequate: the reader is asked to focus instead on the examples that follow.) Status properties that can matter for local justice include age; gender; race and ethnicity; bodily features, such as height, weight, and color of eyes or hair; physical and mental disabilities; literacy; sexual orientation; wealth; residence; occupation; trade union membership; religion; citizenship; marital status; and kinship. Sometimes these criteria are used because they are believed to be appropriate in their own right: American citizens ought to understand English. In other cases, they are used as observable proxies for other criteria: kin-based immigration is based on an assumption that close relatives have a need to be with each other.

The active reliance on status criteria is most prominent in immigration policy. The main status criteria used to screen immigrants have been literacy, national origin, place of birth (not—as Mackie explains—always the same as national origin), occupation, marital status, and kinship. In other arenas, status is today increasingly seen as irrelevant. Whereas earlier generations of applicants to college could be asked to send a photo or indicate their height

and weight (to screen out the socially and ethnically undesirable), such practices have been discontinued. But children of alumni and staff still get preferential treatment in admission, as do in-state residents when applying to state universities.

Yet these are secondary matters compared to the overwhelming importance of race and, to a somewhat lesser extent, gender. A full discussion of the role of these criteria is postponed until the Conclusion. Here I note only that they can enter the picture in three different ways. First, there is the principle that allocation shall be status-blind: it ought to proceed as if the applicants' race or sex were unknown. Next, there is the idea that allocation shall be status-representative, in the sense that the status composition of the recipients should match that of the population at large or, more frequently, of the applicants. Third, there is the idea of affirmative action: the allocation of goods such as college places and jobs shall be status-offsetting, to improve the position of less-favored groups in society.

Desert, need, and efficiency. I group these criteria together, perhaps somewhat artificially, because they can be placed on a temporal continuum. Allocation according to desert looks to what the applicants have done in the past. Allocation according to need considers their situation in the present. Allocation according to efficiency compares how well they can use the scarce good to benefit themselves or others in the future. In addition, one may combine backward-looking and forward-looking schemes in *incentive systems,* which operate by telling individuals at time t_1 that their reward at time t_3 will depend on their performance at time t_2.

Layoffs according to seniority is often presented as a desert-based system: workers who have devoted their life to the firm deserve to be kept on. The practice can also be seen as an incentive system: to reduce turnover, management tells the workers that they are more likely to be protected from layoffs if they stay in the firm. Basing college admission on high-school grades can also be justified in both perspectives. Thus, nearly one-third of the admission officers in Conley's survey said that they used grades as an admission criterion to set the right incentives for high-school students. There may have been an element of desert reasoning when a large number of Vietnamese who had taken part in the

American war effort were accepted as refugees. (As far as I know no Vietnamese were induced to collaborate by promises of American citizenship in case the United States lost the war.) Neither desert reasoning nor incentive reasoning is used in the allocation of kidneys for transplantation. Although one might imagine a system in which people were turned down for transplantation if they get ESRD as a predictable outcome of their life style, proposals to this effect have not (yet) had much impact.

In our case studies, the most prominent use of need as an allocative criterion is found in kin-based immigration. There are also elements of need-based reasoning in the allocation of kidneys, notably when priority is given to patients who need a kidney urgently because of failure on dialysis. In such cases, however, the same priority can be justified on the basis of forward-looking reasoning. A kidney is used more efficiently if it raises the survival chances of one patient from 0 to, say, 50 percent than if it is used to get another patient who could survive on dialysis off that treatment.[9] In current American layoff practices, need—as measured by the number of family dependents—is not a major consideration, although management will sometimes take it into account.

