

to which men might reasonably be expected to agree. But if the notion of pure procedural justice is to succeed, it is necessary, as I have said, to set up and to administer impartially a just system of surrounding institutions. The reliance on pure procedural justice presupposes that the basic structure satisfies the two principles.

This account of distributive shares is simply an elaboration of the familiar idea that income and wages will be just once a (workably) competitive price system is properly organized and embedded in a just basic structure. These conditions are sufficient. The distribution that results is a case of background justice on the analogy with the outcome of a fair game. But we need to consider whether this conception fits our intuitive ideas of what is just and unjust. In particular we must ask how well it accords with common sense precepts of justice. It seems as if we have ignored these notions altogether. I now wish to show that they can be accounted for and their subordinate place explained.

The problem may be stated in the following way. Mill argued correctly that so long as one remains at the level of common sense precepts, no reconciliation of these maxims of justice is possible. For example, in the case of wages, the precepts to each according to his effort and to each according to his contribution are contrary injunctions taken by themselves. Moreover, if we wish to assign them certain weights, they provide no way to determine how their relative merits are to be ascertained. Thus common sense precepts do not express a determinate theory of just or fair wages.³¹ It does not follow, though, as Mill seems to suppose, that one can find a satisfactory conception only by adopting the utilitarian principle. Some higher principle is indeed necessary; but there are other alternatives than that of utility. It is even possible to elevate one of these precepts, or some combination of them, to the level of a first principle, as when it is said: to each according to his needs.³² From the standpoint of the theory of justice, the two principles of justice define the correct higher criterion. Therefore the problem is to consider whether the common sense precepts of justice would arise in a well-ordered society and how they would receive their appropriate weights.

Consider the case of wages in a perfectly competitive economy surrounded by a just basic structure. Assume that each firm (whether pub-

31. *Utilitarianism*, ch. V, par. 30.

32. This precept is cited by Marx in his *Critique of the Gotha Program*, in *Karl Marx and Frederick Engels, Selected Works* (Moscow, Foreign Languages Publishing House, 1955), vol. II, p. 24.

licly or privately owned) must adjust its rates of pay to the long-run forces of supply and demand. The rates firms pay cannot be so high that they cannot afford paying those rates or so low that a sufficient number will not offer their skills in view of the other opportunities available. In equilibrium the relative attractiveness of different jobs will be equal, all things considered. It is easy, then, to see how the various precepts of justice arise. They simply identify features of jobs that are significant on either the demand or the supply side of the market, or both. A firm's demand for workers is determined by the marginal productivity of labor, that is, by the net value of the contribution of a unit of labor measured by the sale price of the commodities that it produces. The worth of this contribution to the firm rests eventually on market conditions, on what households are willing to pay for various goods. Experience and training, natural ability and special know-how, tend to earn a premium. Firms are willing to pay more to those with these characteristics because their productivity is greater. This fact explains and gives weight to the precept to each according to his contribution, and as special cases, we have the norms to each according to his training, or his experience, and the like. But also, viewed from the supply side, a premium must be paid if those who may later offer their services are to be persuaded to undertake the costs of training and postponement. Similarly jobs which involve uncertain or unstable employment, or which are performed under hazardous and unpleasantly strenuous conditions, tend to receive more pay. Otherwise men cannot be found to fill them. From this circumstance arise such precepts as to each according to his effort, or the risks he bears, and so on. Even when individuals are assumed to be of the same natural ability, these norms will still arise from the requirements of economic activity. Given the aims of productive units and of those seeking work, certain characteristics are singled out as relevant. At any time the wage practices of firms tend to recognize these precepts and, allowing time for adjustment, assign them the weights called for by market conditions.

All of this seems reasonably clear. More important are several further points. For one thing, different conceptions of justice are likely to generate much the same common sense precepts. Thus in a society regulated by the principle of utility all of the above norms would most likely be recognized. So long as the aims of economic agents are sufficiently similar, these precepts are bound to be appealed to, and wage practices will explicitly take them into account. On the other hand, the weights that are assigned to these precepts will not in general be the same. It is here that conceptions of justice diverge. Not only will there be a tendency to

operate wage practices in other ways, but the long-term trend of economic events will almost certainly take another course. When the family of background institutions is governed by distinct conceptions, the market forces to which firms and workers have to adjust will not be the same. A different balance of supply and demand will see to it that the various precepts are balanced differently. Thus the contrast between conceptions of justice does not show up at the level of common sense norms but rather in the relative and changing emphasis that these norms receive over time. In no case can the customary or conventional notion of a fair or just balancing be taken as fundamental, since it will depend upon the principles regulating the background system and the adjustments which they require to current conditions.

An example may clarify this point. Suppose that the basic structure of one society provides for fair equality of opportunity while that of a second society does not. Then in the first society the precept to each according to his contribution in the particular form of each according to his training and education will probably receive much less weight. This is likely to be true even if we suppose, as the facts suggest, that persons have different natural abilities. The reason for this is that with many more persons receiving the benefits of training and education, the supply of qualified individuals in the first society is much greater. When there are no restrictions on entry or imperfections in the capital market for loans (or subsidies) for education, the premium earned by those better endowed is far less. The relative difference in earnings between the more favored and the lowest income class tends to close; and this tendency is even stronger when the difference principle is followed. Thus the precept to each according to his training and education is weighted less in the first than in the second society and the precept to each according to his effort is weighted more. Of course, a conception of justice requires that when social conditions change the appropriate balance of precepts normally changes as well. Over time the consistent application of its principles gradually reshapes the social structure so that market forces also shift, thereby resetting the weight of precepts. There is nothing sacrosanct about the existing balance even if it is correct.

Moreover, it is essential to keep in mind the subordinate place of common sense norms. Doing this is sometimes difficult because they are familiar from everyday life and therefore they are likely to have a prominence in our thinking that their derivative status does not justify. None of these precepts can be plausibly raised to a first principle. Each has presumably arisen in answer to a relevant feature connected with certain

particular institutions, this feature being but one among many and these institutions of a special kind. Adopting one of them as a first principle is sure to lead to the neglect of other things that should be taken into account. And if all or many precepts are treated as first principles, there is no gain in systematic clarity. Common sense precepts are at the wrong level of generality. In order to find suitable first principles one must step behind them. Admittedly some precepts appear quite general at first. For example, the precept to each according to his contribution covers many cases of distribution in a perfectly competitive economy. Accepting the marginal productivity theory of distribution, each factor of production receives an income according to how much it adds to output (assuming private property in the means of production). In this sense, a worker is paid the full value of the results of his labor, no more and no less. Offhand this strikes us as fair. It appeals to a traditional idea of the natural right of property in the fruits of our labor. Therefore to some writers the precept of contribution has seemed satisfactory as a principle of justice.³³

It is easy to see, however, that this is not the case. The marginal product of labor depends upon supply and demand. What an individual contributes by his work varies with the demand of firms for his skills, and this in turn varies with the demand for the products of firms. An individual's contribution is also affected by how many offer similar talents. There is no presumption, then, that following the precept of contribution leads to a just outcome unless the underlying market forces, and the availability of opportunities which they reflect, are appropriately regulated. And this implies, as we have seen, that the basic structure as a whole is just. There is no way, then, to give a proper weight to the precepts of justice except by instituting the surrounding arrangements required by the principles of justice. Some institutions may indeed give a special prominence to certain precepts, in the way for example that a competitive economy emphasizes the precept of contribution. But no inference about the justice of the final distribution can be drawn from viewing the use of any precept in isolation. The overall weighting of the many precepts is done by the whole system. Thus the precept of need is left to the transfer branch; it does not serve as a precept of wages at all. To assess the justice of distributive shares, we must note the total working of the background arrangements, the proportion of income and wealth deriving from each branch.³⁴

33. J. B. Clark is often cited as an example. But see the discussion by J. M. Clark in *The Development of Economic Thought*, ed. H. W. Spiegel (New York, John Wiley and Sons, 1952), pp. 598–612.

34. Thus J. B. Clark's mistake in his reply to Marx is his failure to consider sufficiently the

It may be objected to the preceding account of the common sense precepts and to the idea of pure procedural justice that a perfectly competitive economy can never be realized. Factors of production never in fact receive their marginal products, and under modern conditions anyway industries soon come to be dominated by a few large firms. Competition is at best imperfect and persons receive less than the value of their contribution, and in this sense they are exploited.³⁵ The reply to this is first that in any case the conception of a suitably regulated competitive economy with the appropriate background institutions is an ideal scheme which shows how the two principles of justice might be realized. It serves to illustrate the content of these principles, and brings out one way in which either a private-property economy or a socialist regime can satisfy this conception of justice. Granting that existing conditions always fall short of the ideal assumptions, we have some notion of what is just. Moreover we are in a better position to assess how serious the existing imperfections are and to decide upon the best way to approximate the ideal.

A second point is this. The sense in which persons are exploited by market imperfections is a highly special one: namely, the precept of contribution is violated, and this happens because the price system is no longer efficient. But as we have just seen, this precept is but one among many secondary norms, and what really counts is the workings of the whole system and whether these defects are compensated for elsewhere. Furthermore, since it is essentially the principle of efficiency that is not fulfilled, one might as well say that the whole community is exploited. But in fact the notion of exploitation is out of place here. It implies a deep injustice in the background system and has little to do with the inefficiencies of markets.³⁶

Finally, in view of the subordinate place of the principle of efficiency in justice as fairness, the inevitable deviations from market perfection are not especially worrisome. It is more important that a competitive scheme gives scope for the principle of free association and individual choice of occupation against a background of fair equality of opportunity, and that it allows the decisions of households to regulate the items to be produced for private purposes. A basic prerequisite is the compatibility of economic

question of background justice. See J. M. Clark, *ibid.*, pp. 610f. Marxian exploitation is compatible with perfect competition, since it is the outcome of a certain structure of property relations.

35. For this definition of exploitation, see A. C. Pigou, *The Economics of Welfare*, 4th ed. (London, Macmillan, 1932), pp. 549–551.

36. See Mark Blaug, *Economic Theory in Retrospect*, pp. 434f.

arrangements with the institutions of liberty and free association. Thus if markets are reasonably competitive and open, the notion of pure procedural justice is a feasible one to follow. It seems more practicable than other traditional ideals, being explicitly framed to coordinate the multitude of possible criteria into one coherent and workable conception.

48. LEGITIMATE EXPECTATIONS AND MORAL DESERT

There is a tendency for common sense to suppose that income and wealth, and the good things in life generally, should be distributed according to moral desert. Justice is happiness according to virtue. While it is recognized that this ideal can never be fully carried out, it is the appropriate conception of distributive justice, at least as a *prima facie* principle, and society should try to realize it as circumstances permit.³⁷ Now justice as fairness rejects this conception. Such a principle would not be chosen in the original position. There seems to be no way of defining the requisite criterion in that situation. Moreover, the notion of distribution according to virtue fails to distinguish between moral desert and legitimate expectations. Thus it is true that as persons and groups take part in just arrangements, they acquire claims on one another defined by the publicly recognized rules. Having done various things encouraged by the existing arrangements, they now have certain rights, and just distributive shares honor these claims. A just scheme, then, answers to what men are entitled to; it satisfies their legitimate expectations as founded upon social institutions. But what they are entitled to is not proportional to nor dependent upon their intrinsic worth. The principles of justice that regulate the basic structure and specify the duties and obligations of individuals do not mention moral desert, and there is no tendency for distributive shares to correspond to it.

This contention is borne out by the preceding account of common sense precepts and their role in pure procedural justice (§47). For example, in determining wages a competitive economy gives weight to the precept of contribution. But as we have seen, the extent of one's contribu-

37. See, for example, W. D. Ross, *The Right and the Good* (Oxford, The Clarendon Press, 1930), pp. 21, 26–28, 35, 57f. Similarly, Leibniz in “On the Ultimate Origin of Things” (1697) speaks of the law of justice which “declares that each one [each individual] participate in the perfection of the universe and in a happiness of his own in proportion to his own virtue and to the good will he entertains toward the common good.” *Leibniz*, ed. P. P. Wiener (New York, Charles Scribner’s Sons, 1951), p. 353.

tion (estimated by one's marginal productivity) depends upon supply and demand. Surely a person's moral worth does not vary according to how many offer similar skills, or happen to want what he can produce. No one supposes that when someone's abilities are less in demand or have deteriorated (as in the case of singers) his moral deservingness undergoes a similar shift. All of this is perfectly obvious and has long been agreed to.³⁸ It simply reflects the fact noted before (§17) that it is one of the fixed points of our moral judgments that no one deserves his place in the distribution of natural assets any more than he deserves his initial starting place in society.

Moreover, none of the precepts of justice aims at rewarding virtue. The premiums earned by scarce natural talents, for example, are to cover the costs of training and to encourage the efforts of learning, as well as to direct ability to where it best furthers the common interest. The distributive shares that result do not correlate with moral worth, since the initial endowment of natural assets and the contingencies of their growth and nurture in early life are arbitrary from a moral point of view. The precept which seems intuitively to come closest to rewarding moral desert is that of distribution according to effort, or perhaps better, conscientious effort.³⁹ Once again, however, it seems clear that the effort a person is willing to make is influenced by his natural abilities and skills and the alternatives open to him. The better endowed are more likely, other things equal, to strive conscientiously, and there seems to be no way to discount for their greater good fortune. The idea of rewarding desert is impracticable. And certainly to the extent that the precept of need is emphasized, moral worth is ignored. Nor does the basic structure tend to balance the precepts of justice so as to achieve the requisite correspondence behind the scenes. It is regulated by the two principles of justice which define other aims entirely.

The same conclusion may be reached in another way. In the preceding remarks the notion of moral worth as distinct from a person's claims based upon his legitimate expectations has not been explained. Suppose, then, that we define this notion and show that it has no correlation with distributive shares. We have only to consider a well-ordered society, that is, a society in which institutions are just and this fact is publicly recognized. Its members also have a strong sense of justice, an effective desire to comply with the existing rules and to give one another that to which

38. See F. H. Knight, *The Ethics of Competition* (New York, Harper and Brothers, 1935), pp. 54–57.

39. See Knight, *ibid.*, p. 56n.

they are entitled. In this case we may assume that everyone is of equal moral worth. We have now defined this notion in terms of the sense of justice, the desire to act in accordance with the principles that would be chosen in the original position (§72). But it is evident that understood in this way, the equal moral worth of persons does not entail that distributive shares are equal. Each is to receive what the principles of justice say he is entitled to, and these do not require equality.

The essential point is that the concept of moral worth does not provide a first principle of distributive justice. This is because it cannot be introduced until after the principles of justice and of natural duty and obligation have been acknowledged. Once these principles are on hand, moral worth can be defined as having a sense of justice; and as I shall discuss later (§66), the virtues can be characterized as desires or tendencies to act upon the corresponding principles. Thus the concept of moral worth is secondary to those of right and justice, and it plays no role in the substantive definition of distributive shares. The case is analogous to the relation between the substantive rules of property and the law of robbery and theft. These offenses and the demerits they entail presuppose the institution of property which is established for prior and independent social ends. For a society to organize itself with the aim of rewarding moral desert as a first principle would be like having the institution of property in order to punish thieves. The criterion to each according to his virtue would not, then, be chosen in the original position. Since the parties desire to advance their conceptions of the good, they have no reason for arranging their institutions so that distributive shares are determined by moral desert, even if they could find an antecedent standard for its definition.

In a well-ordered society individuals acquire claims to a share of the social product by doing certain things encouraged by the existing arrangements. The legitimate expectations that arise are the other side, so to speak, of the principle of fairness and the natural duty of justice. For in the way that one has a duty to uphold just arrangements, and an obligation to do one's part when one has accepted a position in them, so a person who has complied with the scheme and done his share has a right to be treated accordingly by others. They are bound to meet his legitimate expectations. Thus when just economic arrangements exist, the claims of individuals are properly settled by reference to the rules and precepts (with their respective weights) which these practices take as relevant. As we have seen, it is incorrect to say that just distributive shares reward individuals according to their moral worth. But what we can say is that, in

the traditional phrase, a just scheme gives each person his due: that is, it allots to each what he is entitled to as defined by the scheme itself. The principles of justice for institutions and individuals establish that doing this is fair.

Now it should be noted that even though a person's claims are regulated by the existing rules, we can still make a distinction between being entitled to something and deserving it in a familiar although nonmoral sense.⁴⁰ To illustrate, after a game one often says that the losing side deserved to win. Here one does not mean that the victors are not entitled to claim the championship, or whatever spoils go to the winner. One means instead that the losing team displayed to a higher degree the skills and qualities that the game calls forth, and the exercise of which gives the sport its appeal. Therefore the losers truly deserved to win but lost out as a result of bad luck, or from other contingencies that caused the contest to miscarry. Similarly even the best economic arrangements will not always lead to the more preferred outcomes. The claims that individuals actually acquire inevitably deviate more or less widely from those that the scheme is designed to allow for. Some persons in favored positions, for example, may not have to a higher degree than others the desired qualities and abilities. All this is evident enough. Its bearing here is that although we can indeed distinguish between the claims that existing arrangements require us to honor, given what individuals have done and how things have turned out, and the claims that would have resulted under more ideal circumstances, none of this implies that distributive shares should be in accordance with moral worth. Even when things happen in the best way, there is still no tendency for distribution and virtue to coincide.

No doubt some may still contend that distributive shares should match moral worth at least to the extent that this is feasible. They may believe that unless those who are better off have superior moral character, their having greater advantages is an affront to our sense of justice. Now this opinion may arise from thinking of distributive justice as somehow the opposite of retributive justice. It is true that in a reasonably well-ordered society those who are punished for violating just laws have normally done something wrong. This is because the purpose of the criminal law is to uphold basic natural duties, those which forbid us to injure other persons in their life and limb, or to deprive them of their liberty and property, and punishments are to serve this end. They are not simply a

40. Here I borrow from Joel Feinberg, *Doing and Deserving* (Princeton, Princeton University Press, 1970), pp. 64f.

scheme of taxes and burdens designed to put a price on certain forms of conduct and in this way to guide men's conduct for mutual advantage. It would be far better if the acts proscribed by penal statutes were never done.⁴¹ Thus a propensity to commit such acts is a mark of bad character, and in a just society legal punishments will only fall upon those who display these faults.

It is clear that the distribution of economic and social advantages is entirely different. These arrangements are not the converse, so to speak, of the criminal law, so that just as the one punishes certain offenses, the other rewards moral worth.⁴² The function of unequal distributive shares is to cover the costs of training and education, to attract individuals to places and associations where they are most needed from a social point of view, and so on. Assuming that everyone accepts the propriety of self- or group-interested motivation duly regulated by a sense of justice, each decides to do those things that best accord with his aims. Variations in wages and income and the perquisites of position are simply to influence these choices so that the end result accords with efficiency and justice. In a well-ordered society there would be no need for the penal law except insofar as the assurance problem made it necessary. The question of criminal justice belongs for the most part to partial compliance theory, whereas the account of distributive shares belongs to strict compliance theory and so to the consideration of the ideal scheme. To think of distributive and retributive justice as converses of one another is completely misleading and suggests a different justification for distributive shares than the one they in fact have.

49. COMPARISON WITH MIXED CONCEPTIONS

While I have often compared the principles of justice with utilitarianism, I have not yet said anything about the mixed conceptions. It will be recalled that these are defined by substituting the standard of utility and other criteria for the second principle of justice (§21). I must now consider these alternatives, especially since many persons may find them more reasonable than the principles of justice which seem at first anyway to impose rather stringent requirements. But it needs to be emphasized straightway that all the mixed conceptions accept the first principle, and

41. See H. L. A. Hart, *The Concept of Law* (Oxford, The Clarendon Press, 1961), p. 39; and Feinberg, *Doing and Deserving*, ch. V.

42. On this point, see Feinberg, *ibid.*, pp. 62, 69n.

therefore recognize the primary place of the equal liberties. None of these views is utilitarian, for even if the principle of utility is substituted for the second principle, or for some part of it, say the difference principle, the conception of utility still has a subordinate place. Thus insofar as one of the chief aims of justice as fairness is to construct an alternative to the classical utilitarian doctrine, this aim is achieved even if we finally accept a mixed conception rather than the two principles of justice. Moreover, given the importance of the first principle, it seems that the essential feature of the contract theory is preserved in these alternatives.

Now it is evident from these remarks that mixed conceptions are much more difficult to argue against than the principle of utility. Many writers who seem to profess a variant of the utilitarian view, even if it is expressed vaguely as the balancing and harmonizing of social interests, clearly presuppose a fixed constitutional system that guarantees the basic freedoms to a certain minimum degree. Thus they actually hold some mixed doctrine, and therefore the strong arguments from liberty cannot be used as before. The main problem, then, is what can still be said in favor of the second principle over that of utility when both are constrained by the principle of equal liberty. We need to examine the reasons for rejecting the standard of utility even in this instance, although it is clear that these reasons will not be as decisive as those for rejecting the classical and average doctrines.

Consider first a mixed conception that is rather close to the principles of justice: namely, the view arising when the principle of average utility constrained by a certain social minimum is substituted for the difference principle, everything else remaining unchanged. Now the difficulty here is the same as that with intuitionist doctrines generally: how is the social minimum to be selected and adjusted to changing circumstances? Anyone using the two principles of justice might also appear to be striking a balance between maximizing average utility and maintaining an appropriate social minimum. If we attended only to his considered judgments and not to his reason for these judgments, his appraisals might be indistinguishable from those of someone following this mixed conception. There is, I assume, sufficient latitude in the determination of the level of the social minimum under varying conditions to bring about this result. How do we know, then, that a person who adopts this mixed view does not in fact rely on the difference principle? To be sure, he is not conscious of invoking it, and indeed he may even repudiate the suggestion that he does so. But it turns out that the level assigned to the required minimum that constrains the principle of average utility leads to precisely the same

consequences that would arise if he were in fact following this criterion. Moreover, he is unable to explain why he chooses the minimum as he does; the best he can say is that he makes the decision that seems most reasonable to him. Now it is going too far to claim that such a person is really using the difference principle, since his judgments may match some other standard. Yet it is true that his conception of justice is still to be identified. The leeway behind the scenes for the determination of the proper minimum leaves the matter unsettled.

Similar things can be said concerning other mixed theories. Thus one might decide to constrain the average principle by setting up some distributional requirement either by itself or in conjunction with some suitably chosen minimum. For example, one might substitute for the difference principle the criterion to maximize the average utility less some fraction (or multiple) of the standard deviation of the resulting distribution.⁴³ Since this deviation is smallest when everyone achieves the same utility, this criterion indicates a greater concern for the less favored than the average principle. Now the intuitionistic features of this view are also clear, for we need to ask how the fraction (or multiple) of the standard deviation is to be selected and how this parameter is to vary with the average itself. Once again the difference principle may stand in the background. This sort of mixed view is on a par with other intuitionistic conceptions that direct us to follow a plurality of ends. For it holds that provided a certain floor is maintained, greater average well-being and a more equal distribution are both desirable ends. One institution is unambiguously preferable to another if it is better on each count.

Different political views, however, balance these ends differently, and we need criteria for determining their relative weights. The fact is that we do not in general agree to very much when we acknowledge ends of this kind. It must be recognized that a fairly detailed weighting of aims is implicit in a reasonably complete conception of justice. In everyday life we often content ourselves with enumerating common sense precepts and objectives of policy, adding that on particular questions we have to balance them in the light of the general facts of the situation. While this is sound practical advice, it does not express an articulated conception of justice. One is being told in effect to exercise one's judgment as best one can within the framework of these ends as guidelines. Only policies preferable on each score are clearly more desirable. By contrast, the

43. For a view of this kind, see Nicholas Rescher, *Distributive Justice* (New York, Bobbs-Merrill, 1966), pp. 35–38.

difference principle is a relatively precise conception, since it ranks all combinations of objectives according to how well they promote the prospects of the least favored.

Thus despite the fact that the difference principle seems offhand to be a somewhat special conception, it may still be the criterion which when adjoined to the other principles of justice stands in the background and controls the weights expressed in our everyday judgments as these would be matched by various mixed principles. Our customary way of relying on intuition guided by lower-order standards may obscure the existence of more basic principles that account for the force of these criteria. Of course, whether the two principles of justice, and especially the difference principle, explicate our judgments of distributive justice can only be decided by developing the consequences of these principles in some detail and noting how far we are prepared to accept the weights to which they lead. Possibly there will be no conflict between these consequences and our considered convictions. Certainly there should be none with those judgments that are fixed points, ones that we seem unwilling to revise under any foreseeable circumstances. Otherwise the two principles are not fully acceptable and some revision has to be made.