Efficiency considerations, in the context of local justice, must not be confused with global ideas of utilitarianism or wealth-maximization. Only in the case of immigration decisions are efficiency criteria—giving preference to workers with rare skills and to people willing to invest in jobs for Americans—interpreted in a global sense. In the context of kidney allocation, efficiency is taken to mean that the kidney should be given to the patient with the smallest probability of graft rejection. We might get quite different choices if we wanted to maximize the number of life years spent off dialysis (a criterion that would favor young patients), and different results still if the idea was to minimize the cost of treatment (a criterion that would favor those who can go back to a paid job). In the layoff context, efficiency is often taken to imply that one should retain the more able workers in order to maximize the profits of the firm. Again, we might get very different choices if we took into account the social costs of caring for workers who are unable to find a job elsewhere (assuming that the firm could afford to retain them). With regard to education, efficiency is sometimes taken to mean that admissions should maximize "social value

added." In practice, this criterion cannot be implemented. To do so, an admissions officer would have to calculate the difference between the social value that would be produced by admitting a given applicant and the value that would be produced if that candidate was accepted by some other college to which he had applied—a task that is essentially impossible, especially if admissions officers at other colleges are supposed to reason in the same manner. In practice, therefore, efficiency is usually taken to imply admission of the students who will graduate with the best grades.

Mixed principles. In practice, as I said, all systems of allocation are mixed. The mix can take a number of different forms. Sometimes a subset of the scarce good is allocated by one pure principle, another subset by a different principle, and so on. Immigration policy is organized along these lines, as detailed in Chapter Five below. In college admissions, an analogy is Harvard's former practice of creating a "happy bottom quarter" by admitting 25 percent of each class on athletic rather than scholastic criteria. The main procedure in admissions, as well as in layoff selection and kidney allocation, is nevertheless to apply uniform but complex principles to all candidates.

A common way of combining different pure principles is by using a point system. Kidney allocation in the United States is currently regulated by a procedure of this kind, with points given for medical efficiency (likelihood of graft survival, which is based on the number of antigen matches) and time on the waiting list. Some state colleges use a point system in their admissions process, with weights assigned to high-school grades and tests according to their ability to predict future performance. Informally, some colleges also assign points for disadvantages of various kinds. As Conley explains, this process may be backward-driven, in the sense that the weights are assigned so as to get a preset number of minority applicants admitted. In Chapter Five, Mackie details the failed attempt by Edward Kennedy and Alan Simpson to introduce point systems in the immigration process. Although such systems are occasionally used to regulate layoffs in other countries, this procedure is not observed in the United States.

Criteria may also be combined lexicographically, with the secondary criterion being used to break ties. If the first criterion is

very coarse-grained, ties may be very common, so that the secondary criterion becomes in fact the more important one. As Romm explains in Chapter Four, this is the case for many layoff procedures that use seniority to break ability ties. Other layoff procedures follow a different two-step sequence. First, one determines the subset of candidates with "minimal qualifications" to do the job. In the next step, seniority is used to select from within this pool. Romm also points to the use of informal tradeoffs between level of qualification and length of seniority. In the college admissions system such tradeoffs are ubiquitous, especially in smaller institutions. Interviews with admissions officers show that the tradeoffs, while informal and intuitive, also have a great deal of structure (see the grid reproduced by Conley as Figure 2.1). A salient finding is that whereas many private institutions trade nonacademic excellence off against academic merit, public institutions are more likely to set up a tradeoff between economic or cultural disadvantage and academic merit.[10]

Further Effects of These Criteria and Mechanisms

The primary effect of an allocative scheme is to channel the good to the group of recipients defined by the criteria and mechanisms embodied in the system. In addition, there are at least two other effects. First, the scheme may have a disparate impact on groups not specifically targeted or excluded by the scheme. If the scheme tells the allocators to direct the good to members of group X, and most X's are also Y's, then a de facto consequence of the scheme is to allocate the good disproportionately to members of Y. When these *secondary effects* are in fact intended by the rule makers, legal scholars refer to them as disparate intent. If a rule maker who wants to target group Y finds himself prevented by the law or public opinion from doing this, he may achieve substantially the same goal by allocating the good to X. (To escape detection, it may then be essential that only most—rather than all—X's are Y's.) Second, an allocative system may change the behavior of first-order and third-order decision makers by virtue of the *incentive effects* created by the scheme.