But perhaps our everyday views do not entail anything very definite about the problem of balancing competing ends. If so, the main question is whether we can assent to the far more exact specification of our conception of justice which the two principles represent. Provided that certain fixed points are preserved, we have to decide the best way to fill in our conception of justice and to extend it to further cases. The two principles of justice may not so much oppose our intuitive convictions as provide a relatively concrete principle for questions that common sense finds unfamiliar and leaves undecided. Thus while the difference principle strikes us as strange at first, reflection upon its implications when it is suitably circumscribed may convince us that it either accords with our considered judgments, or else projects these convictions to new situations in an acceptable way.

In line with these remarks, we may note that it is a political convention of a democratic society to appeal to the common interest. No political party publicly admits to pressing for legislation to the disadvantage of any recognized social group. But how is this convention to be understood? Surely it is something more than the principle of efficiency, and we cannot assume that government affects everyone's interest equally. Since it is impossible to maximize with respect to more than one point of view, it is natural, given the ethos of a democratic society, to single out

that of the least advantaged and to further their long-term prospects in the best manner consistent with the equal liberties and fair opportunity. It seems that the policies in the justice of which we have the greatest confidence do at least tend in this direction in the sense that this sector of society would be worse off should they be curtailed. These policies are just throughout even if they are not perfectly just. The difference principle can therefore be interpreted as a reasonable extension of the political convention of a democracy once we face up to the necessity of adopting a reasonably complete conception of justice.

In noting that the mixed conceptions have intuitionistic features, I do not mean that this fact is a decisive objection to them. As I have already observed (§7), such combinations of principles are certainly of great practical value. There is no question but that these conceptions identify plausible standards by reference to which policies may be appraised, and given the appropriate background institutions, they may guide us to sound conclusions. For example, a person who accepts the mixed conception to maximize average well-being less some fraction (or multiple) of the standard deviation will presumably favor fair equality of opportunity, for it seems that having more equal chances for all both raises the average (via increases in efficiency) and decreases inequality. In this instance the substitute for the difference principle supports the other part of the second principle. Furthermore it is evident that at some point we cannot avoid relying upon our intuitive judgments. The difficulty with the mixed conceptions is that they may resort to these judgments too soon and fail to define a clear alternative to the difference principle. In the absence of a procedure for assigning the appropriate weights (or parameters), it is possible that the balance is actually determined by the principles of justice, unless of course these principles yield conclusions that we cannot accept. Should this happen, then some mixed conception despite its appeal to intuition may be preferable, especially if its use helps to introduce order and agreement into our considered convictions.

Another consideration favoring the difference principle is the comparative ease with which it can be interpreted and applied. Indeed to some, part of the attractiveness of mixed criteria is that they are a way to avoid the relatively sharp demands of the difference principle. It is fairly straightforward to ascertain what things will advance the interests of the least favored. This group can be identified by its index of primary goods, and policy questions can be settled by asking how the relevant representative man suitably situated would choose. But to the extent that the principle of utility is given a role, the vagueness in the idea of average (or

total) well-being is troublesome. It is necessary to arrive at some estimate of utility functions for different representative persons and to set up an interpersonal correspondence between them, and so on. The problems in doing this are so great and the approximations are so rough that deeply conflicting opinions may seem equally plausible to different persons. Some may claim that the gains of one group outweigh the losses of another, while others may deny it. No one can say what underlying principles account for these differences or how they can be resolved. It is easier for those with the stronger social positions to advance their interests unjustly without being shown to be clearly out of bounds. Of course all this is obvious, and it has always been recognized that ethical principles are vague. Nevertheless they are not all equally imprecise, and the two principles of justice have an advantage in the greater clarity of their demands and in what needs to be done to satisfy them.

It might be thought that the vagueness of the principle of utility can be overcome by a better account of how to measure and to aggregate well-being. I do not wish to stress these much discussed technical problems, since the more important objections to utilitarianism are at another level. But a brief mention of these matters will clarify the contract doctrine. Now there are several ways of establishing an interpersonal measure of utility. One of these (going back at least to Edgeworth) is to suppose that an individual is able to distinguish only a finite number of utility levels.⁴⁴ A person is said to be indifferent between alternatives that belong to the same discrimination level, and the cardinal measure of the utility difference between any two alternatives is defined by the number of distinguishable levels that separate them. The cardinal scale that results is unique, as it must be, up to a positive linear transformation. To set up a measure between persons one might assume that the difference between adjacent levels is the same for all individuals and the same between all levels. With this interpersonal correspondence rule the calculations are extremely simple. In comparing alternatives we ascertain the number of levels between them for each individual and then sum, taking account of the pluses and minuses.

This conception of cardinal utility suffers from well-known difficulties. Leaving aside the obvious practical problems and the fact that the detection of a person's discrimination levels depends upon the alternatives actually available, it seems impossible to justify the assumption that

44. See A. K. Sen, *Collective Choice and Social Welfare* (San Francisco, Holden-Day, 1970), pp. 93f; for Edgeworth, see *Mathematical Psychics* (London, 1888), pp. 7–9, 60f.

the social utility of a shift from one level to another is the same for all individuals. On the one hand, this procedure would weigh identically those changes involving the same number of discriminations that individuals felt differently about, some having stronger feelings than others; while on the other hand, it would count more heavily the changes experienced by those individuals who appear to make more discriminations. Surely it is unsatisfactory to discount the strength of attitudes, and especially to reward so highly the capacity for noting distinctions which may vary systematically with temperament and training.⁴⁵ Indeed, the whole procedure seems arbitrary. It has the merit, however, of illustrating the way in which the principle of utility is likely to contain implicit ethical assumptions in the method chosen for establishing the required measure of utility. The concept of happiness and well-being is not sufficiently determinate, and even to define a suitable cardinal measure we may have to look at the moral theory in which it will be used.

Analogous difficulties arise with the Neumann-Morgenstern definition.⁴⁶ It can be shown that if an individual's choices between risky prospects satisfy certain postulates, then there exist utility numbers corresponding to the alternatives in such a way that his decisions can be interpreted as maximizing expected utility. He chooses as if he were guided by the mathematical expectation of these utility numbers; and these assignments of utility are unique up to a positive linear transformation. Of course, it is not maintained that the individual himself uses an assignment of utilities in making his decisions. These numbers do not guide his choices, nor do they provide a first-person procedure of deliberation. Rather, given that a person's preferences among prospects fulfill certain conditions, the observing mathematician can, theoretically at least, compute numbers that describe these preferences as maximizing expected utility in the sense defined. So far nothing follows about the actual course of reflection, or the criteria, if any, that the individual relies upon; nor is anything implied about what features of the alternatives the utility numbers correspond to or represent.

Now assuming that we can set up a cardinal utility for each person, how is the interpersonal measure to be established? A familiar proposal is the zero-one rule: assign the value zero to the individual's worst possible situation and value one to his best situation. Offhand this seems fair,

45. For these difficulties, see Sen, *ibid.*, pp. 94f; and W. S. Vickrey, "Utility, Strategy, and Social Decision Rules," *Quarterly Journal of Economics*, vol. 74 (1960), pp. 519-522.

46. For an account of this, see Baumol, *Economic Theory and Operations Analysis*, pp. 512-528; and Luce and Raiffa, *Games and Decisions*, pp. 12-38.

perhaps expressing in another way the idea that each is to count for one and no more than one. Yet there are other proposals with comparable symmetry, for example, that which assigns the value zero to the worst alternative and the value one to the sum of the utilities from all alternatives.⁴⁷ Both of these rules seem equally just, since the first postulates equal maximum utility for everyone, the latter equal average utility; but they may lead to different social decisions. Furthermore, these proposals postulate in effect that all individuals have similar capacities for satisfaction, and this seems like an unusual price to pay merely to define an interpersonal measure. These rules clearly determine the concept of well-being in a special way, for the ordinary notion would appear to allow for variations in the sense that a different interpretation of the concept would be equally if not more compatible with common sense. Thus for example the zero-one rule implies that, other things equal, greater social utility results from educating people to have simple desires and to be easily satisfied; and that such persons will generally have the stronger claims. They are pleased with less and so presumably can be brought closer to their highest utility. If one cannot accept these consequences but still wishes to hold the utilitarian view, some other interpersonal measure must be found.

Further, we should observe that while the Neumann-Morgenstern postulates assume that individuals do not enjoy the experience of risk, the actual process of gambling, the resulting measure is nevertheless influenced by attitudes toward uncertainty as defined by the overall probability distribution.⁴⁸ Thus if this definition of utility is used in social decisions, men's feelings about taking chances will affect the criterion of well-being that is to be maximized. Once again we see that the conventions defining interpersonal comparisons have unexpected moral consequences. As before the measure of utility is influenced by contingencies that are arbitrary from a moral point of view. The situation is very different from that of justice as fairness as shown by its Kantian interpretation, the embedding of ideals in its principles, and its reliance upon primary goods for the necessary interpersonal comparisons.

It would appear, then, that the vagueness of the utilitarian principle is not likely to be satisfactorily removed simply by a more precise measure of utility. To the contrary, once the conventions required for interpersonal comparisons are examined, we see that there are various methods for

47. See Sen, *Collective Choice and Social Welfare*, p. 98.

48. See Arrow, *Social Choice and Individual Values*, p. 10; and Sen, *ibid.*, pp. 96f.

defining these comparisons. Yet these methods involve strikingly different assumptions and presumably have very different consequences. It is a moral question which of these definitions and correspondence rules, if any, are appropriate for a conception of justice. This is what is meant, I believe, when it is said that interpersonal comparisons depend upon value judgments. While it is obvious that the acceptance of the principle of utility is a matter for moral theory, it is less evident that the very procedures for measuring well-being raise similar problems. Since there is more than one such procedure, the choice depends upon the use to which the measure is to be put; and this means that ethical considerations will eventually be decisive.

Maine's comments on the standard utilitarian assumptions are apropos here. He suggests that the grounds for these assumptions are clear once we see that they are simply a working rule of legislation, and that this is how Bentham regarded them.⁴⁹ Given a populous and reasonably homogeneous society and an energetic modern legislature, the only principle that can guide legislation on a large scale is the principle of utility. The necessity to neglect differences between persons, even very real ones, leads to the maxim to count all equally, and to the similarity and marginal postulates. Surely the conventions for interpersonal comparisons are to be judged in the same light. The contract doctrine holds that once we see this, we shall also see that the idea of measuring and summing well-being is best abandoned entirely. Viewed from the perspective of the original position, it is not part of a feasible conception of social justice. Instead the two principles of justice are preferable and far simpler to apply. All things considered, there are still reasons for choosing the difference principle, or the second principle as a whole, over that of utility even in the restricted context of a mixed conception.

50. THE PRINCIPLE OF PERFECTION

So far I have said very little about the principle of perfection. But having just considered mixed views, I should now like to examine this conception. There are two variants: in the first it is the sole principle of a teleological theory directing society to arrange institutions and to define the duties and obligations of individuals so as to maximize the

49. These remarks are found in H. S. Maine, *The Early History of Institutions* (London, 1897), pp. 399f.

achievement of human excellence in art, science, and culture. The principle obviously is more demanding the higher the relevant ideal is pitched. The absolute weight that Nietzsche sometimes gives the lives of great men such as Socrates and Goethe is unusual. At places he says that mankind must continually strive to produce great individuals. We give value to our lives by working for the good of the highest specimens.⁵⁰ The second variant found in Aristotle among others has far stronger claims.

This more moderate doctrine is one in which a principle of perfection is accepted as but one standard among several in an intuitionist theory. The principle is to be balanced against others by intuition. The extent to which such a view is perfectionist depends, then, upon the weight given to the claims of excellence and culture. If for example it is maintained that in themselves the achievements of the Greeks in philosophy, science, and art justified the ancient practice of slavery (assuming that this practice was necessary for these achievements), surely the conception is highly perfectionist. The requirements of perfection override the strong claims of liberty. On the other hand, one may use the criterion simply to limit the redistribution of wealth and income under a constitutional regime. In this case it serves as a counterpoise to egalitarian ideas. Thus it may be said that distribution should indeed be more equal if this is essential for meeting the basic needs of those less favored and only diminishes the enjoyments and pleasures of those better off. But the greater happiness of the less fortunate does not in general justify curtailing the expenditures required to preserve cultural values. These forms of life have greater intrinsic worth than the lesser pleasures, however widely the latter are enjoyed. Under normal conditions a certain minimum of social resources must be kept aside to advance the ends of perfection. The only exception is when these claims clash with the demands of the basic needs. Thus given improving circumstances, the principle of perfection acquires an increasing weight relative to a greater satisfaction of desire. No doubt many have accepted perfectionism in this intuitionist form. It

50. See the passages cited in G. A. Morgan, *What Nietzsche Means* (Cambridge, Harvard University Press, 1941), pp. 40–42, 369–376. Particularly striking is Nietzsche's statement: "Mankind must work continually to produce individual great human beings—this and nothing else is the task . . . for the question is this: how can your life, the individual life, retain the highest value, the deepest significance? . . . Only by your living for the good of the rarest and most valuable specimens." *Untimely Meditations: Third Essay: Schopenhauer as Educator*, sec. 6, cited from J. R. Hollingsdale, *Nietzsche: The Man and His Philosophy* (Baton Rouge, Louisiana State University Press, 1965), p. 127.

allows for a range of interpretations and seems to express a far more reasonable view than the strict perfectionist theory.⁵¹

Before considering why the principle of perfection would be rejected, I should like to comment on the relation between the principles of justice and the two kinds of teleological theories, perfectionism and utilitarianism. We may define ideal-regarding principles as those which are not want-regarding principles.⁵² That is, they do not take as the only relevant features the overall amount of want-satisfaction and the way in which it is distributed among persons. Now in terms of this distinction, the principles of justice as well as the principle of perfection (either variant) are ideal-regarding principles. They do not abstract from the aims of desires and hold that satisfactions are of equal value when they are equally intense and pleasurable (the meaning of Bentham's remark that, other things equal, pushpin is as good as poetry). As we have seen (§41), a certain ideal is embedded in the principles of justice, and the fulfillment of desires incompatible with these principles has no value at all. Moreover we are to encourage certain traits of character, especially a sense of justice. Thus the contract doctrine is similar to perfectionism in that it takes into account other things than the net balance of satisfaction and how it is shared. In fact, the principles of justice do not even mention the amount or the distribution of welfare but refer only to the distribution of liberties and the other primary goods. At the same time, they manage to define an ideal of the person without invoking a prior standard of human excellence. The contract view occupies, therefore, an intermediate position between perfectionism and utilitarianism.

Turning to the question whether a perfectionist standard would be adopted, we may consider first the strict perfectionist conception, since here the problems are more obvious. Now in order to have a clear sense, this criterion must provide some way of ranking different kinds of achievements and summing their values. Of course this assessment may not be very exact, but it should be accurate enough to guide the main decisions concerning the basic structure. It is at this point that the princi-

51. For this kind of view, see Bertrand de Jouvenal, *The Ethics of Redistribution* (Cambridge, The University Press, 1951), pp. 53–56, 62–65. See also Hastings Rashdall, *The Theory of Good and Evil* (London, Oxford University Press, 1907), vol. I, pp. 235–243, who argues for the principle that everyone's good is to count for as much as the like good of anyone else, the criteria of perfection being relevant in determining when persons' goods are equal. The capacity for a higher life is a ground for treating men unequally. See pp. 240–242. A similar view is implicit in G. E. Moore, *Principia Ethica*, ch. VI.

52. The definition is from Barry, *Political Argument*, pp. 39f.

ple of perfection gets into difficulty. For while the persons in the original position take no interest in one another's interests, they know that they have (or may have) certain moral and religious interests and other cultural ends which they cannot put in jeopardy. Moreover, they are assumed to be committed to different conceptions of the good and they think that they are entitled to press their claims on one another to further their separate aims. The parties do not share a conception of the good by reference to which the fruition of their powers or even the satisfaction of their desires can be evaluated. They do not have an agreed criterion of perfection that can be used as a principle for choosing between institutions. To acknowledge any such standard would be, in effect, to accept a principle that might lead to a lesser religious or other liberty, if not to a loss of freedom altogether to advance many of one's spiritual ends. If the standard of excellence is reasonably clear, the parties have no way of knowing that their claims may not fall before the higher social goal of maximizing perfection. Thus it seems that the only understanding that the persons in the original position can reach is that everyone should have the greatest equal liberty consistent with a similar liberty for others. They cannot risk their freedom by authorizing a standard of value to define what is to be maximized by a teleological principle of justice. This case is entirely different from that of agreeing to an index of primary goods as a basis of interpersonal comparisons. The index plays a subordinate role in any event, and primary goods are things that men generally want in order to achieve their ends whatever they are. Wanting these goods does not distinguish between one person and another. But of course accepting them for the purpose of an index does not establish a standard of excellence.

It is evident, then, that much the same argument that led to the principle of equal liberty requires the rejection of the principle of perfection. But in making this argument I have not contended that the criteria of excellence lack a rational basis from the standpoint of everyday life. Clearly there are standards in the arts and sciences for appraising creative efforts, at least within particular styles and traditions of thought. Very often it is beyond question that the work of one person is superior to that of another. Indeed, the freedom and well-being of individuals, when measured by the excellence of their activities and works, is vastly different in value. This is true not only of actual performance but of potential performance as well. Comparisons of intrinsic value can obviously be made; and although the standard of perfection is not a principle of justice, judgments of value have an important place in human affairs. They are not necessarily so vague that they must fail as a workable basis for

assigning rights. The argument is rather that in view of their disparate aims the parties have no reason to adopt the principle of perfection given the conditions of the original position.

In order to arrive at the ethic of perfectionism, we should have to attribute to the parties a prior acceptance of some natural duty, say the duty to develop human persons of a certain style and aesthetic grace, and to advance the pursuit of knowledge and the cultivation of the arts. But this assumption would drastically alter the interpretation of the original position. While justice as fairness allows that in a well-ordered society the values of excellence are recognized, the human perfections are to be pursued within the limits of the principle of free association. Persons join together to further their cultural and artistic interests in the same way that they form religious communities. They do not use the coercive apparatus of the state to win for themselves a greater liberty or larger distributive shares on the grounds that their activities are of more intrinsic value. Perfectionism is denied as a political principle. Thus the social resources necessary to support associations dedicated to advancing the arts and sciences and culture generally are to be won as a fair return for services rendered, or from such voluntary contributions as citizens wish to make, all within a regime regulated by the two principles of justice.

On the contract doctrine, then, the equal liberty of citizens does not presuppose that the ends of different persons have the same intrinsic value, nor that their freedom and well-being is of the same worth. It is postulated though that the parties are moral persons, rational individuals with a coherent system of ends and a capacity for a sense of justice. Since they have the requisite defining properties, it would be superfluous to add that the parties are equally moral persons. We can say if we wish that men have equal dignity, meaning by this simply that they all satisfy the conditions of moral personality expressed by the interpretation of the initial contractual situation. And being alike in this respect, they are to be treated as the principles of justice require (§77). But none of this implies that their activities and accomplishments are of equal excellence. To think this is to conflate the notion of moral personality with the various perfections that fall under the concept of value.

I have just noted that persons' being of equal value is not necessary for equal liberty. It should also be observed that their being of equal value is not sufficient either. Sometimes it is said that equality of basic rights follows from the equal capacity of individuals for the higher forms of life; but it is not clear why this should be so. Intrinsic worth is a notion falling under the concept of value, and whether equal liberty or some other

principle is appropriate depends upon the conception of right. Now the criterion of perfection insists that rights in the basic structure be assigned so as to maximize the total of intrinsic value. Presumably the configuration of rights and opportunities enjoyed by individuals affects the degree to which they bring to fruition their latent powers and excellences. But it does not follow that an equal distribution of basic freedoms is the best solution.

The situation resembles that of classical utilitarianism: we require postulates parallel to the standard assumptions. Thus even if the latent abilities of individuals were similar, unless the assignment of rights is governed by a principle of diminishing marginal value (estimated in this case by the criteria for excellence), equal rights would not be insured. Indeed, unless there are bountiful resources, the sum of value might be best increased by very unequal rights and opportunities favoring a few. Doing this is not unjust on the perfectionist view provided that it is necessary to produce a greater sum of human excellence. Now a principle of diminishing marginal value is certainly questionable, although perhaps not as much so as that of equal value. There is little reason to suppose that, in general, rights and resources allocated to encourage and to cultivate highly talented persons contribute less and less to the total beyond some point in the relevant range. To the contrary, this contribution may grow (or stay constant) indefinitely. The principle of perfection provides, then, an insecure foundation for the equal liberties and it would presumably depart widely from the difference principle. The assumptions required for equality seem extremely implausible. To find a firm basis for equal liberty, it seems that we must reject the traditional teleological principles, both perfectionist and utilitarian.

So far I have been discussing perfectionism as a single-principle teleological theory. With this variant the difficulties are most evident. The intuitionistic forms are much more plausible, and when the claims of perfection are weighted with moderation, these views are not easy to argue against. The discrepancy from the two principles of justice is much less. Nevertheless similar problems do arise, for each principle of an intuitionistic view must be chosen, and while the consequences are not likely to be so great in this case, there is as before no basis for acknowledging a principle of perfection as a standard of social justice. In addition, criteria of excellence are imprecise as political principles, and their application to public questions is bound to be unsettled and idiosyncratic, however reasonably they may be invoked and accepted within narrower traditions and communities of thought. It is for this reason, among others,

that justice as fairness requires us to show that modes of conduct interfere with the basic liberties of others or else violate some obligation or natural duty before they can be restricted. For it is when arguments to this conclusion fail that individuals are tempted to appeal to perfectionist criteria in an ad hoc manner. When it is said, for example, that certain kinds of sexual relationships are degrading and shameful, and should be prohibited on this basis, if only for the sake of the individuals in question irrespective of their wishes, it is often because a reasonable case cannot be made in terms of the principles of justice. Instead we fall back on notions of excellence. But in these matters we are likely to be influenced by subtle aesthetic preferences and personal feelings of propriety; and individual, class, and group differences are often sharp and irreconcilable. Since these uncertainties plague perfectionist criteria and jeopardize individual liberty, it seems best to rely entirely on the principles of justice which have a more definite structure.⁵³ Thus even in its intuitionistic form, perfectionism would be rejected as not defining a feasible basis of social justice.

Eventually of course we would have to check whether the consequences of doing without a standard of perfection are acceptable, since offhand it may seem as if justice as fairness does not allow enough scope for ideal-regarding considerations. At this point I can only note that public funds for the arts and sciences may be provided through the exchange branch (§43). In this instance there are no restrictions on the reasons citizens may have for imposing upon themselves the requisite taxes. They may assess the merits of these public goods on perfectionist principles, since the coercive machinery of government is used in this case only to overcome the problems of isolation and assurance, and no one is taxed without his consent. The criterion of excellence does not serve here as a political principle; and so, if it wishes, a well-ordered society can devote a sizable fraction of its resources to expenditures of this kind. But while the claims of culture can be met in this way, the principles of justice do not permit subsidizing universities and institutes, or opera and the theater, on the grounds that these institutions are intrinsically valuable, and that those who engage in them are to be supported

53. Illustrative of this point is the controversy concerning the so-called enforcement of morals, morality often having the narrow sense of sexual morality. See Patrick Devlin, *The Enforcement of Morals* (London, Oxford University Press, 1965), and H. L. A. Hart, *Law, Liberty and Morality* (Stanford, Calif., Stanford University Press, 1963), who take different positions on this issue. For further discussion see Brian Barry, *Political Argument*, pp. 66–69; Ronald Dworkin, “Lord Devlin and the Enforcement of Morals,” *Yale Law Journal*, vol. 75 (1966); and A. R. Louch, “Sins and Crimes,” *Philosophy*, vol. 43 (1968).

even at some significant expense to others who do not receive compensating benefits. Taxation for these purposes can be justified only as promoting directly or indirectly the social conditions that secure the equal liberties and as advancing in an appropriate way the long-term interests of the least advantaged. This seems to authorize those subsidies the justice of which is least in dispute, and so in these cases anyway there is no evident need for a principle of perfection.

With these remarks I conclude the discussion of how the principles of justice apply to institutions. Clearly there are many further questions that should be considered. Other forms of perfectionism are possible and each problem has been examined only briefly. I should emphasize that my intention is solely to indicate that the contract doctrine may serve well enough as an alternative moral conception. When we check its consequences for institutions, it appears to match our common sense convictions more accurately than its traditional rivals, and to extrapolate to previously unsettled cases in a reasonable way.