The theme of secondary effects or disparate impact has an

important place in all the case studies. On its face, Yale's rule of refusing financial aid to poor applicants (see above) was racially neutral. De facto, however, it had a disproportionate (and intended) impact on Jewish applicants. Another facially neutral Yale policy from the 1920s was that of seeking geographical diversity in the incoming class, thus (intentionally) reducing the number of applicants admitted from the populous and heavily Jewish state of New York. Today, as mentioned above, some schools will target the educationally disadvantaged as an indirect way of raising the number of minority students admitted.

In the layoff arena, Romm explains in some detail how seniority systems can have a disparate impact on women and members of minority groups. Although these systems are not in themselves discriminatory, their disparate impact is a result of past discrimination, which has prevented these employees from accumulating the necessary seniority. As a result, some courts have found that seniority systems are in violation of Title VII of the Civil Rights Act of 1964. (Most court decisions, however, go the other way.) Romm's interviews also show that managers are very aware of the possibility that a layoff scheme might have a disparate impact on members of protected groups.

With regard to transplantation, Dennis shows how the facially neutral principle of allocating kidneys according to medical efficiency leads to disproportionately fewer blacks receiving transplants. This outcome comes about for three reasons: blacks have different antigen patterns than those of the white population; they are overrepresented in the population of patients; and they are underrepresented in the population of donors. By contrast, he reports that the underrepresentation of women among the recipients of kidney transplants does not seem to be a secondary effect of the allocative system.

As Mackie explains, disparate intent is rampant in immigration. The facially neutral literacy test was intended to screen out immigrants from Southern and Eastern Europe and was abandoned when literacy in this region rose. Similarly, the search for a national-origins baseline to be used in calculating immigration quotas was blatantly influenced by the ethnic biases that would be produced.

I have already mentioned that local justice schemes can be

set up for the purpose of generating certain desirable incentives among the potential recipients. Many incentive effects, however, are entirely unintended and unforeseen. Often, rule makers naively assume that applicants for the scarce good are naive, in the sense that they will act as if they were ignorant of the rules. When applicants adapt strategically to the rules, the intentions of the rule makers may be thwarted. While quite important in many other arenas—for instance, if married men or students are exempt from military service, more will marry or go to college—this problem seems to be somewhat marginal in the four arenas under consideration here, which is not to say that it is totally absent. It is plausible that layoff systems based on seniority set up a disincentive for acquiring skills, but I have no information about the magnitude of this effect. If spouses are exempt from immigration quotas, some arranged marriages may result. Applicants to college may try to manipulate residence requirements to their advantage. If kidneys are made available to foreign nationals, some patients from abroad may visit the country who wouldn't otherwise have done so.

I have also mentioned that allocative rules can set up incentive effects for the first-order actors who determine the amount of the scarce good that will be available for allocation. In the four case studies, this connection seems to be important only in the field of kidney transplantation. According to Dennis, there is not much truth in the once-popular idea that allowing transplantation to foreign nationals reduces the willingness to donate organs. He argues, however, that there is a close connection between the system of kidney allocation and the level of kidney procurement. Because the actual extraction of a kidney has to be done by a transplant surgeon, allowing him to keep one of the two kidneys for use in his own transplantation center tends to stimulate his procurement efforts.

Explaining Local Justice

Once we have identified the mechanisms and criteria used to allocate a particular scarce good in particular time and place, we can move on to try to explain why that particular scheme was adopted.

Some writers, notably Calabresi and Bobbitt, rely heavily on functional explanation. I have discussed their view, and my disagreements with it, in *Local Justice*. In my opinion, intentional explanations are more promising. Such explanations have two main elements: the preferences of the various actors involved and the constraints that they face.

Beginning with the constraints, let me first note that these can be more on the hard side or on the soft side of the spectrum. In transplantation there are a number of absolute constraints that are not the object of choice and deliberation. Compatibility of blood type must be respected. Transplantation is doomed to fail if the recipient has preformed antibodies against the antigens of the kidney. By contrast, the importance attached to goodness of match between antigens is a matter of opinion and indeed of controversy, partly about factual matters (does goodness of match matter in the post-cyclosporine era?) and partly about normative matters (is emphasis on matching unfair to some patient groups?).