CHAPTER VI. DUTY AND OBLIGATION

In the two preceding chapters I have discussed the principles of justice for institutions. I now wish to take up the principles of natural duty and obligation that apply to individuals. The first two sections examine the reasons why these principles would be chosen in the original position and their role in making social cooperation stable. A brief discussion of promising and the principle of fidelity is included. For the most part, however, I shall study the implications of these principles for the theory of political duty and obligation within a constitutional framework. This seems the best way to explain their sense and content for the purposes of a theory of justice. In particular, an account of the special case of civil disobedience is sketched which connects it with the problem of majority rule and the grounds for complying with unjust laws. Civil disobedience is contrasted with other forms of noncompliance such as conscientious refusal in order to bring out its special role in stabilizing a nearly just democratic regime.

51. THE ARGUMENTS FOR THE PRINCIPLES OF NATURAL DUTY

In an earlier chapter (§§18–19) I described briefly the principles of natural duty and obligation that apply to individuals. We must now consider why these principles would be chosen in the original position. They are an essential part of a conception of right: they define our institutional ties and how we become bound to one another. The conception of justice as fairness is incomplete until these principles have been accounted for.

From the standpoint of the theory of justice, the most important natural duty is that to support and to further just institutions. This duty has two parts: first, we are to comply with and to do our share in just institutions when they exist and apply to us; and second, we are to assist in the

establishment of just arrangements when they do not exist, at least when this can be done with little cost to ourselves. It follows that if the basic structure of society is just, or as just as it is reasonable to expect in the circumstances, everyone has a natural duty to do what is required of him. Each is bound irrespective of his voluntary acts, performative or otherwise. Now our question is why this principle rather than some other would be adopted. As in the case of institutions, there is no way, let us assume, for the parties to examine all the possible principles that might be proposed. The many possibilities are not clearly defined and among them there may be no best choice. To avoid these difficulties I suppose, as before, that the choice is to be made from a short list of traditional and familiar principles. To expedite matters, I shall mention here only the utilitarian alternative for purposes of clarification and contrast, and very much abbreviate the argument.

Now the choice of principles for individuals is greatly simplified by the fact that the principles for institutions have already been adopted. The feasible alternatives are straightway narrowed down to those that constitute a coherent conception of duty and obligation when taken together with the two principles of justice.¹ This restriction is bound to be particularly important in connection with those principles definitive of our institutional ties. Thus let us suppose that the persons in the original position, having agreed to the two principles of justice, entertain the choice of the principle of utility (either variant) as the standard for the acts of individuals. Even if there is no contradiction in this supposition, the adoption of the utilitarian principle would lead to an incoherent conception of right. The criteria for institutions and individuals do not fit together properly. This is particularly clear in situations in which a person holds a social position regulated by the principles of justice. For example, consider the case of a citizen deciding how to vote between political parties, or the case of a legislator wondering whether to favor a certain statute. The assumption is that these individuals are members of a well-ordered society that has adopted the two principles of justice for institutions and the principle of utility for individuals. How are they to act? As a rational citizen or legislator, a person should, it seems, support that party or favor that statute which best conforms to the two principles of justice. This means that he should vote accordingly, urge others to do likewise, and so on. The existence of institutions involves certain patterns of individual conduct in accordance with publicly recognized rules. The principles for

1. For clarification on this point I am indebted to Allan Gibbard.

institutions have, then, consequences for the acts of persons holding positions in these arrangements. But these persons must also regard their actions as governed by the principle of utility. In this case the rational citizen or legislator should support the party or statute whose victory or enactment is most likely to maximize the net balance (or average) of satisfaction. The choice of the utility principle as the standard for individuals leads to contrary directives. To avoid this conflict it is necessary, at least when the individual holds an institutional position, to choose a principle that matches in some suitable way the two principles of justice. Only in noninstitutional situations is the utilitarian view compatible with the agreements already made. Although the principle of utility may have a place in certain duly circumscribed contexts, it is already excluded as a general account of duty and obligation.

The simplest thing to do, then, is to use the two principles of justice as a part of the conception of right for individuals. We can define the natural duty of justice as that to support and to further the arrangements that satisfy these principles; in this way we arrive at a principle that coheres with the criteria for institutions. There is still the question whether the parties in the original position would not do better if they made the requirement to comply with just institutions conditional upon certain voluntary acts on their part, for example, upon their having accepted the benefits of these arrangements, or upon their having promised or otherwise undertaken to abide by them. Offhand a principle with this kind of condition seems more in accordance with the contract idea with its emphasis upon free consent and the protection of liberty. But, in fact, nothing would be gained by this proviso. In view of the lexical ordering of the two principles, the full complement of the equal liberties is already guaranteed. No further assurances on this score are necessary. Moreover, there is every reason for the parties to secure the stability of just institutions, and the easiest and most direct way to do this is to accept the requirement to support and to comply with them irrespective of one's voluntary acts.

These remarks can be strengthened by recalling our previous discussion of public goods (§42). We noted that in a well-ordered society the public knowledge that citizens generally have an effective sense of justice is a very great social asset. It tends to stabilize just social arrangements. Even when the isolation problem is overcome and fair large-scale schemes already exist for producing public goods, there are two sorts of tendencies leading to instability. From a self-interested point of view each person is tempted to shirk doing his share. He benefits from the public good in any case; and even though the marginal social value of his tax

dollar is much greater than that of the marginal dollar spent on himself, only a small fraction thereof redounds to his advantage. These tendencies arising from self-interest lead to instability of the first kind. But since even with a sense of justice men's compliance with a cooperative venture is predicated on the belief that others will do their part, citizens may be tempted to avoid making a contribution when they believe, or with reason suspect, that others are not making theirs. These tendencies arising from apprehensions about the faithfulness of others lead to instability of the second kind. This instability is particularly likely to be strong when it is dangerous to stick to the rules when others are not. It is this difficulty that plagues disarmament agreements; given circumstances of mutual fear, even just men may be condemned to a condition of permanent hostility. The assurance problem, as we have seen, is to maintain stability by removing temptations of the first kind, and since this is done by public institutions, those of the second kind also disappear, at least in a well-ordered society.

The bearing of these remarks is that basing our political ties upon a principle of obligation would complicate the assurance problem. Citizens would not be bound to even a just constitution unless they have accepted and intend to continue to accept its benefits. Moreover this acceptance must be in some appropriate sense voluntary. But what is this sense? It is difficult to find a plausible account in the case of the political system into which we are born and begin our lives.² And even if such an account could be given, citizens might still wonder about one another whether they were bound, or so regarded themselves. The public conviction that all are tied to just arrangements would be less firm, and a greater reliance on the coercive powers of the sovereign might be necessary to achieve stability. But there is no reason to run these risks. Therefore the parties in the original position do best when they acknowledge the natural duty of justice. Given the value of a public and effective sense of justice, it is important that the principle defining the duties of individuals be simple and clear, and that it insure the stability of just arrangements. I assume, then, that the natural duty of justice would be agreed to rather than a principle of utility, and that from the standpoint of the theory of justice, it is the fundamental requirement for individuals. Principles of obligation, while compatible with it, are not alternatives but rather have a complementary role.

2. I do not accept the whole of Hume's argument in "Of the Original Contract," but I believe it is correct on this count as applied to political duty for citizens generally. See *Essays: Moral, Political, and Literary*, ed. T. H. Green and T. H. Grose (London, 1875), vol. I, pp. 450–452.

There are, of course, other natural duties. A number of these were mentioned earlier (§19). Instead of taking up all of these, it may be more instructive to examine a few cases, beginning with the duty of mutual respect, not previously referred to. This is the duty to show a person the respect which is due to him as a moral being, that is, as a being with a sense of justice and a conception of the good. (In some instances these features may be potentialities only, but I leave this complication aside here; see §77.) Mutual respect is shown in several ways: in our willingness to see the situation of others from their point of view, from the perspective of their conception of their good; and in our being prepared to give reasons for our actions whenever the interests of others are materially affected.³

These two ways correspond to the two aspects of moral personality. When called for, reasons are to be addressed to those concerned; they are to be offered in good faith, in the belief that they are sound reasons as defined by a mutually acceptable conception of justice which takes the good of everyone into account. Thus to respect another as a moral person is to try to understand his aims and interests from his standpoint and to present him with considerations that enable him to accept the constraints on his conduct. Since another wishes, let us suppose, to regulate his actions on the basis of principles to which all could agree, he should be acquainted with the relevant facts which explain the restrictions in this way. Also respect is shown in a willingness to do small favors and courtesies, not because they are of any material value, but because they are an appropriate expression of our awareness of another person's feelings and aspirations. Now the reason why this duty would be acknowledged is that although the parties in the original position take no interest in each other's interests, they know that in society they need to be assured by the esteem of their associates. Their self-respect and their confidence in the value of their own system of ends cannot withstand the indifference much less the contempt of others. Everyone benefits then from living in a society where the duty of mutual respect is honored. The cost to self-interest is minor in comparison with the support for the sense of one's own worth.

Similar reasoning supports the other natural duties. Consider, for example, the duty of mutual aid. Kant suggests, and others have followed him here, that the ground for proposing this duty is that situations may

3. On the notion of respect, see B. A. O. Williams, "The Idea of Equality," *Philosophy, Politics, and Society*, Second Series, ed. Peter Laslett and W. O. Runciman (Oxford, Basil Blackwell, 1962), pp. 118f.

arise in which we will need the help of others, and not to acknowledge this principle is to deprive ourselves of their assistance.⁴ While on particular occasions we are required to do things not in our own interests, we are likely to gain on balance at least over the longer run under normal circumstances. In each single instance the gain to the person who needs help far outweighs the loss of those required to assist him, and assuming that the chances of being the beneficiary are not much smaller than those of being the one who must give aid, the principle is clearly in our interest. But this is not the only argument for the duty of mutual aid, or even the most important one. A sufficient ground for adopting this duty is its pervasive effect on the quality of everyday life. The public knowledge that we are living in a society in which we can depend upon others to come to our assistance in difficult circumstances is itself of great value. It makes little difference that we never, as things turn out, need this assistance and that occasionally we are called on to give it. The balance of gain, narrowly interpreted, may not matter. The primary value of the principle is not measured by the help we actually receive but rather by the sense of confidence and trust in other men's good intentions and the knowledge that they are there if we need them. Indeed, it is only necessary to imagine what a society would be like if it were publicly known that this duty was rejected. Thus while the natural duties are not special cases of a single principle (or so I have assumed), similar reasons no doubt support many of them when one considers the underlying attitudes they represent. Once we try to picture the life of a society in which no one had the slightest desire to act on these duties, we see that it would express an indifference if not disdain for human beings that would make a sense of our own worth impossible. Once again we should note the great importance of publicity effects.

Taking any natural duty by itself, the reasons favoring its adoption are fairly obvious. At least it is evident why these duties are preferable to no similar requirements at all. Although their definition and systematic arrangement are untidy, there is little question that they would be acknowledged. The real difficulty lies in their more detailed specification and with questions of priority: how are these duties to be balanced when they come into conflict, either with each other or with obligations, and with the good that can be achieved by supererogatory actions? There are no

4. See *The Foundations of the Metaphysics of Morals*, Academy edition, vol. 4, p. 423. There is a fuller discussion in *The Metaphysics of Morals*, pt. II (*Tugendlehre*), §30, vol. 6, pp. 451f. Kant notes here that the duty of beneficence (as he calls it) is to be public, that is, a universal law. See §23, note 8.

obvious rules for settling these questions. We cannot say, for example, that duties are lexically prior with respect to supererogatory actions, or to obligations. Nor can we simply invoke the utilitarian principle to set things straight. Requirements for individuals so often oppose each other that this would come to much the same thing as adopting the standard of utility for individuals; and, as we have seen, this is ruled out as leading to an incoherent conception of right. I do not know how this problem is to be settled, or even whether a systematic solution formulating useful and practicable rules is possible. It would seem that the theory for the basic structure is actually simpler. Since we are dealing with a comprehensive scheme of general rules, we can rely on certain procedures of aggregation to cancel out the significance of the complicating elements of particular situations once we take the larger long-term view. Therefore in this book I shall not attempt to discuss these questions of priority in full generality. What I shall do is to examine a few special cases in connection with civil disobedience and conscientious refusal under circumstances of what I shall call a nearly just regime. A satisfactory account of these matters is at best only a start; but it may give us some idea of the kinds of obstacles we face and help to focus our intuitive judgments on the right questions.

It seems appropriate at this juncture to note the familiar distinction between a duty other things equal (a so-called *prima facie* duty), and a duty all things considered. (A parallel distinction holds for obligations.) The formulation of this notion is due to Ross and we may follow him in the main lines.⁵ Thus suppose that the full system of principles that would be chosen in the original position is known. It will contain principles for institutions and individuals, and also, of course, priority rules to weigh these principles when they favor contrary sides in given cases. I further suppose that this full conception of right is finite: it consists of a finite number of principles and priority rules. Although there is a sense in which the number of moral principles (virtues of institutions and individuals) is infinite, or indefinitely large, the full conception is approximately complete: that is, the moral considerations that it fails to cover are for the most part of minor importance. Normally they can be neglected without serious risk of error. The significance of the moral reasons that are not accounted for becomes negligible as the conception of right is more fully worked out. Now adjoined to this full conception (finite yet complete in the sense defined) there is a principle asserting its completeness, and, if we like, also a principle enjoining the agent to perform that

5. See *The Right and the Good* (Oxford, The Clarendon Press, 1930), pp. 18–33, 41f.

action which of all those available to him is reasonably judged the right one (or a best one) in the light of the full system (including the priority rules). Here I imagine that the priority rules are sufficient to resolve conflicts of principles, or at least to guide the way to a correct assignment of weights. Obviously, we are not yet in a position to state these rules for more than a few cases; but since we manage to make these judgments, useful rules exist (unless the intuitionist is correct and there are only descriptions). In any case, the full system directs us to act in the light of all the available relevant reasons (as defined by the principles of the system) as far as we can or should ascertain them.

Now with these stipulations in mind, the phrases “other things equal” and “all things considered” (and other related expressions) indicate the extent to which a judgment is based upon the whole system of principles. A principle taken alone does not express a universal statement which always suffices to establish how we should act when the conditions of the antecedent are fulfilled. Rather, first principles single out relevant features of moral situations such that the exemplification of these features lends support to, provides a reason for making, a certain ethical judgment. The correct judgment depends upon all the relevant features as these are identified and tallied up by the complete conception of right. We claim to have surveyed each of these aspects of the case when we say that something is our duty all things considered; or else we imply that we know (or have reason for believing) how this broader inquiry would turn out. By contrast, in speaking of some requirement as a duty other things equal (a so-called *prima facie* duty), we are indicating that we have so far only taken certain principles into account, that we are making a judgment based on only a subpart of the larger scheme of reasons. I shall not usually signal the distinction between something’s being a person’s duty (or obligation) other things equal, and its being his duty all things considered. Ordinarily the context can be relied upon to gather what is meant.

I believe that these remarks express the essentials of Ross’s concept of *prima facie* duty. The important thing is that such riders as “other things equal” and “all things considered” (and of course “*prima facie*”) are not operators on single sentences, much less on predicates of actions. Rather they express a relation between sentences, a relation between a judgment and its grounds; or as I have put it above, they express a relation between a judgment and a part or the whole of the system of principles that defines its grounds.⁶ This interpretation allows for the point of Ross’s notion. For

6. Here I follow Donald Davidson, “How Is Weakness of the Will Possible?” in *Moral Concepts*,

he introduced it as a way of stating first principles so as to allow the reasons they define to support contrary lines of action in particular cases, as indeed they so often do, without involving us in a contradiction. A traditional doctrine found in Kant, or so Ross believed, is to divide the principles that apply to individuals into two groups, those of perfect and imperfect obligation, and then to rank those of the first kind as lexically prior (to use my term) to those of the second kind. Yet not only is it in general false that imperfect obligations (for example, that of beneficence) should always give way to perfect ones (for example, that of fidelity), but we have no answer if perfect obligations conflict.⁷ Maybe Kant's theory permits a way out; but in any case, he left this problem aside. It is convenient to use Ross's notion for this purpose. These remarks do not, of course, accept his contention that first principles are self-evident. This thesis concerns how these principles are known, and what sort of derivation they admit of. This question is independent of how principles hang together in one system of reasons and lend support to particular judgments of duty and obligation.

52. THE ARGUMENTS FOR THE PRINCIPLE OF FAIRNESS

Whereas there are various principles of natural duty, all obligations arise from the principle of fairness (as defined in §18). It will be recalled that this principle holds that a person is under an obligation to do his part as specified by the rules of an institution whenever he has voluntarily accepted the benefits of the scheme or has taken advantage of the opportunities it offers to advance his interests, provided that this institution is just or fair, that is, satisfies the two principles of justice. As noted before, the intuitive idea here is that when a number of persons engage in a mutually advantageous cooperative venture according to certain rules and thus voluntarily restrict their liberty, those who have submitted to these restrictions have a right to a similar acquiescence on the part of those who have benefited from their submission.⁸ We are not to gain from the cooperative efforts of others without doing our fair share.

ed. Joel Feinberg (London, Oxford University Press, 1969), see p. 109. The whole discussion on pp. 105–110 is relevant here.

7. See *The Right and the Good*, pp. 18f, and *The Foundations of Ethics* (Oxford, The Clarendon Press, 1939), pp. 173, 187.

8. I am indebted here to H. L. A. Hart, "Are There Any Natural Rights?" *Philosophical Review*, vol. 64 (1955), pp. 185f.

It must not be forgotten that the principle of fairness has two parts: one which states how we acquire obligations, namely, by doing various things voluntarily; and another which lays down the condition that the institution in question be just, if not perfectly just, at least as just as it is reasonable to expect under the circumstances. The purpose of this second clause is to insure that obligations arise only if certain background conditions are satisfied. Acquiescence in, or even consent to, clearly unjust institutions does not give rise to obligations. It is generally agreed that extorted promises are void *ab initio*. But similarly, unjust social arrangements are themselves a kind of extortion, even violence, and consent to them does not bind. The reason for this condition is that the parties in the original position would insist upon it.

Before discussing the derivation of the principle, there is a preliminary matter to straighten out. It may be objected that since the principles of natural duty are on hand, there is no necessity for the principle of fairness. Obligations can be accounted for by the natural duty of justice, for when a person avails himself of an institutional set up, its rules then apply to him and the duty of justice holds. Now this contention is, indeed, sound enough. We can, if we like, explain obligations by invoking the duty of justice. It suffices to construe the requisite voluntary acts as acts by which our natural duties are freely extended. Although previously the scheme in question did not apply to us, and we had no duties in regard to it other than that of not seeking to undermine it, we have now by our deeds enlarged the bonds of natural duty. But it seems appropriate to distinguish between those institutions or aspects thereof which must inevitably apply to us since we are born into them and they regulate the full scope of our activity, and those that apply to us because we have freely done certain things as a rational way of advancing our ends. Thus we have a natural duty to comply with the constitution, say, or with the basic laws regulating property (assuming them to be just), whereas we have an obligation to carry out the duties of an office that we have succeeded in winning, or to follow the rules of associations or activities that we have joined. Sometimes it is reasonable to weigh obligations and duties differently when they conflict precisely because they do not arise in the same way. In some cases at least, the fact that obligations are freely assumed is bound to affect their assessment when they conflict with other moral requirements. It is also true that the better-placed members of society are more likely than others to have political obligations as distinct from political duties. For by and large it is these persons who are best able to gain political office and to take advantage of the opportunities offered by

the constitutional system. They are, therefore, bound even more tightly to the scheme of just institutions. To mark this fact, and to emphasize the manner in which many ties are freely assumed, it is useful to have the principle of fairness. This principle should enable us to give a more discriminating account of duty and obligation. The term “obligation” will be reserved, then, for moral requirements that derive from the principle of fairness, while other requirements are called “natural duties.”

Since in later sections the principle of fairness is mentioned in connection with political affairs, I shall discuss here its relation to promises. Now the principle of fidelity is but a special case of the principle of fairness applied to the social practice of promising. The argument for this begins with the observation that promising is an action defined by a public system of rules. These rules are, as in the case of institutions generally, a set of constitutive conventions. Just as the rules of games do, they specify certain activities and define certain actions.⁹ In the case of promising, the basic rule is that governing the use of the words “I promise to do X.” It reads roughly as follows: if one says the words “I promise to do X” in the appropriate circumstances, one is to do X, unless certain excusing conditions obtain. This rule we may think of as the rule of promising; it may be taken as representing the practice as a whole. It is not itself a moral principle but a constitutive convention. In this respect it is on a par with legal rules and statutes, and rules of games; as these do, it exists in a society when it is more or less regularly acted upon.

The way in which the rule of promising specifies the appropriate circumstances and excusing conditions determines whether the practice it represents is just. For example, in order to make a binding promise, one must be fully conscious, in a rational frame of mind, and know the meaning of the operative words, their use in making promises, and so on. Furthermore, these words must be spoken freely or voluntarily, when one is not subject to threats or coercion, and in situations where one has a reasonably fair bargaining position, so to speak. A person is not required to perform if the operative words are uttered while he is asleep, or suffering delusions, or if he was forced to promise, or if pertinent information was deceitfully withheld from him. In general, the circumstances giving rise to a promise and the excusing conditions must be defined so as to preserve the equal liberty of the parties and to make the practice a rational means whereby men can enter into and stabilize cooperative agreements

9. On constitutive rules, see J. R. Searle, *Speech Acts* (Cambridge, The University Press, 1969), pp. 33–42. Promising is discussed in ch. III, esp. pp. 57–62.

for mutual advantage. Unavoidably the many complications here cannot be considered. It must suffice to remark that the principles of justice apply to the practice of promising in the same way that they apply to other institutions. Therefore the restrictions on the appropriate conditions are necessary in order to secure equal liberty. It would be wildly irrational in the original position to agree to be bound by words uttered while asleep, or extorted by force. No doubt it is so irrational that we are inclined to exclude this and other possibilities as inconsistent with the concept (meaning) of promising. However, I shall not regard promising as a practice which is just by definition, since this obscures the distinction between the rule of promising and the obligation derived from the principle of fairness. There are many variations of promising just as there are of the law of contract. Whether the particular practice as it is understood by a person, or group of persons, is just remains to be determined by the principles of justice.

With these remarks as a background, we may introduce two definitions. First, a bona fide promise is one which arises in accordance with the rule of promising when the practice it represents is just. Once a person says the words "I promise to do X" in the appropriate circumstances as defined by a just practice, he has made a bona fide promise. Next, the principle of fidelity is the principle that bona fide promises are to be kept. It is essential, as noted above, to distinguish between the rule of promising and the principle of fidelity. The rule is simply a constitutive convention, whereas the principle of fidelity is a moral principle, a consequence of the principle of fairness. For suppose that a just practice of promising exists. Then in making a promise, that is, in saying the words "I promise to do X" in the appropriate circumstances, one knowingly invokes the rule and accepts the benefits of a just arrangement. There is no obligation to make a promise, let us assume; one is at liberty to do so or not. But since by hypothesis the practice is just, the principle of fairness applies and one is to do as the rule specifies, that is, one is to do X. The obligation to keep a promise is a consequence of the principle of fairness.

I have said that by making a promise one invokes a social practice and accepts the benefits that it makes possible. What are these benefits and how does the practice work? To answer this question, let us assume that the standard reason for making promises is to set up and to stabilize small-scale schemes of cooperation, or a particular pattern of transactions. The role of promises is analogous to that which Hobbes attributed to the sovereign. Just as the sovereign maintains and stabilizes the system

of social cooperation by publicly maintaining an effective schedule of penalties, so men in the absence of coercive arrangements establish and stabilize their private ventures by giving one another their word. Such ventures are often hard to initiate and to maintain. This is especially evident in the case of covenants, that is, in those instances where one person is to perform before the other. For this person may believe that the second party will not do his part, and therefore the scheme never gets going. It is subject to instability of the second kind even though the person to perform later would in fact carry through. Now in these situations there may be no way of assuring the party who is to perform first except by giving him a promise, that is, by putting oneself under an obligation to carry through later. Only in this way can the scheme be made secure so that both can gain from the benefits of their cooperation. The practice of promising exists for precisely this purpose; and so while we normally think of moral requirements as bonds laid upon us, they are sometimes deliberately self-imposed for our advantage. Thus promising is an act done with the public intention of deliberately incurring an obligation the existence of which in the circumstances will further one's ends. We want this obligation to exist and to be known to exist, and we want others to know that we recognize this tie and intend to abide by it. Having, then, availed ourselves of the practice for this reason, we are under an obligation to do as we promised by the principle of fairness.

In this account of how promising (or entering into covenants) is used to initiate and to stabilize forms of cooperation I have largely followed Prichard.¹⁰ His discussion contains all the essential points. I have also assumed, as he does, that each person knows, or at least reasonably believes, that the other has a sense of justice and so a normally effective desire to carry out his bona fide obligations. Without this mutual confidence nothing is accomplished by uttering words. In a well-ordered society, however, this knowledge is present: when its members give promises there is a reciprocal recognition of their intention to put themselves under an obligation and a shared rational belief that this obligation is honored. It is this reciprocal recognition and common knowledge that enables an arrangement to get started and preserves it in being.

There is no need to comment further on the extent to which a common conception of justice (including the principles of fairness and natural duty), and the public awareness of men's willingness to act in accordance

10. See H. A. Prichard, "The Obligation To Keep a Promise," (c. 1940) in *Moral Obligation* (Oxford, The Clarendon Press, 1949), pp. 169–179.

with it, are a great collective asset. I have already noted the many advantages from the standpoint of the assurance problem. It is now equally evident that, having trust and confidence in one another, men can use their public acceptance of these principles enormously to extend the scope and value of mutually advantageous schemes of cooperation. From the standpoint of the original position, then, it is clearly rational for the parties to agree to the principle of fairness. This principle can be used to secure these ventures in ways consistent with freedom of choice and without unnecessarily multiplying moral requirements. At the same time, given the principle of fairness, we see why there should exist the practice of promising as a way of freely establishing an obligation when this is to the mutual advantage of both parties. Such an arrangement is obviously in the common interest. I shall suppose that these considerations are sufficient to argue for the principle of fairness.

Before taking up the question of political duty and obligation, I should note several further points. First of all, as the discussion of promises illustrates, the contract doctrine holds that no moral requirements follow from the existence of institutions alone. Even the rule of promising does not give rise to a moral obligation by itself. To account for fiduciary obligations we must take the principle of fairness as a premise. Thus along with most other ethical theories, justice as fairness holds that natural duties and obligations arise only in virtue of ethical principles. These principles are those that would be chosen in the original position. Together with the relevant facts of the circumstances at hand, it is these criteria that determine our obligations and duties, and single out what count as moral reasons. A (sound) moral reason is a fact which one or more of these principles identifies as supporting a judgment. The correct moral decision is the one most in line with the dictates of this system of principles when it is applied to all the facts it deems to be relevant. Thus the reason identified by one principle may be supported, overridden, or even canceled (brought to naught) by reasons identified by one or more other principles. I assume, though, that out of the totality of facts, presumably in some sense infinite, a finite or surveyable number are selected as those that bear upon any particular case so that the full system enables us to reach a judgment, all things considered.

By contrast, institutional requirements, and those deriving from social practices generally, can be ascertained from the existing rules and how they are to be interpreted. For example, as citizens our legal duties and obligations are settled by what the law is, insofar as it can be ascertained. The norms applying to persons who are players in a game depend upon

the rules of the game. Whether these requirements are connected with moral duties and obligations is a separate question. This is so even if the standards used by judges and others to interpret and to apply the law resemble the principles of right and justice, or are identical with them. It may be, for example, that in a well-ordered society the two principles of justice are used by courts to interpret those parts of the constitution regulating freedom of thought and conscience, and guaranteeing equal protection of the laws.¹¹ Although in this case it is clear that, should the law satisfy its own standards, we are morally bound, other things equal, to comply with it, the questions what the law demands and what justice requires are still distinct. The tendency to conflate the rule of promising and the principle of fidelity (as a special case arising from the principle of fairness) is particularly strong. At first sight they may seem to be the same thing; but one is defined by the existing constitutive conventions, while the other is explained by the principles that would be chosen in the original position. In this way, then, we can distinguish two kinds of norms. The terms “duty” and “obligation” are used in the context of both kinds; but the ambiguities stemming from this usage should be easy enough to resolve.

Finally, I should like to remark that the preceding account of the principle of fidelity answers a question posed by Prichard. He wondered how it is possible, without appealing to a prior general promise, or agreement to keep agreements, to explain the fact that by uttering certain words (by availing oneself of a convention) one becomes bound to do something, particularly when the action whereby one becomes bound is publicly performed with the very intention, which one wants others to recognize, of bringing about this obligation. Or as Prichard expressed it: what is the something implied in there being bona fide agreements which looks much like an agreement to keep agreements and yet which, strictly speaking, cannot be one (since no such agreement has been entered into)?¹² Now the existence of a just practice of promising as a system of public constitutive rules and the principle of fairness suffice for a theory of fiduciary obligations. And neither implies the existence of an actual prior agreement to keep agreements. The adoption of the principle of fairness is purely hypothetical; we only need the fact that this principle would be acknowledged. For the rest, once we assume that a just practice of promising obtains, however it may have come to be established, the

11. On this point, see Ronald Dworkin, “The Model of Rules,” *University of Chicago Law Review*, vol. 35 (1967), esp. pp. 21–29.

12. See “The Obligation To Keep a Promise,” pp. 172, 178f.

principle of fairness is enough to bind those who take advantage of it, given the appropriate conditions already described. Thus what corresponds to the something, which to Prichard looked like a prior agreement but is not, is the just practice of giving one's word in conjunction with the hypothetical agreement on the principle of fairness. Of course, another ethical theory might derive this principle without using the conception of the original position. For the moment I need not maintain that fiduciary ties cannot be explained in some other way. Rather, what I am concerned to show is that even though justice as fairness uses the notion of an original agreement, it is still able to give a satisfactory answer to Prichard's question.

53. THE DUTY TO COMPLY WITH AN UNJUST LAW

There is quite clearly no difficulty in explaining why we are to comply with just laws enacted under a just constitution. In this case the principles of natural duty and the principle of fairness establish the requisite duties and obligations. Citizens generally are bound by the duty of justice, and those who have assumed favored offices and positions, or who have taken advantage of certain opportunities to further their interests, are in addition obligated to do their part by the principle of fairness. The real question is under which circumstances and to what extent we are bound to comply with unjust arrangements. Now it is sometimes said that we are never required to comply in these cases. But this is a mistake. The injustice of a law is not, in general, a sufficient reason for not adhering to it any more than the legal validity of legislation (as defined by the existing constitution) is a sufficient reason for going along with it. When the basic structure of society is reasonably just, as estimated by what the current state of things allows, we are to recognize unjust laws as binding provided that they do not exceed certain limits of injustice. In trying to discern these limits we approach the deeper problem of political duty and obligation. The difficulty here lies in part in the fact that there is a conflict of principles in these cases. Some principles counsel compliance while others direct us the other way. Thus the claims of political duty and obligation must be balanced by a conception of the appropriate priorities.

There is, however, a further problem. As we have seen, the principles of justice (in lexical order) belong to ideal theory (§39). The persons in the original position assume that the principles they acknowledge, what-

ever they are, will be strictly complied with and followed by everyone. Thus the principles of justice that result are those defining a perfectly just society, given favorable conditions. With the presumption of strict compliance, we arrive at a certain ideal conception. When we ask whether and under what circumstances unjust arrangements are to be tolerated, we are faced with a different sort of question. We must ascertain how the ideal conception of justice applies, if indeed it applies at all, to cases where rather than having to make adjustments to natural limitations, we are confronted with injustice. The discussion of these problems belongs to the partial compliance part of nonideal theory. It includes, among other things, the theory of punishment and compensatory justice, just war and conscientious objection, civil disobedience and militant resistance. These are among the central issues of political life, yet so far the conception of justice as fairness does not directly apply to them. Now I shall not attempt to discuss these matters in full generality. In fact, I shall take up but one fragment of partial compliance theory: namely, the problem of civil disobedience and conscientious refusal. And even here I shall assume that the context is one of a state of near justice, that is, one in which the basic structure of society is nearly just, making due allowance for what it is reasonable to expect in the circumstances. An understanding of this admittedly special case may help to clarify the more difficult problems. However, in order to consider civil disobedience and conscientious refusal, we must first discuss several points concerning political duty and obligation.

For one thing, it is evident that our duty or obligation to accept existing arrangements may sometimes be overridden. These requirements depend upon the principles of right, which may justify noncompliance in certain situations, all things considered. Whether noncompliance is justified depends on the extent to which laws and institutions are unjust. Unjust laws do not all stand on a par, and the same is true of policies and institutions. Now there are two ways in which injustice can arise: current arrangements may depart in varying degrees from publicly accepted standards that are more or less just; or these arrangements may conform to a society's conception of justice, or to the view of the dominant class, but this conception itself may be unreasonable, and in many cases clearly unjust. As we have seen, some conceptions of justice are more reasonable than others (see §49). While the two principles of justice and the related principles of natural duty and obligation define the most reasonable view among those on the list, other principles are not unreasonable. Indeed,

some mixed conceptions are certainly adequate enough for many purposes. As a rough rule a conception of justice is reasonable in proportion to the strength of the arguments that can be given for adopting it in the original position. This criterion is, of course, perfectly natural if the original position incorporates the various conditions which are to be imposed on the choice of principles and which lead to a match with our considered judgments.

Although it is easy enough to distinguish these two ways in which existing institutions can be unjust, a workable theory of how they affect our political duty and obligation is another matter. When laws and policies deviate from publicly recognized standards, an appeal to the society's sense of justice is presumably possible to some extent. I argue below that this condition is presupposed in undertaking civil disobedience. If, however, the prevailing conception of justice is not violated, then the situation is very different. The course of action to be followed depends largely on how reasonable the accepted doctrine is and what means are available to change it. Doubtless one can manage to live with a variety of mixed and intuitionistic conceptions, and with utilitarian views when they are not too rigorously interpreted. In other cases, though, as when a society is regulated by principles favoring narrow class interests, one may have no recourse but to oppose the prevailing conception and the institutions it justifies in such ways as promise some success.

Secondly, we must consider the question why, in a situation of near justice anyway, we normally have a duty to comply with unjust, and not simply with just, laws. While some writers have questioned this contention, I believe that most would accept it; only a few think that any deviation from justice, however small, nullifies the duty to comply with existing rules. How, then, is this fact to be accounted for? Since the duty of justice and the principle of fairness presuppose that institutions are just, some further explanation is required.¹³ Now one can answer this question if we postulate a nearly just society in which there exists a viable constitutional regime more or less satisfying the principles of justice. Thus I suppose that for the most part the social system is well-ordered, although not of course perfectly ordered, for in this event the question of

13. I did not note this fact in my essay "Legal Obligation and the Duty of Fair Play" in *Law and Philosophy*, ed. Sidney Hook (New York, New York University Press, 1964). In this section I have tried to make good this defect. The view argued for here is different, however, in that the natural duty of justice is the main principle of political duty for citizens generally, the principle of fairness having a secondary role.

whether to comply with unjust laws and policies would not arise. Under these assumptions, the earlier account of a just constitution as an instance of imperfect procedural justice (§31) provides an answer.

It will be recalled that in the constitutional convention the aim of the parties is to find among the just constitutions (those satisfying the principle of equal liberty) the one most likely to lead to just and effective legislation in view of the general facts about the society in question. The constitution is regarded as a just but imperfect procedure framed as far as the circumstances permit to insure a just outcome. It is imperfect because there is no feasible political process which guarantees that the laws enacted in accordance with it will be just. In political affairs perfect procedural justice cannot be achieved. Moreover, the constitutional process must rely, to a large degree, on some form of voting. I assume for simplicity that a variant of majority rule suitably circumscribed is a practical necessity. Yet majorities (or coalitions of minorities) are bound to make mistakes, if not from a lack of knowledge and judgment, then as a result of partial and self-interested views. Nevertheless, our natural duty to uphold just institutions binds us to comply with unjust laws and policies, or at least not to oppose them by illegal means as long as they do not exceed certain limits of injustice. Being required to support a just constitution, we must go along with one of its essential principles, that of majority rule. In a state of near justice, then, we normally have a duty to comply with unjust laws in virtue of our duty to support a just constitution. Given men as they are, there are many occasions when this duty will come into play.

The contract doctrine naturally leads us to wonder how we could ever consent to a constitutional rule that would require us to comply with laws that we think are unjust. One might ask: how is it possible that when we are free and still without chains, we can rationally accept a procedure that may decide against our own opinion and give effect to that of others?¹⁴ Once we take up the point of view of the constitutional convention, the answer is clear enough. First, among the very limited number of feasible procedures that have any chance of being accepted at all, there are none that would always decide in our favor. And second, consenting to one of these procedures is surely preferable to no agreement at all. The situation is analogous to that of the original position where the parties give up any

14. The metaphor of being free and still without chains is from I. M. D. Little's review of K. J. Arrow, *Social Choice and Individual Values*, in *The Journal of Political Economy*, vol. 60 (1952), p. 431. My remarks here follow Little.

hope of free-rider egoism: this alternative is each person's best (or second best) candidate (leaving aside the constraint of generality), but it is obviously not acceptable to anyone else. Similarly, although at the stage of the constitutional convention the parties are now committed to the principles of justice, they must make some concession to one another to operate a constitutional regime. Even with the best of intentions, their opinions of justice are bound to clash. In choosing a constitution, then, and in adopting some form of majority rule, the parties accept the risks of suffering the defects of one another's knowledge and sense of justice in order to gain the advantages of an effective legislative procedure. There is no other way to manage a democratic regime.

Nevertheless, when they adopt the majority principle the parties agree to put up with unjust laws only on certain conditions. Roughly speaking, in the long run the burden of injustice should be more or less evenly distributed over different groups in society, and the hardship of unjust policies should not weigh too heavily in any particular case. Therefore the duty to comply is problematic for permanent minorities that have suffered from injustice for many years. And certainly we are not required to acquiesce in the denial of our own and others' basic liberties, since this requirement could not have been within the meaning of the duty of justice in the original position, nor consistent with the understanding of the rights of the majority in the constitutional convention. Instead, we submit our conduct to democratic authority only to the extent necessary to share equitably in the inevitable imperfections of a constitutional system. Accepting these hardships is simply recognizing and being willing to work within the limits imposed by the circumstances of human life. In view of this, we have a natural duty of civility not to invoke the faults of social arrangements as a too ready excuse for not complying with them, nor to exploit inevitable loopholes in the rules to advance our interests. The duty of civility imposes a due acceptance of the defects of institutions and a certain restraint in taking advantage of them. Without some recognition of this duty mutual trust and confidence are liable to break down. Thus in a state of near justice at least, there is normally a duty (and for some also the obligation) to comply with unjust laws provided that they do not exceed certain bounds of injustice. This conclusion is not much stronger than that asserting our duty to comply with just laws. It does, however, take us a step further, since it covers a wider range of situations; but more important, it gives some idea of the questions that are to be asked in ascertaining our political duty.

54. THE STATUS OF MAJORITY RULE

It is evident from the preceding remarks that the procedure of majority rule, however it is defined and circumscribed, has a subordinate place as a procedural device. The justification for it rests squarely on the political ends that the constitution is designed to achieve, and therefore on the two principles of justice. I have assumed that some form of majority rule is justified as the best available way of insuring just and effective legislation. It is compatible with equal liberty (§36) and possesses a certain naturalness; for if minority rule is allowed, there is no obvious criterion to select which one is to decide and equality is violated. A fundamental part of the majority principle is that the procedure should satisfy the conditions of background justice. In this case these conditions are those of political liberty—freedom of speech and assembly, freedom to take part in public affairs and to influence by constitutional means the course of legislation—and the guarantee of the fair value of these freedoms. When this background is absent, the first principle of justice is not satisfied; yet even when it is present, there is no assurance that just legislation will be enacted.¹⁵

There is nothing to the view, then, that what the majority wills is right. In fact, none of the traditional conceptions of justice have held this doctrine, maintaining always that the outcome of the voting is subject to political principles. Although in given circumstances it is justified that the majority (suitably defined and circumscribed) has the constitutional right to make law, this does not imply that the laws enacted are just. The dispute of substance about majority rule concerns how it is best defined and whether constitutional constraints are effective and reasonable devices for strengthening the overall balance of justice. These limitations may often be used by entrenched minorities to preserve their illicit advantages. This question is one of political judgment and does not belong to the theory of justice. It suffices to note that while citizens normally submit their conduct to democratic authority, that is, recognize the out-

15. For further discussion of majority rule see Herbert McCloskey, "The Fallacy of Majority Rule," *Journal of Politics*, vol. II (1949), and J. R. Pennock, *Liberal Democracy* (New York, Rinehart, 1950), pp. 112–114, 117f. For some of the attractive features of the majority principle from the standpoint of social choice, see A. K. Sen, *Collective Choice and Social Welfare* (San Francisco, Holden-Day, 1970), pp. 68–70, 71–73, 161–186. One problem with this procedure is that it may allow cyclical majorities. But the primary defect from the point of view of justice is that it permits the violation of liberty. Also see Sen, pp. 79–83, 87–89, where his paradox of liberalism is discussed.

come of a vote as establishing a binding rule, other things equal, they do not submit their judgment to it.

I now wish to take up the place of the principle of majority rule in the ideal procedure that forms a part of the theory of justice. A just constitution is defined as a constitution that would be agreed upon by rational delegates in a constitutional convention who are guided by the two principles of justice. When we justify a constitution, we present considerations to show that it would be adopted under these conditions. Similarly, just laws and policies are those that would be enacted by rational legislators at the legislative stage who are constrained by a just constitution and who are conscientiously trying to follow the principles of justice as their standard. When we criticize laws and policies we try to show that they would not be chosen under this ideal procedure. Now since even rational legislators would often reach different conclusions, there is a necessity for a vote under ideal conditions. The restrictions on information will not guarantee agreement, since the tendencies of the general social facts will often be ambiguous and difficult to assess.

A law or policy is sufficiently just, or at least not unjust, if when we try to imagine how the ideal procedure would work out, we conclude that most persons taking part in this procedure and carrying out its stipulations would favor that law or policy. In the ideal procedure, the decision reached is not a compromise, a bargain struck between opposing parties trying to advance their ends. The legislative discussion must be conceived not as a contest between interests, but as an attempt to find the best policy as defined by the principles of justice. I suppose, then, as part of the theory of justice, that an impartial legislator's only desire is to make the correct decision in this regard, given the general facts known to him. He is to vote solely according to his judgment. The outcome of the vote gives an estimate of what is most in line with the conception of justice.

If we ask how likely it is that the majority opinion will be correct, it is evident that the ideal procedure bears a certain analogy to the statistical problem of pooling the views of a group of experts to arrive at a best judgment.¹⁶ Here the experts are rational legislators able to take an objective perspective because they are impartial. The suggestion goes back to Condorcet that if the likelihood of a correct judgment on the part of the representative legislator is greater than that of an incorrect one, the prob-

16. On this point, see K. J. Arrow, *Social Choice and Individual Values*, 2nd ed. (New York, John Wiley and Sons, 1963), pp. 85f. For the notion of legislative discussion as an objective inquiry and not a contest between interests, see F. H. Knight, *The Ethics of Competition* (New York, Harper and Brothers, 1935), pp. 296, 345–347. In both cases see the footnotes.

ability that the majority vote is correct increases as the likelihood of a correct decision by the representative legislator increases.¹⁷ Thus we might be tempted to suppose that if many rational persons were to try to simulate the conditions of the ideal procedure and conducted their reasoning and discussion accordingly, a large majority anyway would be almost certainly right. This would be a mistake. We must not only be sure that there is a greater chance of a correct than of an incorrect judgment on the part of the representative legislator, but it is also clear that the votes of different persons are not independent. Since their views will be influenced by the course of the discussion, the simpler sorts of probabilistic reasoning do not apply.

Nevertheless, we normally assume that an ideally conducted discussion among many persons is more likely to arrive at the correct conclusion (by a vote if necessary) than the deliberations of any one of them by himself. Why should this be so? In everyday life the exchange of opinion with others checks our partiality and widens our perspective; we are made to see things from their standpoint and the limits of our vision are brought home to us. But in the ideal process the veil of ignorance means that the legislators are already impartial. The benefits from discussion lie in the fact that even representative legislators are limited in knowledge and the ability to reason. No one of them knows everything the others know, or can make all the same inferences that they can draw in concert. Discussion is a way of combining information and enlarging the range of arguments. At least in the course of time, the effects of common deliberation seem bound to improve matters.

Thus we arrive at the problem of trying to formulate an ideal constitution of public deliberation in matters of justice, a set of rules well-designed to bring to bear the greater knowledge and reasoning powers of the group so as best to approximate if not to reach the correct judgment. I shall not, however, pursue this question. The important point here is that the idealized procedure is part of the theory of justice. I have mentioned some of its features in order to elucidate to some degree what is meant by it. The more definite our conception of this procedure as it might be realized under favorable conditions, the more firm the guidance that the four-stage sequence gives to our reflections. For we then have a more precise idea of how laws and policies would be assessed in the light of general facts about society. Often we can make good intuitive sense of the

17. See Duncan Black, *Theory of Committee and Elections*, 2nd ed. (Cambridge, The University Press, 1963), pp. 159–165.

question how deliberations at the legislative stage, when properly conducted, would turn out.

The ideal procedure is further clarified by noting that it stands in contrast to the ideal market process. Thus, granting that the classical assumptions for perfect competition hold, and that there are no external economies or diseconomies, and the like, an efficient economic configuration results. The ideal market is a perfect procedure with respect to efficiency. A peculiarity of the ideal market process, as distinct from the ideal political process conducted by rational and impartial legislators, is that the market achieves an efficient outcome even if everyone pursues his own advantage. Indeed, the presumption is that this is how economic agents normally behave. In buying and selling to maximize satisfaction or profits, households and firms are not giving a judgment as to what is from a social point of view the most efficient economic configuration, given the initial distribution of assets. Rather they are advancing their ends as the rules allow, and any judgment they make is from their own point of view. It is the system as a whole, so to speak, that makes the judgment of efficiency, this judgment being derived from the many separate sources of information provided by the activities of firms and households. The system provides an answer, even though individuals have no opinion of this question, and often do not know what it means.

Thus despite certain resemblances between markets and elections, the ideal market process and the ideal legislative procedure are different in crucial respects. They are designed to achieve distinct ends, the first leading to efficiency, the latter if possible to justice. And while the ideal market is a perfect process with regard to its objective, even the ideal legislature is an imperfect procedure. There seems to be no way to characterize a feasible procedure guaranteed to lead to just legislation. One consequence of this fact is that whereas a citizen may be bound to comply with the policies enacted, other things equal, he is not required to think that these policies are just, and it would be mistaken of him to submit his judgment to the vote. But in a perfect market system, an economic agent, so far as he has any opinion at all, must suppose that the resulting outcome is indeed efficient. Although the household or firm has gotten everything that it wanted, it must concede that, given the initial distribution, an efficient situation has been attained. But the parallel recognition of the outcome of the legislative process concerning questions of justice cannot be demanded, for although, of course, actual constitutions should be designed as far as possible to make the same determinations as the ideal legislative procedure, they are bound in practice to fall short of what

is just. This is not only because, as existing markets do, they fail to conform to their ideal counterpart, but also because this counterpart is that of an imperfect procedure. A just constitution must rely to some extent on citizens and legislators adopting a wider view and exercising good judgment in applying the principles of justice. There seems to be no way of allowing them to take a narrow or group-interested standpoint and then regulating the process so that it leads to a just outcome. So far at least there does not exist a theory of just constitutions as procedures leading to just legislation which corresponds to the theory of competitive markets as procedures resulting in efficiency. And this would seem to imply that the application of economic theory to the actual constitutional process has grave limitations insofar as political conduct is affected by men's sense of justice, as it must be in any viable society, and just legislation is the primary social end (§76). Certainly economic theory does not fit the ideal procedure.¹⁸

These remarks are confirmed by a further contrast. In the ideal market process some weight is given to the relative intensity of desire. A person can spend a greater part of his income on things he wants more of and in this way, together with other buyers, he encourages the use of resources in ways he most prefers. The market allows for finely graded adjustments in answer to the overall balance of preferences and the relative dominance of certain wants. There is nothing corresponding to this in the ideal legislative procedure. Each rational legislator is to vote his opinion as to which laws and policies best conform to principles of justice. No special weight is or should be given to opinions that are held with greater confidence, or to the votes of those who let it be known that their being in the minority will cause them great displeasure (§37). Of course, such a voting rule is conceivable, but there are no grounds for adopting it in the ideal procedure. Even among rational and impartial persons, those with greater confidence in their opinion are not, it seems, more likely to be right. Some may be more sensitive to the complexities of the case than others. In defining the criterion for just legislation one should stress the weight of considered collective judgment arrived at when each person does his best under ideal conditions to apply the correct principles. The

18. For the economic theory of democracy, see J. A. Schumpeter, *Capitalism, Socialism and Democracy*, 3rd ed. (New York, Harper and Brothers, 1950), chs. 21–23, and Anthony Downs, *An Economic Theory of Democracy* (New York, Harper and Brothers, 1957). The pluralist account of democracy, insofar as the rivalry between interests is believed to regulate the political process, is open to similar objection. See R. A. Dahl, *A Preface to Democratic Theory* (Chicago, University of Chicago Press, 1956), and more recently, *Pluralist Democracy in the United States* (Chicago, Rand McNally, 1967).

intensity of desire or the strength of conviction is irrelevant when questions of justice arise.