To be able to practice selective layoffs, rather than closing down altogether, a firm must survive to make a profit. In some contexts this is a hard constraint, which will rule out layoff schemes that have very generous severance payments and early retirement schemes or that emphasize seniority at the expense of ability. In other contexts, state willingness to subsidize ailing enterprises makes for "soft budget constraints." Although that phrase was originally coined by Janos Kornai to describe how firms operate under Communism, it also applies to many non-Communist societies with active industrial policies. Many U.S. firms would have to lay off more workers or go out of business completely were it not for protectionist government policies. With soft budget constraints, more generous layoff schemes become feasible. I do not mean that the subsidies are made for the purposes of supporting a particular layoff scheme. Rather, generous or costly schemes emerge because the subsidizing authorities are unable to determine the exact cause of the firm's losses and because managers do not want to confront the workers if they do not have to.

Rule makers are to some extent also constrained by public opinion. The history of local justice is filled with scandals caused by revelations that the allocating institutions practice discrimination, tolerate waste, allow for strategic manipulation by recipients, or

use discretionary procedures that invite bribes, caprice, and irrelevant intrusion of personal values. The reader will notice that the history of immigration and ESRD technology, in particular, is littered by scandals that have forced radical changes in first- and second-order decisions. It is of no particular importance whether we refer to the need to avoid scandals as a constraint or a preference. (Is the manager's need to avoid bankruptcy a constraint or a preference?) What matters is that we recognize the elusive but important role of public opinion in the shaping of local justice practices.

Given the constraints, we can turn to the preferences of the relevant actors: first-order authorities, professional allocators, and potential recipients or applicants. Here the explanation must address two issues: mechanisms of preference formation and mechanisms of preference aggregation.

Consider first preference formation. As an approximate first cut we may assert that first-order authorities are moved by global efficiency concerns, second-order allocators by conceptions of local equity and local efficiency, and third-order recipients by their self-interest. These differences are largely an effect of the more-or-less inclusive scopes of the actors. To the extent that first-order authorities consider allocative matters, in addition to their constraint-setting role in fixing the total to be allocated, these authorities tend to look at all the ramifications of a rule system, including incentive effects and how allocation may benefit those other than the recipients. One would expect, for instance, Medicare administrators to have a preference for transplanting patients who will be able to go back to work. Second-order authorities tend to form their preferences with a view to choosing among applicants, partly to ensure fairness and partly to make sure that the good goes to someone who can really benefit from it. To use again the example of transplantation, doctors are partly concerned with patients who suffer from some kind of "medical bad luck" and partly with the need to ensure graft survival. Individual applicants have no systematic tendency or reason to look beyond themselves, which is not to say that they never do.

Consider next preference aggregation. The first question is whether all the actors are likely to have an impact on this process. Specifically, it might seem as if third-order actors—with the obvi-

ous exception of organized labor in the selection of layoff principles—have little influence in shaping the rules. Would-be immigrants, for example, do not even belong to the society in which the rules evolve. Their preferences are, as it were, an idling part of the machinery. As is clear from Mackie's chapter, however, U.S. citizens lobby actively for immigration rights for applicants from their country of national origin. Similarly, even though inner-city black candidates for transplantation lack the resources to promote their interests, doctors do it in their place. To some extent, those who have been accepted at a selective college also see themselves as representing the interests of future applicants. The rule seems to be, therefore, that third-order actors are represented, if at all, by proxy. Although, as I said, the role of labor unions in the determination of layoff principles is a major exception, a form of virtual representation can occur in this arena as well, when nonunionized firms adopt the seniority principle for layoffs to forestall drives for unionization.

The diversity of aggregation mechanisms in the four arenas is such as to suggest that no general theory—similar to the theories of utility maximization and profit maximization to explain the behavior of consumers and firms—will ever be available. It is nevertheless possible to identify two main types of mechanism: bargaining and coalition formation.

A paradigmatic case of bargaining is the determination of layoff principles as part of a collective bargaining agreement. It would be tempting but wrong to expect the relative importance of ability and seniority to reflect the relative bargaining power of management and labor. First, management may have its own reasons for preferring seniority. Second, and more importantly, the union may want to use its bargaining power to get its way on wages or working conditions rather than on layoff procedures. The latter point also applies to congressional logrolling over immigration. It is not simply a question of group A accepting a quota for an immigrant class favored by group B in exchange for the support of group B for a class favored by A. In addition to—or instead of—such exchanges, A might demand the support of B on some issue not related to immigration. When local justice issues are embedded in a larger context and are resolved as part of a package solution, the outcome may be very different from what it would have been

if these issues had been the subject of separate negotiations. In the latter case, the bargaining powers of the parties would be directly reflected in the outcome. One might expect, for instance, to see the adoption of a point system that incorporated the criteria favored by the different parties, with weights corresponding to their bargaining power. Note, however, that point systems can also owe their origin to a consensual process, if all parties agree on the relevant criteria and their weighting.

Coalition building occurs when different parties, for different reasons, agree on some specific allocative procedure. As Mackie demonstrates, alliance formation is very common in immigration policies. Often, one group will support a given principle because of its primary consequences, whereas another will look more to the secondary consequences. Some will support literacy tests because they resonate with the idea of individual merit, and others because it will tend to exclude immigrants from certain regions of the world. Some will support kin-based immigration because it resonates with family values, and others because it has desirable geographical implications. In the arena of transplantation, some doctors will emphasize time on the waiting list because they believe in the inherent fairness of queuing, and others because they want to promote the chances of black patients. As mentioned above, employers no less than employees may agree on the regulation of layoffs by seniority.

In the most general case, we observe coalition building as well as bargaining among coalitions, both within the more-or-less hard constraints of first-order supply and public opinion. Although the case of immigration stands out in its complexity, I believe that all the case studies demonstrate that the processes that shape allocative procedures resist any attempt to reduce them to some overarching general mechanism.

NOTES

1. See notably D. Kahneman, J. Knetsch, and R. Thaler, "Experimental Tests of the Endowment Effect and the Coase Theorem," *Journal of Political Economy* 98 (1990), 1325–1348.

2. In "Allocation by Lot," *Social Science Information* 29 (1990), 745–763, W. Hofstee shows, for instance, that people are more willing to accept the use of lotteries to allocate bads than to allocate goods.

3. Note that this possibility does not depend on the soldiers being so destitute that they could not afford to gamble: that would bring us back to the case of decreasing marginal utility. It is enough to assume that they are so conservative that they did not want to gamble.

4. See A. Hylland and R. Zeckhauser, "The Efficient Allocation of Individuals to Positions," *Journal of Political Economy* 87 (1979), 293–314.

5. Hofstee, "Allocation by Lot."

6. G. Calabresi and P. Bobbitt, *Tragic Choices*, New York: Norton, 1978, p. 19.

7. For a general discussion of the use of lotteries in allocation and decision making, see Chapter 2 of my *Solomonic Judgments*, Cambridge: Cambridge University Press, 1989.

8. H. P. Young, "Equitable Selection of Kidney Recipients," *Journal of American Medical Association* 26 (1989), 2957.

9. In *Local Justice*, pp. 92–93 I argue that in liver and heart transplantations there is a tendency for the need and efficiency criteria to diverge from one another.

10. Although point systems (linear and additive) also define tradeoffs, they are of a special kind. If we consider a two-dimensional process, a point system will assign a sum $ax+by$ to a candidate who scores x and y on the two criteria. The indifference curves will be straight lines. In a multiplicative tradeoff, he might be assigned a number $x^a y^{1-a}$ $(0 < a < 1)$, with the indifference curves curving away from the origin. Assume that the two criteria are level of skill and number of years of seniority. If we want to express the intuition that it takes a larger difference in seniority to compensate for a given difference in skill at low-skill levels than at high-skill levels, we must use the multiplicative rather than the additive formulation. Or to take a more graphic example: We need the multiplicative expression if we believe that in a tradeoff between need and ability as determinants of layoffs it would take a greater difference in ability to offset the difference between no children and one child than to offset the difference between nine and ten children.