So much for several differences between the ideal legislative and the ideal market process. I now wish to note the use of the procedure of majority rule as a way of achieving a political settlement. As we have seen, majority rule is adopted as the most feasible way to realize certain ends antecedently defined by the principles of justice. Sometimes however these principles are not clear or definite as to what they require. This is not always because the evidence is complicated and ambiguous, or difficult to survey and assess. The nature of the principles themselves may leave open a range of options rather than singling out any particular alternative. The rate of savings, for example, is specified only within certain limits; the main idea of the just savings principle is to exclude certain extremes. Eventually in applying the difference principle we wish to include in the prospects of the least advantaged the primary good of self-respect; and there are a variety of ways of taking account of this value consistent with the difference principle. How heavily this good and others related to it should count in the index is to be decided in view of the general features of the particular society and by what it is rational for its least favored members to want as seen from the legislative stage. In such cases as these, then, the principles of justice set up a certain range within which the rate of savings or the emphasis given to self-respect should lie. But they do not say where in this range the choice should fall.

Now for these situations the principle of political settlement applies: if the law actually voted is, so far as one can ascertain, within the range of those that could reasonably be favored by rational legislators conscientiously trying to follow the principles of justice, then the decision of the majority is practically authoritative, though not definitive. The situation is one of quasi-pure procedural justice. We must rely on the actual course of discussion at the legislative stage to select a policy within the allowed bounds. These cases are not instances of pure procedural justice because the outcome does not literally define the right result. It is simply that those who disagree with the decision made cannot convincingly establish their point within the framework of the public conception of justice. The question is one that cannot be sharply defined. In practice political parties will no doubt take different stands on these kinds of issues. The aim of constitutional design is to make sure, if possible, that the self-interest of social classes does not so distort the political settlement that it is made outside the permitted limits.

55. THE DEFINITION OF CIVIL DISOBEDIENCE

I now wish to illustrate the content of the principles of natural duty and obligation by sketching a theory of civil disobedience. As I have already indicated, this theory is designed only for the special case of a nearly just society, one that is well-ordered for the most part but in which some serious violations of justice nevertheless do occur. Since I assume that a state of near justice requires a democratic regime, the theory concerns the role and the appropriateness of civil disobedience to legitimately established democratic authority. It does not apply to the other forms of government nor, except incidentally, to other kinds of dissent or resistance. I shall not discuss this mode of protest, along with militant action and resistance, as a tactic for transforming or even overturning an unjust and corrupt system. There is no difficulty about such action in this case. If any means to this end are justified, then surely nonviolent opposition is justified. The problem of civil disobedience, as I shall interpret it, arises only within a more or less just democratic state for those citizens who recognize and accept the legitimacy of the constitution. The difficulty is one of a conflict of duties. At what point does the duty to comply with laws enacted by a legislative majority (or with executive acts supported by such a majority) cease to be binding in view of the right to defend one's liberties and the duty to oppose injustice? This question involves the nature and limits of majority rule. For this reason the problem of civil disobedience is a crucial test case for any theory of the moral basis of democracy.

A constitutional theory of civil disobedience has three parts. First, it defines this kind of dissent and separates it from other forms of opposition to democratic authority. These range from legal demonstrations and infractions of law designed to raise test cases before the courts to militant action and organized resistance. A theory specifies the place of civil disobedience in this spectrum of possibilities. Next, it sets out the grounds of civil disobedience and the conditions under which such action is justified in a (more or less) just democratic regime. And finally, a theory should explain the role of civil disobedience within a constitutional system and account for the appropriateness of this mode of protest within a free society.

Before I take up these matters, a word of caution. We should not expect too much of a theory of civil disobedience, even one framed for special circumstances. Precise principles that straightway decide actual cases are

clearly out of the question. Instead, a useful theory defines a perspective within which the problem of civil disobedience can be approached; it identifies the relevant considerations and helps us to assign them their correct weights in the more important instances. If a theory about these matters appears to us, on reflection, to have cleared our vision and to have made our considered judgments more coherent, then it has been worthwhile. The theory has done what, for the present, one may reasonably expect it to do: namely, to narrow the disparity between the conscientious convictions of those who accept the basic principles of a democratic society.

I shall begin by defining civil disobedience as a public, nonviolent, conscientious yet political act contrary to law usually done with the aim of bringing about a change in the law or policies of the government.¹⁹ By acting in this way one addresses the sense of justice of the majority of the community and declares that in one's considered opinion the principles of social cooperation among free and equal men are not being respected. A preliminary gloss on this definition is that it does not require that the civilly disobedient act breach the same law that is being protested.²⁰ It allows for what some have called indirect as well as direct civil disobedience. And this a definition should do, as there are sometimes strong reasons for not infringing on the law or policy held to be unjust. Instead, one may disobey traffic ordinances or laws of trespass as a way of presenting one's case. Thus, if the government enacts a vague and harsh statute against treason, it would not be appropriate to commit treason as a way of objecting to it, and in any event, the penalty might be far more than one should reasonably be ready to accept. In other cases there is no way to violate the government's policy directly, as when it concerns foreign affairs, or affects another part of the country. A second gloss is that the civilly disobedient act is indeed thought to be contrary to law, at least in the sense that those engaged in it are not simply presenting a test

19. Here I follow H. A. Bedau's definition of civil disobedience. See his "On Civil Disobedience," *Journal of Philosophy*, vol. 58 (1961), pp. 653–661. It should be noted that this definition is narrower than the meaning suggested by Thoreau's essay, as I note in the next section. A statement of a similar view is found in Martin Luther King's "Letter from Birmingham City Jail" (1963), reprinted in H. A. Bedau, ed., *Civil Disobedience* (New York, Pegasus, 1969), pp. 72–89. The theory of civil disobedience in the text tries to set this sort of conception into a wider framework. Some recent writers have also defined civil disobedience more broadly. For example, Howard Zinn, *Disobedience and Democracy* (New York, Random House, 1968), pp. 119f, defines it as "the deliberate, discriminate violation of law for a vital social purpose." I am concerned with a more restricted notion. I do not at all mean to say that only this form of dissent is ever justified in a democratic state.

20. This and the following gloss are from Marshall Cohen, "Civil Disobedience in a Constitutional Democracy," *The Massachusetts Review*, vol. 10 (1969), pp. 224–226, 218–221, respectively.

case for a constitutional decision; they are prepared to oppose the statute even if it should be upheld. To be sure, in a constitutional regime, the courts may finally side with the dissenters and declare the law or policy objected to unconstitutional. It often happens, then, that there is some uncertainty as to whether the dissenters' action will be held illegal or not. But this is merely a complicating element. Those who use civil disobedience to protest unjust laws are not prepared to desist should the courts eventually disagree with them, however pleased they might have been with the opposite decision.

It should also be noted that civil disobedience is a political act not only in the sense that it is addressed to the majority that holds political power, but also because it is an act guided and justified by political principles, that is, by the principles of justice which regulate the constitution and social institutions generally. In justifying civil disobedience one does not appeal to principles of personal morality or to religious doctrines, though these may coincide with and support one's claims; and it goes without saying that civil disobedience cannot be grounded solely on group or self-interest. Instead one invokes the commonly shared conception of justice that underlies the political order. It is assumed that in a reasonably just democratic regime there is a public conception of justice by reference to which citizens regulate their political affairs and interpret the constitution. The persistent and deliberate violation of the basic principles of this conception over any extended period of time, especially the infringement of the fundamental equal liberties, invites either submission or resistance. By engaging in civil disobedience a minority forces the majority to consider whether it wishes to have its actions construed in this way, or whether, in view of the common sense of justice, it wishes to acknowledge the legitimate claims of the minority.

A further point is that civil disobedience is a public act. Not only is it addressed to public principles, it is done in public. It is engaged in openly with fair notice; it is not covert or secretive. One may compare it to public speech, and being a form of address, an expression of profound and conscientious political conviction, it takes place in the public forum. For this reason, among others, civil disobedience is nonviolent. It tries to avoid the use of violence, especially against persons, not from the abhorrence of the use of force in principle, but because it is a final expression of one's case. To engage in violent acts likely to injure and to hurt is incompatible with civil disobedience as a mode of address. Indeed, any interference with the civil liberties of others tends to obscure the civilly disobedient quality of one's act. Sometimes if the appeal fails in its

purpose, forceful resistance may later be entertained. Yet civil disobedience is giving voice to conscientious and deeply held convictions; while it may warn and admonish, it is not itself a threat.

Civil disobedience is nonviolent for another reason. It expresses disobedience to law within the limits of fidelity to law, although it is at the outer edge thereof.²¹ The law is broken, but fidelity to law is expressed by the public and nonviolent nature of the act, by the willingness to accept the legal consequences of one's conduct.²² This fidelity to law helps to establish to the majority that the act is indeed politically conscientious and sincere, and that it is intended to address the public's sense of justice. To be completely open and nonviolent is to give bond of one's sincerity, for it is not easy to convince another that one's acts are conscientious, or even to be sure of this before oneself. No doubt it is possible to imagine a legal system in which conscientious belief that the law is unjust is accepted as a defense for noncompliance. Men of great honesty with full confidence in one another might make such a system work. But as things are, such a scheme would presumably be unstable even in a state of near justice. We must pay a certain price to convince others that our actions have, in our carefully considered view, a sufficient moral basis in the political convictions of the community.

Civil disobedience has been defined so that it falls between legal protest and the raising of test cases on the one side, and conscientious refusal and the various forms of resistance on the other. In this range of possibilities it stands for that form of dissent at the boundary of fidelity to law. Civil disobedience, so understood, is clearly distinct from militant action and obstruction; it is far removed from organized forcible resistance. The militant, for example, is much more deeply opposed to the existing political system. He does not accept it as one which is nearly just or reasonably so; he believes either that it departs widely from its professed principles or that it pursues a mistaken conception of justice altogether. While his action is conscientious in its own terms, he does not appeal to the sense of

21. For a fuller discussion of this point, see Charles Fried, "Moral Causation," *Harvard Law Review*, vol. 77 (1964), pp. 1268f. For clarification below of the notion of militant action, I am indebted to Gerald Loew.

22. Those who define civil disobedience more broadly might not accept this description. See, for example, Zinn, *Disobedience and Democracy*, pp. 27–31, 39, 119f. Moreover he denies that civil disobedience need be nonviolent. Certainly one does not accept the punishment as right, that is, as deserved for an unjustified act. Rather one is willing to undergo the legal consequences for the sake of fidelity to law, which is a different matter. There is room for latitude here in that the definition allows that the charge may be contested in court, should this prove appropriate. But there comes a point beyond which dissent ceases to be civil disobedience as defined here.

justice of the majority (or those having effective political power), since he thinks that their sense of justice is erroneous, or else without effect. Instead, he seeks by well-framed militant acts of disruption and resistance, and the like, to attack the prevalent view of justice or to force a movement in the desired direction. Thus the militant may try to evade the penalty, since he is not prepared to accept the legal consequences of his violation of the law; this would not only be to play into the hands of forces that he believes cannot be trusted, but also to express a recognition of the legitimacy of the constitution to which he is opposed. In this sense militant action is not within the bounds of fidelity to law, but represents a more profound opposition to the legal order. The basic structure is thought to be so unjust or else to depart so widely from its own professed ideals that one must try to prepare the way for radical or even revolutionary change. And this is to be done by trying to arouse the public to an awareness of the fundamental reforms that need to be made. Now in certain circumstances militant action and other kinds of resistance are surely justified. I shall not, however, consider these cases. As I have said, my aim here is the limited one of defining a concept of civil disobedience and understanding its role in a nearly just constitutional regime.

56. THE DEFINITION OF CONSCIENTIOUS REFUSAL

Although I have distinguished civil disobedience from conscientious refusal, I have yet to explain the latter notion. This will now be done. It must be recognized, however, that to separate these two ideas is to give a narrower definition to civil disobedience than is traditional; for it is customary to think of civil disobedience in a broader sense as any noncompliance with law for conscientious reasons, at least when it is not covert and does not involve the use of force. Thoreau's essay is characteristic, if not definitive, of the traditional meaning.²³ The usefulness of the narrower sense will, I believe, be clear once the definition of conscientious refusal is examined.

Conscientious refusal is noncompliance with a more or less direct legal injunction or administrative order. It is refusal since an order is addressed to us and, given the nature of the situation, whether we accede to it is known to the authorities. Typical examples are the refusal of the

23. See Henry David Thoreau, "Civil Disobedience" (1848), reprinted in H. A. Bedau, ed., *Civil Disobedience*, pp. 27–48. For a critical discussion, see Bedau's remarks, pp. 15–26.

early Christians to perform certain acts of piety prescribed by the pagan state, and the refusal of the Jehovah's Witnesses to salute the flag. Other examples are the unwillingness of a pacifist to serve in the armed forces, or of a soldier to obey an order that he thinks is manifestly contrary to the moral law as it applies to war. Or again, in Thoreau's case, the refusal to pay a tax on the grounds that to do so would make him an agent of grave injustice to another. One's action is assumed to be known to the authorities, however much one might wish, in some cases, to conceal it. Where it can be covert, one might speak of conscientious evasion rather than conscientious refusal. Covert infractions of a fugitive slave law are instances of conscientious evasion.²⁴

There are several contrasts between conscientious refusal (or evasion) and civil disobedience. First of all, conscientious refusal is not a form of address appealing to the sense of justice of the majority. To be sure, such acts are not generally secretive or covert, as concealment is often impossible anyway. One simply refuses on conscientious grounds to obey a command or to comply with a legal injunction. One does not invoke the convictions of the community, and in this sense conscientious refusal is not an act in the public forum. Those ready to withhold obedience recognize that there may be no basis for mutual understanding; they do not seek out occasions for disobedience as a way to state their cause. Rather, they bide their time hoping that the necessity to disobey will not arise. They are less optimistic than those undertaking civil disobedience and they may entertain no expectation of changing laws or policies. The situation may allow no time for them to make their case, or again there may not be any chance that the majority will be receptive to their claims.

Conscientious refusal is not necessarily based on political principles; it may be founded on religious or other principles at variance with the constitutional order. Civil disobedience is an appeal to a commonly shared conception of justice, whereas conscientious refusal may have other grounds. For example, assuming that the early Christians would not justify their refusal to comply with the religious customs of the Empire by reasons of justice but simply as being contrary to their religious convictions, their argument would not be political; nor, with similar qualifications, are the views of a pacifist, assuming that wars of self-defense at least are recognized by the conception of justice that underlies a constitutional regime. Conscientious refusal may, however, be grounded on political principles. One may decline to go along with a law thinking that

24. For these distinctions I am indebted to Burton Dreben.

it is so unjust that complying with it is simply out of the question. This would be the case if, say, the law were to enjoin our being the agent of enslaving another, or to require us to submit to a similar fate. These are patent violations of recognized political principles.

It is a difficult matter to find the right course when some men appeal to religious principles in refusing to do actions which, it seems, are required by principles of political justice. Does the pacifist possess an immunity from military service in a just war, assuming that there are such wars? Or is the state permitted to impose certain hardships for noncompliance? There is a temptation to say that the law must always respect the dictates of conscience, but this cannot be right. As we have seen in the case of the intolerant, the legal order must regulate men's pursuit of their religious interests so as to realize the principle of equal liberty; and it may certainly forbid religious practices such as human sacrifice, to take an extreme case. Neither religiosity nor conscientiousness suffices to protect this practice. A theory of justice must work out from its own point of view how to treat those who dissent from it. The aim of a well-ordered society, or one in a state of near justice, is to preserve and strengthen the institutions of justice. If a religion is denied its full expression, it is presumably because it is in violation of the equal liberties of others. In general, the degree of tolerance accorded opposing moral conceptions depends upon the extent to which they can be allowed an equal place within a just system of liberty.

If pacifism is to be treated with respect and not merely tolerated, the explanation must be that it accords reasonably well with the principles of justice, the main exception arising from its attitude toward engaging in a just war (assuming here that in some situations wars of self-defense are justified). The political principles recognized by the community have a certain affinity with the doctrine the pacifist professes. There is a common abhorrence of war and the use of force, and a belief in the equal status of men as moral persons. And given the tendency of nations, particularly great powers, to engage in war unjustifiably and to set in motion the apparatus of the state to suppress dissent, the respect accorded to pacifism serves the purpose of alerting citizens to the wrongs that governments are prone to commit in their name. Even though his views are not altogether sound, the warnings and protests that a pacifist is disposed to express may have the result that on balance the principles of justice are more rather than less secure. Pacifism as a natural departure from the correct doctrine conceivably compensates for the weakness of men in living up to their professions.

It should be noted that there is, of course, in actual situations no sharp distinction between civil disobedience and conscientious refusal. Moreover the same action (or sequence of actions) may have strong elements of both. While there are clear cases of each, the contrast between them is intended as a way of elucidating the interpretation of civil disobedience and its role in a democratic society. Given the nature of this way of acting as a special kind of political appeal, it is not usually justified until other steps have been taken within the legal framework. By contrast this requirement often fails in the obvious cases of legitimate conscientious refusal. In a free society no one may be compelled, as the early Christians were, to perform religious acts in violation of equal liberty, nor must a soldier comply with inherently evil commands while awaiting an appeal to higher authority. These remarks lead up to the question of justification.

57. THE JUSTIFICATION OF CIVIL DISOBEDIENCE

With these various distinctions in mind, I shall consider the circumstances under which civil disobedience is justified. For simplicity I shall limit the discussion to domestic institutions and so to injustices internal to a given society. The somewhat narrow nature of this restriction will be mitigated a bit by taking up the contrasting problem of conscientious refusal in connection with the moral law as it applies to war. I shall begin by setting out what seem to be reasonable conditions for engaging in civil disobedience, and then later connect these conditions more systematically with the place of civil disobedience in a state of near justice. Of course, the conditions enumerated should be taken as presumptions; no doubt there will be situations when they do not hold, and other arguments could be given for civil disobedience.

The first point concerns the kinds of wrongs that are appropriate objects of civil disobedience. Now if one views such disobedience as a political act addressed to the sense of justice of the community, then it seems reasonable, other things equal, to limit it to instances of substantial and clear injustice, and preferably to those which obstruct the path to removing other injustices. For this reason there is a presumption in favor of restricting civil disobedience to serious infringements of the first principle of justice, the principle of equal liberty, and to blatant violations of the second part of the second principle, the principle of fair equality of opportunity. Of course, it is not always easy to tell whether these principles are satisfied. Still, if we think of them as guaranteeing the basic

liberties, it is often clear that these freedoms are not being honored. After all, they impose certain strict requirements that must be visibly expressed in institutions. Thus when certain minorities are denied the right to vote or to hold office, or to own property and to move from place to place, or when certain religious groups are repressed and others denied various opportunities, these injustices may be obvious to all. They are publicly incorporated into the recognized practice, if not the letter, of social arrangements. The establishment of these wrongs does not presuppose an informed examination of institutional effects.

By contrast infractions of the difference principle are more difficult to ascertain. There is usually a wide range of conflicting yet rational opinion as to whether this principle is satisfied. The reason for this is that it applies primarily to economic and social institutions and policies. A choice among these depends upon theoretical and speculative beliefs as well as upon a wealth of statistical and other information, all of this seasoned with shrewd judgment and plain hunch. In view of the complexities of these questions, it is difficult to check the influence of self-interest and prejudice; and even if we can do this in our own case, it is another matter to convince others of our good faith. Thus unless tax laws, for example, are clearly designed to attack or to abridge a basic equal liberty, they should not normally be protested by civil disobedience. The appeal to the public's conception of justice is not sufficiently clear. The resolution of these issues is best left to the political process provided that the requisite equal liberties are secure. In this case a reasonable compromise can presumably be reached. The violation of the principle of equal liberty is, then, the more appropriate object of civil disobedience. This principle defines the common status of equal citizenship in a constitutional regime and lies at the basis of the political order. When it is fully honored the presumption is that other injustices, while possibly persistent and significant, will not get out of hand.

A further condition for civil disobedience is the following. We may suppose that the normal appeals to the political majority have already been made in good faith and that they have failed. The legal means of redress have proved of no avail. Thus, for example, the existing political parties have shown themselves indifferent to the claims of the minority or have proved unwilling to accommodate them. Attempts to have the laws repealed have been ignored and legal protests and demonstrations have had no success. Since civil disobedience is a last resort, we should be sure that it is necessary. Note that it has not been said, however, that legal means have been exhausted. At any rate, further normal appeals can be

repeated; free speech is always possible. But if past actions have shown the majority immovable or apathetic, further attempts may reasonably be thought fruitless, and a second condition for justified civil disobedience is met. This condition is, however, a presumption. Some cases may be so extreme that there may be no duty to use first only legal means of political opposition. If, for example, the legislature were to enact some outrageous violation of equal liberty, say by forbidding the religion of a weak and defenseless minority, we surely could not expect that sect to oppose the law by normal political procedures. Indeed, even civil disobedience might be much too mild, the majority having already convicted itself of wantonly unjust and overtly hostile aims.

The third and last condition I shall discuss can be rather complicated. It arises from the fact that while the two preceding conditions are often sufficient to justify civil disobedience, this is not always the case. In certain circumstances the natural duty of justice may require a certain restraint. We can see this as follows. If a certain minority is justified in engaging in civil disobedience, then any other minority in relevantly similar circumstances is likewise justified. Using the two previous conditions as the criteria of relevantly similar circumstances, we can say that, other things equal, two minorities are similarly justified in resorting to civil disobedience if they have suffered for the same length of time from the same degree of injustice and if their equally sincere and normal political appeals have likewise been to no avail. It is conceivable, however, even if it is unlikely, that there should be many groups with an equally sound case (in the sense just defined) for being civilly disobedient; but that, if they were all to act in this way, serious disorder would follow which might well undermine the efficacy of the just constitution. I assume here that there is a limit on the extent to which civil disobedience can be engaged in without leading to a breakdown in the respect for law and the constitution, thereby setting in motion consequences unfortunate for all. There is also an upper bound on the ability of the public forum to handle such forms of dissent; the appeal that civilly disobedient groups wish to make can be distorted and their intention to appeal to the sense of justice of the majority lost sight of. For one or both of these reasons, the effectiveness of civil disobedience as a form of protest declines beyond a certain point; and those contemplating it must consider these constraints.

The ideal solution from a theoretical point of view calls for a cooperative political alliance of the minorities to regulate the overall level of dissent. For consider the nature of the situation: there are many groups each equally entitled to engage in civil disobedience. Moreover they all

wish to exercise this right, equally strong in each case; but if they all do so, lasting injury may result to the just constitution to which they each recognize a natural duty of justice. Now when there are many equally strong claims which if taken together exceed what can be granted, some fair plan should be adopted so that all are equitably considered. In simple cases of claims to goods that are indivisible and fixed in number, some rotation or lottery scheme may be the fair solution when the number of equally valid claims is too great.²⁵ But this sort of device is completely unrealistic here. What seems called for is a political understanding among the minorities suffering from injustice. They can meet their duty to democratic institutions by coordinating their actions so that while each has an opportunity to exercise its right, the limits on the degree of civil disobedience are not exceeded. To be sure, an alliance of this sort is difficult to arrange; but with perceptive leadership, it does not appear impossible.

Certainly the situation envisaged is a special one, and it is quite possible that these sorts of considerations will not be a bar to justified civil disobedience. There are not likely to be many groups similarly entitled to engage in this form of dissent while at the same time recognizing a duty to a just constitution. One should note, however, that an injured minority is tempted to believe its claims as strong as those of any other; and therefore even if the reasons that different groups have for engaging in civil disobedience are not equally compelling, it is often wise to presume that their claims are indistinguishable. Adopting this maxim, the circumstance imagined seems more likely to happen. This kind of case is also instructive in showing that the exercise of the right to dissent, like the exercise of rights generally, is sometimes limited by others having the very same right. Everyone's exercising this right would have deleterious consequences for all, and some equitable plan is called for.

Suppose that in the light of the three conditions, one has a right to appeal one's case by civil disobedience. The injustice one protests is a clear violation of the liberties of equal citizenship, or of equality of opportunity, this violation having been more or less deliberate over an extended period of time in the face of normal political opposition, and

25. For a discussion of the conditions when some fair arrangement is called for, see Kurt Baier, *The Moral Point of View* (Ithaca, N.Y., Cornell University Press, 1958), pp. 207–213; and David Lyons, *Forms and Limits of Utilitarianism* (Oxford, The Clarendon Press, 1965), pp. 160–176. Lyons gives an example of a fair rotation scheme and he also observes that (waiving costs of setting them up) such fair procedures may be reasonably efficient. See pp. 169–171. I accept the conclusions of his account, including his contention that the notion of fairness cannot be explained by assimilating it to utility, pp. 176f. The earlier discussion by C. D. Broad, "On the Function of False Hypotheses in Ethics," *International Journal of Ethics*, vol. 26 (1916), esp. pp. 385–390, should also be noted here.

any complications raised by the question of fairness are met. These conditions are not exhaustive; some allowance still has to be made for the possibility of injury to third parties, to the innocent, so to speak. But I assume that they cover the main points. There is still, of course, the question whether it is wise or prudent to exercise this right. Having established the right, one is now free, as one is not before, to let these matters decide the issue. We may be acting within our rights but nevertheless unwisely if our conduct only serves to provoke the harsh retaliation of the majority. To be sure, in a state of near justice, vindictive repression of legitimate dissent is unlikely, but it is important that the action be properly designed to make an effective appeal to the wider community. Since civil disobedience is a mode of address taking place in the public forum, care must be taken to see that it is understood. Thus the exercise of the right to civil disobedience should, like any other right, be rationally framed to advance one's ends or the ends of those one wishes to assist. The theory of justice has nothing specific to say about these practical considerations. In any event questions of strategy and tactics depend upon the circumstances of each case. But the theory of justice should say at what point these matters are properly raised.

Now in this account of the justification of civil disobedience I have not mentioned the principle of fairness. The natural duty of justice is the primary basis of our political ties to a constitutional regime. As we noted before (§52), only the more favored members of society are likely to have a clear political obligation as opposed to a political duty. They are better situated to win public office and find it easier to take advantage of the political system. And having done so, they have acquired an obligation owed to citizens generally to uphold the just constitution. But members of subjected minorities, say, who have a strong case for civil disobedience will not generally have a political obligation of this sort. This does not mean, however, that the principle of fairness will not give rise to important obligations in their case.²⁶ For not only do many of the requirements of private life derive from this principle, but it comes into force when persons or groups come together for common political purposes. Just as we acquire obligations to others with whom we have joined in various private associations, those who engage in political action assume obligatory ties to one another. Thus while the political obligation of dissenters to citizens generally is problematical, bonds of loyalty and fidelity still

26. For a discussion of these obligations, see Michael Walzer, *Obligations: Essays on Disobedience, War, and Citizenship* (Cambridge, Harvard University Press, 1970), ch. III.

develop between them as they seek to advance their cause. In general, free association under a just constitution gives rise to obligations provided that the ends of the group are legitimate and its arrangements fair. This is as true of political as it is of other associations. These obligations are of immense significance and they constrain in many ways what individuals can do. But they are distinct from an obligation to comply with a just constitution. My discussion of civil disobedience is in terms of the duty of justice alone; a fuller view would note the place of these other requirements.

58. THE JUSTIFICATION OF CONSCIENTIOUS REFUSAL

In examining the justification of civil disobedience I assumed for simplicity that the laws and policies protested concerned domestic affairs. It is natural to ask how the theory of political duty applies to foreign policy. Now in order to do this it is necessary to extend the theory of justice to the law of nations. I shall try to indicate how this can be done. To fix ideas I shall consider briefly the justification of conscientious refusal to engage in certain acts of war, or to serve in the armed forces. I assume that this refusal is based upon political and not upon religious or other principles; that is, the principles cited by way of justification are those of the conception of justice underlying the constitution. Our problem, then, is to relate the just political principles regulating the conduct of states to the contract doctrine and to explain the moral basis of the law of nations from this point of view.

Let us assume that we have already derived the principles of justice as these apply to societies as units and to the basic structure. Imagine also that the various principles of natural duty and of obligation that apply to individuals have been adopted. Thus the persons in the original position have agreed to the principles of right as these apply to their own society and to themselves as members of it. Now at this point one may extend the interpretation of the original position and think of the parties as representatives of different nations who must choose together the fundamental principles to adjudicate conflicting claims among states. Following out the conception of the initial situation, I assume that these representatives are deprived of various kinds of information. While they know that they represent different nations each living under the normal circumstances of human life, they know nothing about the particular circumstances of their own society, its power and strength in comparison with other nations, nor

do they know their place in their own society. Once again the contracting parties, in this case representatives of states, are allowed only enough knowledge to make a rational choice to protect their interests but not so much that the more fortunate among them can take advantage of their special situation. This original position is fair between nations; it nullifies the contingencies and biases of historical fate. Justice between states is determined by the principles that would be chosen in the original position so interpreted. These principles are political principles, for they govern public policies toward other nations.

I can give only an indication of the principles that would be acknowledged. But, in any case, there would be no surprises, since the principles chosen would, I think, be familiar ones.²⁷ The basic principle of the law of nations is a principle of equality. Independent peoples organized as states have certain fundamental equal rights. This principle is analogous to the equal rights of citizens in a constitutional regime. One consequence of this equality of nations is the principle of self-determination, the right of a people to settle its own affairs without the intervention of foreign powers. Another consequence is the right of self-defense against attack, including the right to form defensive alliances to protect this right. A further principle is that treaties are to be kept, provided they are consistent with the other principles governing the relations of states. Thus treaties for self-defense, suitably interpreted, would be binding, but agreements to cooperate in an unjustified attack are void *ab initio*.

These principles define when a nation has a just cause in war or, in the traditional phrase, its *jus ad bellum*. But there are also principles regulating the means that a nation may use to wage war, its *jus in bello*.²⁸ Even in a just war certain forms of violence are strictly inadmissible; and where a country's right to war is questionable and uncertain, the constraints on the means it can use are all the more severe. Acts permissible in a war of legitimate self-defense, when these are necessary, may be flatly excluded in a more doubtful situation. The aim of war is a just peace, and therefore the means employed must not destroy the possibility of peace or encourage a contempt for human life that puts the safety of ourselves and of mankind in jeopardy. The conduct of war is to be constrained and adjusted to this end. The representatives of states would recognize that their

27. See. J. L. Brierly, *The Law of Nations*, 6th ed. (Oxford, The Clarendon Press, 1963), esp. chs. IV–V. This work contains all that we need here.

28. For a recent discussion, see Paul Ramsey, *War and the Christian Conscience* (Durham, N.C., The Duke University Press, 1961); and also R. B. Potter, *War and Moral Discourse* (Richmond, Va., John Knox Press, 1969). The latter contains a useful bibliographical essay, pp. 87–123.

national interest, as seen from the original position, is best served by acknowledging these limits on the means of war. This is because the national interest of a just state is defined by the principles of justice that have already been acknowledged. Therefore such a nation will aim above all to maintain and to preserve its just institutions and the conditions that make them possible. It is not moved by the desire for world power or national glory; nor does it wage war for purposes of economic gain or the acquisition of territory. These ends are contrary to the conception of justice that defines a society's legitimate interest, however prevalent they have been in the actual conduct of states. Granting these presumptions, then, it seems reasonable to suppose that the traditional prohibitions incorporating the natural duties that protect human life would be chosen.

Now if conscientious refusal in time of war appeals to these principles, it is founded upon a political conception, and not necessarily upon religious or other notions. While this form of denial may not be a political act, since it does not take place in the public forum, it is based upon the same theory of justice that underlies the constitution and guides its interpretation. Moreover, the legal order itself presumably recognizes in the form of treaties the validity of at least some of these principles of the law of nations. Therefore if a soldier is ordered to engage in certain illicit acts of war, he may refuse if he reasonably and conscientiously believes that the principles applying to the conduct of war are plainly violated. He can maintain that, all things considered, his natural duty not to be made the agent of grave injustice and evil to another outweighs his duty to obey. I cannot discuss here what constitutes a manifest violation of these principles. It must suffice to note that certain clear cases are perfectly familiar. The essential point is that the justification cites political principles that can be accounted for by the contract doctrine. The theory of justice can be developed, I believe, to cover this case.

A somewhat different question is whether one should join the armed forces at all during some particular war. The answer is likely to depend upon the aim of the war as well as upon its conduct. In order to make the situation definite, let us suppose that conscription is in force and that the individual has to consider whether to comply with his legal duty to enter military service. Now I shall assume that since conscription is a drastic interference with the basic liberties of equal citizenship, it cannot be justified by any needs less compelling than those of national security.²⁹ In a well-ordered society (or in one nearly just) these needs are determined

29. I am indebted to R. G. Albritton for clarification on this and other matters in this paragraph.

by the end of preserving just institutions. Conscription is permissible only if it is demanded for the defense of liberty itself, including here not only the liberties of the citizens of the society in question, but also those of persons in other societies as well. Therefore if a conscript army is less likely to be an instrument of unjustified foreign adventures, it may be justified on this basis alone despite the fact that conscription infringes upon the equal liberties of citizens. But in any case, the priority of liberty (assuming serial order to obtain) requires that conscription be used only as the security of liberty necessitates. Viewed from the standpoint of the legislature (the appropriate stage for this question), the mechanism of the draft can be defended only on this ground. Citizens agree to this arrangement as a fair way of sharing in the burdens of national defense. To be sure, the hazards that any particular individual must face are in part the result of accident and historical happenstance. But in a well-ordered society anyway, these evils arise externally, that is, from unjustified attacks from the outside. It is impossible for just institutions to eliminate these hardships entirely. The most that they can do is to try to make sure that the risks of suffering from these imposed misfortunes are more or less evenly shared by all members of society over the course of their life, and that there is no avoidable class bias in selecting those who are called for duty.

Imagine, then, a democratic society in which conscription exists. A person may conscientiously refuse to comply with his duty to enter the armed forces during a particular war on the ground that the aims of the conflict are unjust. It may be that the objective sought by war is economic advantage or national power. The basic liberty of citizens cannot be interfered with to achieve these ends. And, of course, it is unjust and contrary to the law of nations to attack the liberty of other societies for these reasons. Therefore a just cause for war does not exist, and this may be sufficiently evident that a citizen is justified in refusing to discharge his legal duty. Both the law of nations and the principles of justice for his own society uphold him in this claim. There is sometimes a further ground for refusal based not on the aim of the war but upon its conduct. A citizen may maintain that once it is clear that the moral law of war is being regularly violated, he has a right to decline military service on the ground that he is entitled to insure that he honors his natural duty. Once he is in the armed forces, and in a situation where he finds himself ordered to do acts contrary to the moral law of war, he may not be able to resist the demand to obey. Actually, if the aims of the conflict are sufficiently dubious and the likelihood of receiving flagrantly unjust com-

mands is sufficiently great, one may have a duty and not only a right to refuse. Indeed, the conduct and aims of states in waging war, especially large and powerful ones, are in some circumstances so likely to be unjust that one is forced to conclude that in the foreseeable future one must abjure military service altogether. So understood a form of contingent pacifism may be a perfectly reasonable position: the possibility of a just war is conceded but not under present circumstances.³⁰

What is needed, then, is not a general pacifism but a discriminating conscientious refusal to engage in war in certain circumstances. States have not been loath to recognize pacifism and to grant it a special status. The refusal to take part in all war under any conditions is an unworldly view bound to remain a sectarian doctrine. It no more challenges the state's authority than the celibacy of priests challenges the sanctity of marriage.³¹ By exempting pacifists from its prescriptions the state may even seem to display a certain magnanimity. But conscientious refusal based upon the principles of justice between peoples as they apply to particular conflicts is another matter. For such refusal is an affront to the government's pretensions, and when it becomes widespread, the continuation of an unjust war may prove impossible. Given the often predatory aims of state power, and the tendency of men to defer to their government's decision to wage war, a general willingness to resist the state's claims is all the more necessary.

59. THE ROLE OF CIVIL DISOBEDIENCE

The third aim of a theory of civil disobedience is to explain its role within a constitutional system and to account for its connection with a democratic polity. As always, I assume that the society in question is one that is nearly just; and this implies that it has some form of democratic government, although serious injustices may nevertheless exist. In such a society I assume that the principles of justice are for the most part publicly recognized as the fundamental terms of willing cooperation among free and equal persons. By engaging in civil disobedience one intends, then, to address the sense of justice of the majority and to serve fair notice that in one's sincere and considered opinion the conditions of free cooperation are being violated. We are appealing to others to reconsider, to put them-

30. See *Nuclear Weapons and Christian Conscience*, ed. Walter Stein (London, The Merlin Press, 1965), for a presentation of this sort of doctrine in connection with nuclear war.

31. I borrow this point from Walzer, *Obligations*, p. 127.

selves in our position, and to recognize that they cannot expect us to acquiesce indefinitely in the terms they impose upon us.

Now the force of this appeal depends upon the democratic conception of society as a system of cooperation among equal persons. If one thinks of society in another way, this form of protest may be out of place. For example, if the basic law is thought to reflect the order of nature and if the sovereign is held to govern by divine right as God's chosen lieutenant, then his subjects have only the right of suppliants. They can plead their cause but they cannot disobey should their appeal be denied. To do this would be to rebel against the final legitimate moral (and not simply legal) authority. This is not to say that the sovereign cannot be in error but only that the situation is not one for his subjects to correct. But once society is interpreted as a scheme of cooperation among equals, those injured by serious injustice need not submit. Indeed, civil disobedience (and conscientious refusal as well) is one of the stabilizing devices of a constitutional system, although by definition an illegal one. Along with such things as free and regular elections and an independent judiciary empowered to interpret the constitution (not necessarily written), civil disobedience used with due restraint and sound judgment helps to maintain and strengthen just institutions. By resisting injustice within the limits of fidelity to law, it serves to inhibit departures from justice and to correct them when they occur. A general disposition to engage in justified civil disobedience introduces stability into a well-ordered society, or one that is nearly just.

It is necessary to look at this doctrine from the standpoint of the persons in the original position. There are two related problems which they must consider. The first is that, having chosen principles for individuals, they must work out guidelines for assessing the strength of the natural duties and obligations, and, in particular, the strength of the duty to comply with a just constitution and one of its basic procedures, that of majority rule. The second problem is that of finding reasonable principles for dealing with unjust situations, or with circumstances in which the compliance with just principles is only partial. Now it seems that, given the assumptions characterizing a nearly just society, the parties would agree to the presumptions (previously discussed) that specify when civil disobedience is justified. They would acknowledge these criteria as spelling out when this form of dissent is appropriate. Doing this would indicate the weight of the natural duty of justice in one important special case. It would also tend to enhance the realization of justice throughout the society by strengthening men's self-esteem as well as their respect for one another. As the contract doctrine emphasizes, the principles of justice are

the principles of willing cooperation among equals. To deny justice to another is either to refuse to recognize him as an equal (one in regard to whom we are prepared to constrain our actions by principles that we would choose in a situation of equality that is fair), or to manifest a willingness to exploit the contingencies of natural fortune and happenstance for our own advantage. In either case deliberate injustice invites submission or resistance. Submission arouses the contempt of those who perpetuate injustice and confirms their intention, whereas resistance cuts the ties of community. If after a decent period of time to allow for reasonable political appeals in the normal way, citizens were to dissent by civil disobedience when infractions of the basic liberties occurred, these liberties would, it seems, be more rather than less secure. For these reasons, then, the parties would adopt the conditions defining justified civil disobedience as a way of setting up, within the limits of fidelity to law, a final device to maintain the stability of a just constitution. Although this mode of action is strictly speaking contrary to law, it is nevertheless a morally correct way of maintaining a constitutional regime.

In a fuller account the same kind of explanation could presumably be given for the justifying conditions of conscientious refusal (again assuming the context of a nearly just state). I shall not, however, discuss these conditions here. I should like to emphasize instead that the constitutional theory of civil disobedience rests solely upon a conception of justice. Even the features of publicity and nonviolence are explained on this basis. And the same is true of the account of conscientious refusal, although it requires a further elaboration of the contract doctrine. At no point has a reference been made to other than political principles; religious or pacifist conceptions are not essential. While those engaging in civil disobedience have often been moved by convictions of this kind, there is no necessary connection between them and civil disobedience. For this form of political action can be understood as a way of addressing the sense of justice of the community, an invocation of the recognized principles of cooperation among equals. Being an appeal to the moral basis of civic life, it is a political and not a religious act. It relies upon common sense principles of justice that men can require one another to follow and not upon the affirmations of religious faith and love which they cannot demand that everyone accept. I do not mean, of course, that nonpolitical conceptions have no validity. They may, in fact, confirm our judgment and support our acting in ways known on other grounds to be just. Nevertheless, it is not these principles but the principles of justice, the fundamental terms of social cooperation between free and equal per-

sons, that underlie the constitution. Civil disobedience as defined does not require a sectarian foundation but is derived from the public conception of justice that characterizes a democratic society. So understood a conception of civil disobedience is part of the theory of free government.

One distinction between medieval and modern constitutionalism is that in the former the supremacy of law was not secured by established institutional controls. The check to the ruler who in his judgments and edicts opposed the sense of justice of the community was limited for the most part to the right of resistance by the whole society, or any part. Even this right seems not to have been interpreted as a corporate act; an unjust king was simply put aside.³² Thus the Middle Ages lacked the basic ideas of modern constitutional government, the idea of the sovereign people who have final authority and the institutionalizing of this authority by means of elections and parliaments, and other constitutional forms. Now in much the same way that the modern conception of constitutional government builds upon the medieval, the theory of civil disobedience supplements the purely legal conception of constitutional democracy. It attempts to formulate the grounds upon which legitimate democratic authority may be dissented from in ways that while admittedly contrary to law nevertheless express a fidelity to law and appeal to the fundamental political principles of a democratic regime. Thus to the legal forms of constitutionalism one may adjoin certain modes of illegal protest that do not violate the aims of a democratic constitution in view of the principles by which such dissent is guided. I have tried to show how these principles can be accounted for by the contract doctrine.

Some may object to this theory of civil disobedience that it is unrealistic. It presupposes that the majority has a sense of justice, and one might reply that moral sentiments are not a significant political force. What moves men are various interests, the desires for power, prestige, wealth, and the like. Although they are clever at producing moral arguments to support their claims, between one situation and another their opinions do not fit into a coherent conception of justice. Rather their views at any given time are occasional pieces calculated to advance certain interests. Unquestionably there is much truth in this contention, and in some societies it is more true than in others. But the essential question is the relative strength of the tendencies that oppose the sense of justice and whether the latter is ever strong enough so that it can be invoked to some significant effect.

32. See J. H. Franklin, ed., *Constitutionalism and Resistance in the Sixteenth Century* (New York, Pegasus, 1969), in the introduction, pp. 11–15.

A few comments may make the account presented more plausible. First of all, I have assumed throughout that we have to do with a nearly just society. This implies that there exists a constitutional regime and a publicly recognized conception of justice. Of course, in any particular situation certain individuals and groups may be tempted to violate its principles but the collective sentiment in their behalf has considerable strength when properly addressed. These principles are affirmed as the necessary terms of cooperation between free and equal persons. If those who perpetrate injustice can be clearly identified and isolated from the larger community, the convictions of the greater part of society may be of sufficient weight. Or if the contending parties are roughly equal, the sentiment of justice of those not engaged can be the deciding factor. In any case, should circumstances of this kind not obtain, the wisdom of civil disobedience is highly problematic. For unless one can appeal to the sense of justice of the larger society, the majority may simply be aroused to more repressive measures if the calculation of advantages points in this direction. Courts should take into account the civilly disobedient nature of the protester's act, and the fact that it is justifiable (or may seem so) by the political principles underlying the constitution, and on these grounds reduce and in some cases suspend the legal sanction.³³ Yet quite the opposite may happen when the necessary background is lacking. We have to recognize then that justifiable civil disobedience is normally a reasonable and effective form of dissent only in a society regulated to some considerable degree by a sense of justice.

There may be some misapprehension about the manner in which the sense of justice is said to work. One may think that this sentiment expresses itself in sincere professions of principle and in actions requiring a considerable degree of self-sacrifice. But this supposition asks too much. A community's sense of justice is more likely to be revealed in the fact that the majority cannot bring itself to take the steps necessary to suppress the minority and to punish acts of civil disobedience as the law allows. Ruthless tactics that might be contemplated in other societies are not entertained as real alternatives. Thus the sense of justice affects, in ways we are often unaware of, our interpretation of political life, our perception of the possible courses of action, our will to resist the justified protests of others, and so on. In spite of its superior power, the majority may abandon its position and acquiesce in the proposals of the dissenters;

33. For a general discussion, see Ronald Dworkin, "On Not Prosecuting Civil Disobedience," *The New York Review of Books*, June 6, 1968.

its desire to give justice weakens its capacity to defend its unjust advantages. The sentiment of justice will be seen as a more vital political force once the subtle forms in which it exerts its influence are recognized, and in particular its role in rendering certain social positions indefensible.

In these remarks I have assumed that in a nearly just society there is a public acceptance of the same principles of justice. Fortunately this assumption is stronger than necessary. There can, in fact, be considerable differences in citizens' conceptions of justice provided that these conceptions lead to similar political judgments. And this is possible, since different premises can yield the same conclusion. In this case there exists what we may refer to as overlapping rather than strict consensus. In general, the overlapping of professed conceptions of justice suffices for civil disobedience to be a reasonable and prudent form of political dissent. Of course, this overlapping need not be perfect; it is enough that a condition of reciprocity is satisfied. Both sides must believe that however much their conceptions of justice differ, their views support the same judgment in the situation at hand, and would do so even should their respective positions be interchanged. Eventually, though, there comes a point beyond which the requisite agreement in judgment breaks down and society splits into more or less distinct parts that hold diverse opinions on fundamental political questions. In this case of strictly partitioned consensus, the basis for civil disobedience no longer obtains. For example, suppose those who do not believe in toleration, and who would not tolerate others had they the power, wish to protest their lesser liberty by appealing to the sense of justice of the majority which holds the principle of equal liberty. While those who accept this principle should, as we have seen, tolerate the intolerant as far as the safety of free institutions permits, they are likely to resent being reminded of this duty by the intolerant who would, if positions were switched, establish their own dominion. The majority is bound to feel that their allegiance to equal liberty is being exploited by others for unjust ends. This situation illustrates once again the fact that a common sense of justice is a great collective asset which requires the cooperation of many to maintain. The intolerant can be viewed as free-riders, as persons who seek the advantages of just institutions while not doing their share to uphold them. Although those who acknowledge the principles of justice should always be guided by them, in a fragmented society as well as in one moved by group egoisms, the conditions for civil disobedience do not exist. Still, it is not necessary to have strict consensus, for often a degree of overlapping consensus allows the reciprocity condition to be fulfilled.

There are, to be sure, definite risks in the resort to civil disobedience. One reason for constitutional forms and their judicial interpretation is to establish a public reading of the political conception of justice and an explanation of the application of its principles to social questions. Up to a certain point it is better that the law and its interpretation be settled than that it be settled rightly. Therefore it may be protested that the preceding account does not determine who is to say when circumstances are such as to justify civil disobedience. It invites anarchy by encouraging everyone to decide for himself, and to abandon the public rendering of political principles. The reply to this is that each person must indeed make his own decision. Even though men normally seek advice and counsel, and accept the injunctions of those in authority when these seem reasonable to them, they are always accountable for their deeds. We cannot divest ourselves of our responsibility and transfer the burden of blame to others. This is true on any theory of political duty and obligation that is compatible with the principles of a democratic constitution. The citizen is autonomous yet he is held responsible for what he does (§78). If we ordinarily think that we should comply with the law, this is because our political principles normally lead to this conclusion. Certainly in a state of near justice there is a presumption in favor of compliance in the absence of strong reasons to the contrary. The many free and reasoned decisions of individuals fit together into an orderly political regime.

But while each person must decide for himself whether the circumstances justify civil disobedience, it does not follow that one is to decide as one pleases. It is not by looking to our personal interests, or to our political allegiances narrowly construed, that we should make up our minds. To act autonomously and responsibly a citizen must look to the political principles that underlie and guide the interpretation of the constitution. He must try to assess how these principles should be applied in the existing circumstances. If he comes to the conclusion after due consideration that civil disobedience is justified and conducts himself accordingly, he acts conscientiously. And though he may be mistaken, he has not done as he pleased. The theory of political duty and obligation enables us to draw these distinctions.

There are parallels with the common understandings and conclusions reached in the sciences. Here, too, everyone is autonomous yet responsible. We are to assess theories and hypotheses in the light of the evidence by publicly recognized principles. It is true that there are authoritative works, but these sum up the consensus of many persons each deciding for himself. The absence of a final authority to decide, and so of an official

interpretation that all must accept, does not lead to confusion, but is rather a condition of theoretical advance. Equals accepting and applying reasonable principles need have no established superior. To the question, who is to decide? the answer is: all are to decide, everyone taking counsel with himself, and with reasonableness, comity, and good fortune, it often works out well enough.

In a democratic society, then, it is recognized that each citizen is responsible for his interpretation of the principles of justice and for his conduct in the light of them. There can be no legal or socially approved rendering of these principles that we are always morally bound to accept, not even when it is given by a supreme court or legislature. Indeed each constitutional agency, the legislature, the executive, and the court, puts forward its interpretation of the constitution and the political ideals that inform it.³⁴ Although the court may have the last say in settling any particular case, it is not immune from powerful political influences that may force a revision of its reading of the constitution. The court presents its doctrine by reason and argument; its conception of the constitution must, if it is to endure, persuade the major part of the citizens of its soundness. The final court of appeal is not the court, nor the executive, nor the legislature, but the electorate as a whole. The civilly disobedient appeal in a special way to this body. There is no danger of anarchy so long as there is a sufficient working agreement in citizens' conceptions of justice and the conditions for resorting to civil disobedience are respected. That men can achieve such an understanding and honor these limits when the basic political liberties are maintained is an assumption implicit in a democratic polity. There is no way to avoid entirely the danger of divisive strife, any more than one can rule out the possibility of profound scientific controversy. Yet if justified civil disobedience seems to threaten civic concord, the responsibility falls not upon those who protest but upon those whose abuse of authority and power justifies such opposition. For to employ the coercive apparatus of the state in order to maintain manifestly unjust institutions is itself a form of illegitimate force that men in due course have a right to resist.

With these remarks we have reached the end of our discussion of the content of the principles of justice. Throughout this part my aim has been to describe a scheme of institutions that satisfies these principles and to indicate how duties and obligations arise. These things must be done to

34. For a presentation of this view to which I am indebted, see A. M. Bickel, *The Least Dangerous Branch* (New York, Bobbs-Merrill, 1962), esp. chs. V and VI.

see if the theory of justice put forward matches our considered judgments and extends them in an acceptable way. We need to check whether it defines a workable political conception and helps to focus our reflections on the most relevant and basic moral concerns. The account in this part is still highly abstract, but I hope to have provided some guidance as to how the principles of justice apply in practice. However, we should not forget the limited scope of the theory presented. For the most part I have tried to develop an ideal conception, only occasionally commenting on the various cases of nonideal theory. To be sure the priority rules suggest directives in many instances, and they may be useful if not pressed too far. Even so, the only question of nonideal theory examined in any detail is that of civil disobedience in the special case of near justice. If ideal theory is worthy of study, it must be because, as I have conjectured, it is the fundamental part of the theory of justice and essential for the nonideal part as well. I shall not pursue these matters further. We have still to complete the theory of justice by seeing how it is rooted in human thought and feeling, and tied in with our ends and aspirations.

PART THREE. ENDS

CHAPTER VII. GOODNESS AS RATIONALITY

In this final part I proceed as follows. First, I present in more detail the theory of the good which has already been used to characterize primary goods and the interests of the persons in the original position. Since a more comprehensive view is required for the subsequent argument, this theory must be given a firmer foundation. The next chapter is largely concerned with moral psychology and the acquisition of the sentiment of justice. Once these matters have been dealt with, we are in a position to discuss the relative stability of justice as fairness and to argue in the last chapter that, in a sense to be defined, justice and goodness are congruent, at least in the circumstances of a well-ordered society. Last of all I explain how the theory of justice connects up with the social values and the good of community. Sometimes in this part the overall direction of the exposition may seem less clear, and the transition from one topic to another more abrupt. It might help to keep in mind that the central aim is to prepare the way to settle the questions of stability and congruence, and to account for the values of society and the good of justice.

60. THE NEED FOR A THEORY OF THE GOOD

So far I have said very little about the concept of goodness. It was briefly mentioned earlier when I suggested that a person's good is determined by what is for him the most rational plan of life given reasonably favorable circumstances (§15). All along I have assumed that in a well-ordered society citizens' conceptions of their good conform to the principles of right publicly recognized and include an appropriate place for the various primary goods. But the concept of goodness has been used only in a rather thin sense. And in fact I shall distinguish between two theories of the good. The reason for doing this is that in justice as fairness the concept of right is prior to that of the good. In contrast with teleological

theories, something is good only if it fits into ways of life consistent with the principles of right already on hand. But to establish these principles it is necessary to rely on some notion of goodness, for we need assumptions about the parties' motives in the original position. Since these assumptions must not jeopardize the prior place of the concept of right, the theory of the good used in arguing for the principles of justice is restricted to the bare essentials. This account of the good I call the thin theory: its purpose is to secure the premises about primary goods required to arrive at the principles of justice. Once this theory is worked out and the primary goods accounted for, we are free to use the principles of justice in the further development of what I shall call the full theory of the good.

In order to clarify these matters, let us recall where a theory of the good has already played a role. First of all, it is used to define the least favored members of society. The difference principle assumes that this can be done. It is true that the theory need not define a cardinal measure of welfare. We do not have to know how disadvantaged the least fortunate are, since once this group is singled out, we can take their ordinal preferences (from the appropriate point of view) as determining the proper arrangement of the basic structure (§15). Nevertheless, we must be able to identify this group. Further, the index of well-being and the expectations of representative men are specified in terms of primary goods. Rational individuals, whatever else they want, desire certain things as prerequisites for carrying out their plans of life. Other things equal, they prefer a wider to a narrower liberty and opportunity, and a greater rather than a smaller share of wealth and income. That these things are good seems clear enough. But I have also said that self-respect and a sure confidence in the sense of one's own worth is perhaps the most important primary good. And this suggestion has been used in the argument for the two principles of justice (§29). Thus the initial definition of expectations solely by reference to such things as liberty and wealth is provisional; it is necessary to include other kinds of primary goods and these raise deeper questions. Obviously an account of the good is required for this; and it must be the thin theory.

Again, some view of goodness is used in defending justice as fairness against various objections. For example, it may be said that the persons in the original position know so little about their situation that a rational agreement upon principles of justice is impossible. Since they do not know what their aims are, they may find their plans utterly ruined by the

principles to which they consent. Therefore how can they reach a sensible decision? One might reply that the rationality of a person's choice does not depend upon how much he knows, but only upon how well he reasons from whatever information he has, however incomplete. Our decision is perfectly rational provided that we face up to our circumstances and do the best we can. Thus the parties can in fact make a rational decision, and surely some of the alternative conceptions of justice are better than others. Nevertheless, the thin theory of the good which the parties are assumed to accept shows that they should try to secure their liberty and self-respect, and that, in order to advance their aims, whatever these are, they normally require more rather than less of the other primary goods. In entering into the original agreement, then, the parties suppose that their conceptions of the good have a certain structure, and this is sufficient to enable them to choose principles on a rational basis.

Summing up these points, we need what I have called the thin theory of the good to explain the rational preference for primary goods and to explicate the notion of rationality underlying the choice of principles in the original position. This theory is necessary to support the requisite premises from which the principles of justice are derived. But looking ahead to other questions yet to be discussed, a more comprehensive account of the good is essential. Thus the definition of beneficent and supererogatory acts depends upon such a theory. So likewise does the definition of the moral worth of persons. This is the third main concept of ethics and we must find a place for it within the contract view. Eventually we shall have to consider whether being a good person is a good thing for that person, if not in general, then under what conditions. In some circumstances at least, for example those of a society well-ordered or in a state of near justice, it turns out, I believe, that being a good person is indeed a good. This fact is intimately connected with the good of justice and the problem of the congruence of a moral theory. We need an account of the good to spell all this out. The characteristic feature of this full theory, as I have said, is that it takes the principles of justice as already secured, and then uses these principles in defining the other moral concepts in which the notion of goodness is involved. Once the principles of right are on hand, we may appeal to them in explaining the concept of moral worth and the good of the moral virtues. Indeed, even rational plans of life which determine what things are good for human beings, the values of human life so to speak, are themselves constrained by the principles of justice. But clearly, to avoid moving in a circle, we must

distinguish between the thin and the full theory, and always keep in mind which one we are relying upon.

Finally, when we come to the explanation of the social values and the stability of a conception of justice, a wider interpretation of the good is required. For example, one basic psychological principle is that we have a tendency to love those who manifestly love us, those who with evident intention advance our good. In this instance our good comprises final ends and not only primary goods. Moreover, in order to account for the social values, we need a theory that explains the good of activities, and in particular the good of everyone's willingly acting from the public conception of justice in affirming their social institutions. When we consider these questions we can work within the full theory. Sometimes we are examining the processes by which the sense of justice and moral sentiments are acquired; or else we are noting that the collective activities of a just society are also good. There is no reason for not using the full theory, since the conception of justice is available.

However, when we ask whether the sense of justice is a good, the important question clearly is that defined by the thin theory. We want to know whether having and maintaining a sense of justice is a good (in the thin sense) for persons who are members of a well-ordered society. Surely if the sentiment of justice is ever a good, it is a good in this special case. And if within the thin theory it turns out that having a sense of justice is indeed a good, then a well-ordered society is as stable as one can hope for. Not only does it generate its own supportive moral attitudes, but these attitudes are desirable from the standpoint of rational persons who have them when they assess their situation independently from the constraints of justice. This match between justice and goodness I refer to as congruence; and I shall examine this relation when we take up the good of justice (§86).

61. THE DEFINITION OF GOOD FOR SIMPLER CASES

Rather than proceeding immediately to the application of the concept of rationality to the assessment of plans, it seems best to illustrate the definition I shall use by first considering simpler cases. Doing this will bring out several distinctions that are necessary for a clear understanding of its sense. Thus I suppose the definition to have three stages as follows (for simplicity these stages are formulated using the concept of goodness rather than that of better than): (1) A is a good X if and only if A has the

properties (to a higher degree than the average¹ or standard X) which it is rational to want in an X, given what X's are used for, or expected to do, and the like (whichever rider is appropriate); (2) A is a good X for K (where K is some person) if and only if A has the properties which it is rational for K to want in an X, given K's circumstances, abilities, and plan of life (his system of aims), and therefore in view of what he intends to do with an X, or whatever; (3) the same as 2 but adding a clause to the effect that K's plan of life, or that part of it relevant in the present instance, is itself rational. What rationality means in the case of plans has yet to be determined and will be discussed later on. But according to the definition, once we establish that an object has the properties that it is rational for someone with a rational plan of life to want, then we have shown that it is good for him. And if certain sorts of things satisfy this condition for persons generally, then these things are human goods. Eventually we want to be assured that liberty and opportunity, and a sense of our own worth, fall into this category.²

Now for a few comments on the first two stages of the definition. We tend to move from the first stage to the second whenever it is necessary to take into account the special features of a person's situation which the definition defines to be relevant. Typically these features are his interests, abilities, and circumstances. Although the principles of rational choice have not yet been set out, the everyday notion seems clear enough for the

1. See W. D. Ross, *The Right and the Good* (Oxford, The Clarendon Press, 1930), p. 67.

2. As I have remarked, there is wide agreement, with many variations, on an account of the good along these lines. See Aristotle, *Nicomachean Ethics*, bks. I and X; and Aquinas, *Summa Theologica*, I-I, q. 5-6, *Summa Contra Gentiles*, bk. III, chs. 1-63, and *Treatise on Happiness*, trans. J. A. Oesterle (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1964). For Kant, *The Fundamental Principles of the Metaphysics of Morals*, Academy Edition, vol. IV, pp. 415-419; and *The Critique of Practical Reason*, first part of ch. II, bk. I of pt. I. See H. J. Paton's discussion of Kant, *In Defense of Reason* (London, George Allen and Unwin, Ltd., 1951), pp. 157-177. For Sidgwick, *Methods of Ethics*, 7th ed. (London, Macmillan, 1907), bk. I, ch. IX, and bk. III, ch. XIV. This kind of view is held by idealists and those influenced by them. See, for example, F. H. Bradley, *Ethical Studies*, 2nd ed. (Oxford, The Clarendon Press, 1926), ch. II; and Josiah Royce, *The Philosophy of Loyalty* (New York, Macmillan, 1908), lect. II. And more recently, H. J. Paton, *The Good Will* (London, George Allen and Unwin, 1927), bks. II and III, esp. chs. VIII and IX; W. D. Lamont, *The Value Judgment* (Edinburgh, The University Press, 1955); and J. N. Findlay, *Values and Intentions* (London, George Allen and Unwin, 1961), ch. V, secs. I and III, and ch. VI. For the so-called naturalists in value theory, see John Dewey, *Human Nature and Conduct* (New York, Henry Holt, 1922), pt. III; R. B. Perry, *General Theory of Value* (New York, Longmans, Green, 1926), chs. XX-XXII; and C. I. Lewis, *An Analysis of Knowledge and Valuation* (LaSalle, Ill., Open Court Publishing Co., 1946), bk. III. My account is indebted to J. O. Urmson, "On Grading," *Mind*, vol. 59 (1950); Paul Ziff, *Semantic Analysis* (Ithaca, N.Y., Cornell University Press, 1960), ch. VI; and Philippa Foot, "Goodness and Choice," *Proceedings of the Aristotelian Society*, supp. vol. 35 (1961), though they may not approve of what I say.

time being. In general, there is a reasonably precise sense in speaking simply of a good object of a certain kind, a sense explained by the first stage, provided that there is enough similarity of interests and circumstances among persons concerned with objects of this kind so that recognized standards can be established. When these conditions are met, saying that something is good conveys useful information. There is sufficient common experience with or knowledge of these things for us to have an understanding of the desired features exemplified by an average or standard object. Often there are conventional criteria founded upon commercial or other practice which define these properties.³ By taking up various examples we could no doubt see how these criteria evolve and the relevant standards determined. The essential point, however, is that these criteria depend upon the nature of the objects in question and upon our experience with them; and therefore we say that certain things are good without further elaboration only when a certain background is presupposed or some particular context is taken for granted. The basic value judgments are those made from the standpoint of persons, given their interests, abilities, and circumstances. Only insofar as a similarity of conditions permits can we safely abstract from anyone's special situation. In cases of any complexity, when the thing to be chosen should be adjusted to specific wants and situations, we move to the second stage of the definition. Our judgments of value are tailored to the agent in question as this stage requires.

These remarks may be illustrated by looking at several examples from certain typical categories: artifacts, functional parts of systems, and occupations and roles. Among artifacts, a good watch, say, is one that has the features which it is rational to want in a watch. There are clearly a number of desired features here, in addition to that of keeping accurate time. It must not be excessively heavy, for example. These features must be measured somehow and assigned appropriate weights in the overall assessment. I shall not consider here how these things are done. It is worth noting, however, that if we take the definition of good in the traditional sense as an analysis, that is, as a statement of concept identity, and if we suppose that by definition a watch is an article used to tell time, and that by definition rationality is taking effective means to achieve one's ends, then it is analytic that a good watch is one that keeps accurate time. This fact is established solely by virtue of truths of logic and definitions of concepts. But since I do not wish to take the definition of good in

3. See Urmson, "On Grading," pp. 148–154.

this sense but rather as a rough guideline for constructing substitute expressions that can be used to say what on reflection we want to say, I do not count this statement as analytic. In fact, for our present purposes I shall sidestep this question entirely and simply take certain facts about watches (or whatever) as common knowledge. There is no occasion to ask whether the statements that express them are analytic. On this account, then, it is certainly true that a good watch keeps accurate time and this correspondence with everyday facts suffices to confirm the propriety of the definition.

Again, it is plain that the letter “X” in the phrase “a good X” often has to be replaced by various noun phrases depending on the context. Thus it is usually not enough to speak of good watches, since we frequently need a more fine-grained classification. We are called upon to assess wrist watches, stop watches, and so on; or even wrist watches to go with a particular kind of evening dress. In all these cases special interests give rise to certain appropriate classifications and standards. These complications are ordinarily gathered from the circumstances and are explicitly mentioned when it seems necessary. With things that are not artifacts some elaboration is usually called for to explain one’s meaning since it is not provided by the reference to the object. Thus, for example, the statement that Wildcat is a good mountain may require the kind of amplification provided by adding that it is a good mountain for skiing. Or the observation that it is a good night may call for the explanation that it is a good night for seeing the stars, since it is a clear and dark night. Some terms suggest the appropriate expansion. Consider an example: if we compare the statement that a body is a good corpse with the statement that it is a good cadaver, the sense of the first is not clear, whereas referring to something as a cadaver conveys its use in the study of anatomy. A good cadaver is presumably a corpse having the properties (whatever they are) which it is rational to want for this purpose.⁴ It may be noted in passing that we can understand at least part of what is meant by calling something good even though we do not know what are the desired features of the object being evaluated.

There always stands in the background a point of view from which an artifact, functional part, or role is being appraised, although of course this point of view need not be made explicit. This perspective is characterized by identifying the persons whose concerns are relevant for making the judgment, and then by describing the interests which they take in the

4. The example is from Ziff, *Semantic Analysis*, p. 211.

object. For example, in the case of parts of the body (functional parts of systems), we normally take up the point of view of the person in question and presume that his interest is the normal one. Thus good eyes and ears are those having the properties that it is rational to want in one's own eyes and ears when one wishes to see and hear well. Similarly with animals and plants: when we say that they have a good coat, or good roots, we appear to adopt the point of view of the animal or plant. No doubt there is some artificiality in doing this, especially in the case of plants. On the other hand, perhaps there are other perspectives that would explain these judgments more naturally. But the definition is likely to be more suitable for some cases than others, and this fact need not worry us too much so long as it is satisfactory for the purposes of the theory of justice. Turning to the category of occupations, in some instances anyway while the desired properties are those of persons belonging to the occupation, the persons whose point of view we take up do not belong to it. Thus a good doctor is one who has the skills and abilities that it is rational for his patients to want in a doctor. The skills and abilities are the doctor's, the interest in the restoration of health by which they are assessed are the patients'. These illustrations show that the point of view varies from case to case and the definition of goodness contains no general formula for determining it. These matters are explained as the occasion arises or gathered from the context.

A further comment is that there is nothing necessarily right, or morally correct, about the point of view from which things are judged to be good or bad.⁵ One may say of a man that he is a good spy, or a good assassin, without approving of his skills. Applying the definition to this case, we would be interpreted as saying that the individual referred to has the attributes that it is rational to want in a spy, or assassin, given what spies and assassins are expected to do. There is no implication that it is proper to want spies and assassins to do what they do. Normally it is governments and conspirators and the like who employ spies and assassins. We are simply evaluating certain proficiencies and talents from the point of view of governments and conspirators. Whether a spy or assassin is a good person is a separate question altogether; to answer it we should have to judge the cause for which he works and his motives for doing so.

Now this moral neutrality of the definition of good is exactly what we should expect. The concept of rationality by itself is not an adequate basis

5. On this point, see Ross, *The Right and the Good*, p. 67. A somewhat different view is expressed by A. E. Duncan-Jones, "Good Things and Good Thieves," *Analysis*, vol. 27 (1966), pp. 113–118.

for the concept of right; and in contract theory the latter is derived in another way. Moreover, to construct the conception of moral goodness, the principles of right and justice must be introduced. It is easy to see that with many occupations and roles moral principles have an important place in characterizing the desired properties. For example, a good judge has a strong desire to give justice, to decide cases fairly in accordance with what the law requires. He possesses the judicial virtues which his position demands: he is impartial, able to assess the evidence fairly, not prejudiced or moved by personal considerations. These attributes may not suffice but they are generally necessary. The characterizations of a good father or wife, friend or associate, and so on indefinitely, rely upon a theory of the virtues and therefore presuppose the principles of right. These matters belong to the full theory. In order for goodness as rationality to hold for the concept of moral worth, it must turn out that the virtues are properties that it is rational for persons to want in one another when they adopt the requisite point of view. I shall try to show in due course that this is in fact the case (§66).

62. A NOTE ON MEANING

I shall supplement this account of the thin theory with a few words about the meaning of judgments of value. These matters are not central to our inquiry but several comments may prevent misunderstanding. Perhaps the chief issue is whether these judgments represent a descriptive or a prescriptive use of language. Unfortunately the notions of a descriptive and a prescriptive use are obscure, but I shall try to come to the main point straightway.⁶ All sides seem to agree upon two general facts. First, the terms “good” and “bad” and the like are typically used in giving advice and counsel, and to praise and extol, and so on. To be sure, these terms are not always used in this manner, since they may appear in conditional statements, in commands and questions, as well as in other remarks that have no practical bearings. Still, their role in giving advice and counsel and in praising and extolling is characteristic. Second, the criteria for evaluation vary from one kind of thing to another. What is wanted in dwellings is not what is wanted in clothes. A satisfactory definition of the goodness must fit these two facts.

6. For the most part my account follows J. R. Searle, “Meaning and Speech Acts,” *Philosophical Review*, vol. 71 (1962). See also his *Meaning and Speech Acts* (Cambridge, The University Press, 1969), ch. VI; and Ziff, *Semantic Analysis*, ch. VI.

Now I shall simply define a descriptive theory as maintaining the following pair of theses. First, despite the variation in criteria from object to object, the term “good” has a constant sense (or meaning) that, for philosophical purposes, is of the same kind as that of other predicates normally counted as descriptive. Indeed, this constant sense enables us to understand why and how the criteria for evaluation vary from one kind of thing to another. The other thesis is that the propriety of using the term “good” (and its relatives) in giving advice and counsel, and in expressions of commendation, is explained by this constant sense together with a general theory of meaning. I assume that this theory includes an account of speech acts and illocutionary forces along the lines suggested by Austin.⁷ A descriptive theory holds that the constant descriptive meaning of good accounts for its being used, when in fact it is properly used, to praise and to advise, and the like. There is no necessity to assign “good” a special kind of meaning which is not already explained by its constant descriptive sense and the general theory of speech acts.

Goodness as rationality is a descriptive theory in this sense. In the required way, it explains the two general facts which everyone recognizes. The constant sense of “good” is characterized by the definition in its several stages. Thus something’s being good is its having the properties that it is rational to want in things of its kind, plus further elaborations depending on the case. In the light of this definition it is easy to account for the fact that the criteria of evaluation differ from one kind of thing to another. Since we want things for different purposes, it is obviously rational to assess them by different features. It is helpful to think of the sense of “good” as being analogous to that of a function sign.⁸ We can then view the definition as assigning to each kind of thing a set of properties by which instances of that kind are to be assessed, namely, the properties which it is rational to want in things of that kind.

Furthermore, the account of goodness as rationality explains why the term “good” appears in statements of advice and counsel, and in remarks of praise and approval. Thus, for example, when we are asked for advice someone wishes to have our opinion as to which course of action, say, is best for him. He wants to know what we think is rational for him to do. A climber who advises another about the equipment and route to use on a difficult pitch takes up the other’s standpoint and recommends what he thinks is a sensible plan of attack. The meaning of “good” and of related

7. See J. L. Austin, *How To Do Things with Words* (Oxford, The Clarendon Press, 1962), esp. pp. 99–109, 113–116, 145f.

8. Here I borrow from P. T. Geach, “Good and Evil,” *Analysis*, vol. 17 (1956), pp. 37f.

expressions does not change in those statements that are counted as advisory. It is the context that converts what we say into advice even though the sense of our words is the same. Climbers, for example, have a duty of mutual aid to help one another, and hence they have a duty to offer their considered opinion in urgent circumstances. In these situations their words become advisory. And so as the situation warrants, what we say may be, and in some cases must be, reckoned as advice and counsel. Accepting the theory of right already sketched, the constant descriptive sense together with the general reasons why persons seek out the views of others explain these characteristic uses of "good." At no point must we appeal to a special kind of prescriptive or emotive meaning.

It may be objected to these remarks that the theory of illocutionary forces allows all that has been claimed by those who have proposed a prescriptive or an emotive theory of meaning. If so, there may be no disagreement. I have not denied that the understanding of the illocutionary forces of the various uses of "good," its being employed in statements of praise and advice, and the like, is relevant to grasping the meaning of the term. Nor do I oppose the view that a certain illocutionary force is central to "good," in the sense that one cannot accept as true the statement that something is good and at the same time dissent from its illocutionary force (assuming this force to obtain in the context).⁹ The question is how these facts are to be explained.

Thus the descriptive theory maintains that "good" is characteristically used with the force of a recommendation or advice, and the like, precisely because of its descriptive sense as given by the definition. The descriptive meaning of "good" is not simply a family of lists of properties, a list for each kind of thing according to convention or preference. Rather in the way that the definition explains, these lists are formed in the light of what it is rational to want in objects of various kinds. Therefore understanding why the word "good" (and its relatives) is employed in these speech acts is part of understanding this constant sense. Similarly, certain illocutionary forces are central to "good" as a result of its descriptive meaning, just as the force of factual narration belongs to some utterances in virtue of their descriptive meaning. For if we assent to the statement that something is best for us when it is offered as advice, say, we will indeed accept this advice and act upon it if we are rational. The dispute, if there is one, is not about these recognized facts but concerns the place of the descrip-

9. For these and other points, see J. O. Urmson, *The Emotive Theory of Ethics* (London, Hutchinson University Library, 1968), pp. 136–145.

tive meaning of “good” in explaining them. The descriptive theory holds that conjoined to a general theory of speech acts the definition of “good” yields an adequate account of these facts. There is no occasion to introduce a distinct kind of meaning.

63. THE DEFINITION OF GOOD FOR PLANS OF LIFE

To this point I have discussed only the first stages of the definition of good in which no questions are raised about the rationality of the ends taken as given. A thing's being a good X for K is treated as equivalent to its having the properties which it is rational for K to want in an X in view of his interests and aims. Yet we often assess the rationality of a person's desires, and the definition must be extended to cover this fundamental case if it is to serve the purposes of the theory of justice. Now the basic idea at the third stage is to apply the definition of good to plans of life. The rational plan for a person determines his good. Here I adapt Royce's thought that a person may be regarded as a human life lived according to a plan. For Royce an individual says who he is by describing his purposes and causes, what he intends to do in his life.¹⁰ If this plan is a rational one, then I shall say that the person's conception of his good is likewise rational. In his case the real and the apparent good coincide. Similarly his interests and aims are rational, and it is appropriate to take them as stopping points in making judgments that correspond to the first two stages of the definition. These suggestions are quite straightforward but unfortunately setting out the details is somewhat tedious. In order to expedite matters I shall start off with a pair of definitions and then explain and comment on them over the next several sections.

These definitions read as follows: first, a person's plan of life is rational if, and only if, (1) it is one of the plans that is consistent with the principles of rational choice when these are applied to all the relevant

10. See *The Philosophy of Loyalty*, lect. IV, sec. IV. Royce uses the notion of a plan to characterize the coherent, systematic purposes of the individual, what makes him a conscious, unified moral person. In this Royce is typical of the philosophical usage found in many of the writers cited in §61, note 2, Dewey and Perry, for example. And I shall do the same. The term is given no technical sense, nor are the structures of plans invoked to get other than obvious common sense results. These are matters I do not investigate. For a discussion of plans, see G. A. Miller, Eugene Galanter, and K. H. Pribram, *Plans and the Structure of Behavior* (New York, Henry Holt, 1960); and also Galanter's *Textbook of Elementary Psychology* (San Francisco, Holden-Day, 1966), ch. IX. The notion of a plan may prove useful in characterizing intentional action. See, for example, Alvin Goldman, *A Theory of Action* (Englewood Cliffs, N.J., Prentice-Hall, 1970), pp. 56–73, 76–80; but I do not consider this question.

features of his situation, and (2) it is that plan among those meeting this condition which would be chosen by him with full deliberative rationality, that is, with full awareness of the relevant facts and after a careful consideration of the consequences.¹¹ (The notion of deliberative rationality is discussed in the next section.) Secondly, a person's interests and aims are rational if, and only if, they are to be encouraged and provided for by the plan that is rational for him. Note that in the first of these definitions I have implied that a rational plan is presumably but one of many possible plans that are consistent with the principles of rational choice. The reason for this complication is that these principles do not single out one plan as the best. We have instead a maximal class of plans: each member of this class is superior to all plans not included in it, but given any two plans in the class, neither is superior or inferior to the other. Thus to identify a person's rational plan, I suppose that it is that plan belonging to the maximal class which he would choose with full deliberative rationality. We criticize someone's plan, then, by showing either that it violates the principles of rational choice, or that it is not the plan that he would pursue were he to assess his prospects with care in the light of a full knowledge of his situation.

Before illustrating the principles of rational choice, I should say a few things about the rather complex notion of a rational plan. It is fundamental for the definition of good, since a rational plan of life establishes the basic point of view from which all judgments of value relating to a particular person are to be made and finally rendered consistent. Indeed, with certain qualifications (§83) we can think of a person as being happy when he is in the way of a successful execution (more or less) of a rational plan of life drawn up under (more or less) favorable conditions, and he is reasonably confident that his plan can be carried through. Someone is happy when his plans are going well, his more important aspirations being fulfilled, and he feels sure that his good fortune will endure. Since plans which it is rational to adopt vary from person to person depending upon their endowments and circumstances, and the like, different individuals find their happiness in doing different things. The gloss concerning favorable circumstances is necessary because even a rational arrangement of one's activities can be a matter of accepting the lesser evil if natural conditions are harsh and the demands of other men oppressive.

11. For simplicity I assume that there is one and only one plan that would be chosen, and not several (or many) between which the agent would be indifferent, or whatever. Thus I speak throughout of the plan that would be adopted with deliberative rationality.

The achievement of happiness in the larger sense of a happy life, or of a happy period of one's life, always presumes a degree of good fortune.

Several further points about long-term plans should be mentioned. The first relates to their time structure. A plan will, to be sure, make some provision for even the most distant future and for our death, but it becomes relatively less specific for later periods. Certain broad contingencies are insured against and general means provided for, but the details are filled in gradually as more information becomes available and our wants and needs are known with greater accuracy. Indeed, one principle of rational choice is that of postponement: if in the future we may want to do one of several things but are unsure which, then, other things equal, we are to plan now so that these alternatives are both kept open. We must not imagine that a rational plan is a detailed blueprint for action stretching over the whole course of life. It consists of a hierarchy of plans, the more specific subplans being filled in at the appropriate time.

The second point is connected with the first. The structure of a plan not only reflects the lack of specific information but it also mirrors a hierarchy of desires proceeding in similar fashion from the more to the less general. The main features of a plan encourage and secure the fulfillment of the more permanent and general aims. A rational plan must, for example, allow for the primary goods, since otherwise no plan can succeed; but the particular form that the corresponding desires will take is usually unknown in advance and can wait for the occasion. Thus while we know that over any extended period of time we shall always have desires for food and drink, it is not until the moment comes that we decide to have a meal consisting of this or that course. These decisions depend on the choices available, on the menu that the situation allows.

Thus planning is in part scheduling.¹² We try to organize our activities into a temporal sequence in which each is carried on for a certain length of time. In this way a family of interrelated desires can be satisfied in an effective and harmonious manner. The basic resources of time and energy are allotted to activities in accordance with the intensity of the wants that they answer to and the contribution that they are likely to make to the fulfillment of other ends. The aim of deliberation is to find that plan which best organizes our activities and influences the formation of our subsequent wants so that our aims and interests can be fruitfully combined into one scheme of conduct. Desires that tend to interfere with

12. See J. D. Mabbott, "Reason and Desire," *Philosophy*, vol. 28 (1953), for a discussion of this and other points to which I am indebted.

other ends, or which undermine the capacity for other activities, are weeded out; whereas those that are enjoyable in themselves and support other aims as well are encouraged. A plan, then, is made up of subplans suitably arranged in a hierarchy, the broad features of the plan allowing for the more permanent aims and interests that complement one another. Since only the outlines of these aims and interests can be foreseen, the operative parts of the subplans that provide for them are finally decided upon independently as we go along. Revisions and changes at the lower levels do not usually reverberate through the entire structure. If this conception of plans is sound, we should expect that the good things in life are, roughly speaking, those activities and relationships which have a major place in rational plans. And primary goods should turn out to be those things which are generally necessary for carrying out such plans successfully whatever the particular nature of the plan and of its final ends.

These remarks are unhappily too brief. But they are intended only to prevent the more obvious misunderstandings of the notion of a rational plan, and to indicate the place of this notion in a theory of the good. I must now try to convey what is meant by the principles of rational choice. These principles are to be given by enumeration so that eventually they replace the concept of rationality. The relevant features of a person's situation are identified by these principles and the general conditions of human life to which plans must be adjusted. At this point I shall mention those aspects of rationality that are most familiar and seem least in dispute. And for the moment I shall assume that the choice situation relates to the short term. The question is how to fill in the more or less final details of a subplan to be executed over a relatively brief period of time, as when we make plans for a holiday. The larger system of desires may not be significantly affected, although of course some desires will be satisfied in this interval and others will not.

Now for short-term questions anyway, certain principles seem perfectly straightforward and not in dispute. The first of these is that of effective means. Suppose that there is a particular objective that is wanted, and that all the alternatives are means to achieve it, while they are in other respects neutral. The principle holds that we are to adopt that alternative which realizes the end in the best way. More fully: given the objective, one is to achieve it with the least expenditure of means (whatever they are); or given the means, one is to fulfill the objective to the fullest possible extent. This principle is perhaps the most natural criterion of rational choice. Indeed, as we shall note later, there is some tendency to suppose that deliberation must always take this form, being regulated

ultimately by a single final end (§83). Otherwise it is thought that there is no rational way to balance a plurality of aims against one another. But this question I leave aside for the present.

The second principle of rational choice is that one (short-term) plan is to be preferred to another if its execution would achieve all of the desired aims of the other plan and one or more further aims in addition. Perry refers to this criterion as the principle of inclusiveness and I shall do the same.¹³ Thus we are to follow the more inclusive plan if such a plan exists. To illustrate, suppose that we are planning a trip and we have to decide whether to go to Rome or Paris. It seems impossible to visit both. If on reflection it is clear that we can do everything in Paris that we want to do in Rome, and some other things as well, then we should go to Paris. Adopting this plan will realize a larger set of ends and nothing is left undone that might have been realized by the other plan. Often, however, neither plan is more inclusive than the other; each may achieve an aim which the other does not. We must invoke some other principle to make up our minds, or else subject our aims to further analysis (§83).

A third principle we may call that of the greater likelihood. Suppose that the aims which may be achieved by two plans are roughly the same. Then it may happen that some objectives have a greater chance of being realized by one plan than the other, yet at the same time none of the remaining aims are less likely to be attained. For example, although one can perhaps do everything one wants to do in both Rome and Paris, some of the things one wishes to do seem more likely to meet with success in Paris, and for the rest it is roughly the same. If so, the principle holds that one should go to Paris. A greater likelihood of success favors a plan just as the more inclusive end does. When these principles work together the choice is as obvious as can be. Suppose that we prefer a Titian to a Tintoretto, and that the first of two lottery tickets gives the larger chance to Titian while the second assigns it to the Tintoretto. Then one must prefer the first ticket.

So far we have been considering the application of the principles of rational choice to the short-term case. I now wish to examine the other extreme in which one has to adopt a long-term plan, even a plan of life, as when we have to choose a profession or occupation. It may be thought that having to make such a decision is a task imposed only by a particular form of culture. In another society this choice might not arise. But in fact the question of what to do with our life is always there, although some

13. See *General Theory of Value* (New York, Longmans, Green, 1926), pp. 645–649.

societies force it upon us more obviously than others and at a different time of life. The limit decision to have no plan at all, to let things come as they may, is still theoretically a plan that may or may not be rational. Accepting the idea of a long-term plan, then, it seems clear that such a scheme is to be assessed by what it will probably lead to in each future period of time. The principle of inclusiveness in this case, therefore, runs as follows: one long-term plan is better than another for any given period (or number of periods) if it allows for the encouragement and satisfaction of all the aims and interests of the other plan and for the encouragement and satisfaction of some further aim or interest in addition. The more inclusive plan, if there is one, is to be preferred: it comprehends all the ends of the first plan and at least one other end as well. If this principle is combined with that of effective means, then together they define rationality as preferring, other things equal, the greater means for realizing our aims, and the development of wider and more varied interests assuming that these aspirations can be carried through. The principle of greater likelihood supports this preference even in situations when we cannot be sure that the larger aims can be executed, provided that the chances of execution are as great as with the less comprehensive plan.

The application of the principles of effective means and the greater likelihood to the long-term case seems sound enough. But the use of the principle of inclusiveness may seem problematical. With a fixed system of ends in the short run, we assume that we already have our desires and given this fact we consider how best to satisfy them. But in long-term choice, although we do not yet have the desires which various plans will encourage, we are nevertheless directed to adopt that plan which will develop the more comprehensive interests on the assumption that these further aims can be realized. Now a person may say that since he does not have the more inclusive interests, he is not missing anything in not deciding to encourage and to satisfy them. He may hold that the possible satisfaction of desires that he can arrange never to have is an irrelevant consideration. Of course, he might also contend that the more inclusive system of interests subjects him to a greater risk of dissatisfaction; but this objection is excluded since the principle assumes that the larger pattern of ends is equally likely to be attained.

There are two considerations that seem to favor the principle of inclusiveness in the long-term case. First of all, assuming that how happy a person is depends in part upon the proportion of his aims that are achieved, the extent to which his plans are carried through, it follows that pursuing the principle of inclusiveness tends to raise this proportion and

thereby enhance a person's happiness. This effect is absent only in the case where all of the aims of the less inclusive plan are already safely provided for. The other consideration is that, in accordance with the Aristotelian Principle (explained below in §65), I assume that human beings have a higher-order desire to follow the principle of inclusiveness. They prefer the more comprehensive long-term plan because its execution presumably involves a more complex combination of abilities. The Aristotelian Principle states that, other things equal, human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and that this enjoyment increases the more the capacity is realized, or the greater its complexity. A person takes pleasure in doing something as he becomes more proficient at it, and of two activities which he performs equally well, he prefers the one that calls upon the greater number of more subtle and intricate discriminations. Thus the desire to carry out the larger pattern of ends which brings into play the more finely developed talents is an aspect of the Aristotelian Principle. And this desire, along with the higher-order desires to act upon other principles of rational choice, is one of the regulative ends that moves us to engage in rational deliberation and to follow its outcome.

Many things in these remarks call for further explanation. It is clear, for example, that these three principles are not in general sufficient to rank the plans open to us. Means may not be neutral, inclusive plans may not exist, the aims achieved may not be sufficiently similar, and so on. To apply these principles we view our aims as we are inclined to describe them, and more or less count the number realized by this or that plan, or estimate the likelihood of success. For this reason I shall refer to these criteria as counting principles. They do not require a further analysis or alteration of our desires, nor a judgment concerning the relative intensity of our wants. These matters I put aside for the discussion of deliberative rationality. It seems best to conclude this preliminary account by noting what seems to be reasonably clear: namely that we can choose between rational plans of life. And this means that we can choose now which desires we shall have at a later time.

One might suppose at first that this is not possible. We sometimes think that our major desires at least are fixed and that we deliberate solely about the means to satisfy them. Of course, it is obvious that deliberation leads us to have some desires that we did not have before, for example, the desire to avail ourselves of certain means that we have on reflection come to see as useful for our purposes. Furthermore, it is clear that taking thought may lead us to make a general desire more specific, as when a

desire for music becomes a desire to hear a particular work. But let us suppose that, except for these sorts of exceptions, we do not choose now what to desire now. Nevertheless, we can certainly decide now to do something that we know will affect the desires that we shall have in the future. At any given time rational persons decide between plans of action in view of their situation and beliefs, all in conjunction with their present major desires and the principles of rational choice. Thus we choose between future desires in the light of our existing desires, including among these the desire to act on rational principles. When an individual decides what to be, what occupation or profession to enter, say, he adopts a particular plan of life. In time his choice will lead him to acquire a definite pattern of wants and aspirations (or the lack thereof), some aspects of which are peculiar to him while others are typical of his chosen occupation or way of life. These considerations appear evident enough, and simply parallel in the case of the individual the deep effects that a choice of a conception of justice is bound to have upon the kinds of aims and interests encouraged by the basic structure of society. Convictions about what sort of person to be are similarly involved in the acceptance of principles of justice.

64. DELIBERATIVE RATIONALITY

I have already noted that the simpler principles of rational choice (the counting principles) do not suffice to order plans. Sometimes they do not apply, since there may be no inclusive plan, say, or else the means are not neutral. Or it often happens that we are left with a maximal class. In these cases further rational criteria may of course be invoked, and some of these I shall discuss below. But I shall suppose that while rational principles can focus our judgments and set up guidelines for reflection, we must finally choose for ourselves in the sense that the choice often rests on our direct self-knowledge not only of what things we want but also of how much we want them. Sometimes there is no way to avoid having to assess the relative intensity of our desires. Rational principles can help us to do this, but they cannot always determine these estimates in a routine fashion. To be sure, there is one formal principle that seems to provide a general answer. This is the principle to adopt that plan which maximizes the expected net balance of satisfaction. Or to express the criterion less hedonistically, if more loosely, one is directed to take that course most likely to realize one's most important aims. But this principle also fails to

provide us with an explicit procedure for making up our minds. It is clearly left to the agent himself to decide what it is that he most wants and to judge the comparative importance of his several ends.

At this point I introduce the notion of deliberative rationality following an idea of Sidgwick's. He characterizes a person's future good on the whole as what he would now desire and seek if the consequences of all the various courses of conduct open to him were, at the present point of time, accurately foreseen by him and adequately realized in imagination. An individual's good is the hypothetical composition of impulsive forces that results from deliberative reflection meeting certain conditions.¹⁴ Adjusting Sidgwick's notion to the choice of plans, we can say that the rational plan for a person is the one (among those consistent with the counting principles and other principles of rational choice once these are established) which he would choose with deliberative rationality. It is the plan that would be decided upon as the outcome of careful reflection in which the agent reviewed, in the light of all the relevant facts, what it would be like to carry out these plans and thereby ascertained the course of action that would best realize his more fundamental desires.

In this definition of deliberative rationality it is assumed that there are no errors of calculation or reasoning, and that the facts are correctly assessed. I suppose also that the agent is under no misconceptions as to what he really wants. In most cases anyway, when he achieves his aim, he does not find that he no longer wants it and wishes that he had done something else instead. Moreover, the agent's knowledge of his situation and the consequences of carrying out each plan is presumed to be accurate and complete. No relevant circumstances are left out of account. Thus the best plan for an individual is the one that he would adopt if he possessed full information. It is the objectively rational plan for him and determines his real good. As things are, of course, our knowledge of what will happen if we follow this or that plan is usually incomplete. Often we do not know what is the rational plan for us; the most that we can have is a reasonable belief as to where our good lies, and sometimes we can only conjecture. But if the agent does the best that a rational person can do with the information available to him, then the plan he follows is a subjectively rational plan. His choice may be an unhappy one, but if so it is because his beliefs are understandably mistaken or his knowledge insufficient, and not because he drew hasty and fallacious inferences or was confused as to what he really wanted. In this case a

14. See *The Methods of Ethics*, 7th ed. (London, Macmillan, 1907), pp. 111f.

person is not to be faulted for any discrepancy between his apparent and his real good.

The notion of deliberative rationality is obviously highly complex, combining many elements. I shall not attempt to enumerate here all the ways in which the process of reflection may go wrong. One could if necessary classify the kinds of mistakes that can be made, the sorts of tests that the agent might apply to see if he has adequate knowledge, and so on. It should be noted, however, that a rational person will not usually continue to deliberate until he has found the best plan open to him. Often he will be content if he forms a satisfactory plan (or subplan), that is, one that meets various minimum conditions.¹⁵ Rational deliberation is itself an activity like any other, and the extent to which one should engage in it is subject to rational decision. The formal rule is that we should deliberate up to the point where the likely benefits from improving our plan are just worth the time and effort of reflection. Once we take the costs of deliberation into account, it is unreasonable to worry about finding the best plan, the one that we would choose had we complete information. It is perfectly rational to follow a satisfactory plan when the prospective returns from further calculation and additional knowledge do not outweigh the trouble. There is even nothing irrational in an aversion to deliberation itself provided that one is prepared to accept the consequences. Goodness as rationality does not attribute any special value to the process of deciding. The importance to the agent of careful reflection will presumably vary from one individual to another. Nevertheless, a person is being irrational if his unwillingness to think about what is the best (or a satisfactory) thing to do leads him into misadventures that on consideration he would concede that he should have taken thought to avoid.

In this account of deliberative rationality I have assumed a certain competence on the part of the person deciding: he knows the general features of his wants and ends both present and future, and he is able to estimate the relative intensity of his desires, and to decide if necessary what he really wants. Moreover, he can envisage the alternatives open to him and establish a coherent ordering of them: given any two plans he can work out which one he prefers or whether he is indifferent between them, and these preferences are transitive. Once a plan is settled upon, he is able to adhere to it and he can resist present temptations and distractions that interfere with its execution. These assumptions accord with the

15. On this point, see H. A. Simon, "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, vol. 69 (1955).

familiar notion of rationality that I have used all along (§25). I shall not examine here these aspects of being rational. It seems more useful to mention briefly some ways of criticizing our ends which may often help us to estimate the relative intensity of our desires. Keeping in mind that our overall aim is to carry out a rational plan (or subplan), it is clear that some features of desires make doing this impossible. For example, we cannot realize ends the descriptions of which are meaningless, or contradict well-established truths. Since π is a transcendental number, it would be pointless to try to prove that it is an algebraic number. To be sure, a mathematician in attempting to prove this proposition might discover by the way many important facts, and this achievement might redeem his efforts. But insofar as his end was to prove a falsehood, his plan would be open to criticism; and once he became aware of this, he would no longer have this aim. The same thing holds for desires that depend upon our having incorrect beliefs. It is not excluded that mistaken opinions may have a beneficial effect by enabling us to proceed with our plans, being so to speak useful illusions. Nevertheless, the desires that these beliefs support are irrational to the degree that the falsehood of these beliefs makes it impossible to execute the plan, or prevents superior plans from being adopted. (I should observe here that in the thin theory the value of knowing the facts is derived from their relation to the successful execution of rational plans. So far at least there are no grounds for attributing intrinsic value to having true beliefs.)

We may also investigate the circumstances under which we have acquired our desires and conclude that some of our aims are in various respects out of line.¹⁶ Thus a desire may spring from excessive generalization, or arise from more or less accidental associations. This is especially likely to be so in the case of aversions developed when we are younger and do not possess enough experience and maturity to make the necessary corrections. Other wants may be inordinate, having acquired their peculiar urgency as an overreaction to a prior period of severe deprivation or anxiety. The study of these processes and their disturbing influence on the normal development of our system of desires is not our concern here. They do however suggest certain critical reflections that are important devices of deliberation. Awareness of the genesis of our wants can often make it perfectly clear to us that we really do desire certain things more than others. As some aims seem less important in the face of critical scrutiny, or even lose their appeal entirely, others may assume

16. For the remarks in this paragraph, I am indebted to R. B. Brandt.

an assured prominence that provides sufficient grounds for choice. Of course, it is conceivable that despite the unfortunate conditions under which some of our desires and aversions have developed, they may still fit into and even greatly enhance the fulfillment of rational plans. If so, they turn out to be perfectly rational after all.

Finally, there are certain time-related principles that also can be used to select among plans. The principle of postponement I have already mentioned. It holds that, other things equal, rational plans try to keep our hands free until we have a clear view of the relevant facts. And the grounds for rejecting pure time preference we have also considered (§45). We are to see our life as one whole, the activities of one rational subject spread out in time. Mere temporal position, or distance from the present, is not a reason for favoring one moment over another. Future aims may not be discounted solely in virtue of being future, although we may, of course, ascribe less weight to them if there are reasons for thinking that, given their relation to other things, their fulfillment is less probable. The intrinsic importance that we assign to different parts of our life should be the same at every moment of time. These values should depend upon the whole plan itself as far as we can determine it and should not be affected by the contingencies of our present perspective.

Two other principles apply to the overall shape of plans through time. One of these is that of continuity.¹⁷ It reminds us that since a plan is a scheduled sequence of activities, earlier and later activities are bound to affect one another. The whole plan has a certain unity, a dominant theme. There is not, so to speak, a separate utility function for each period. Not only must effects between periods be taken into account, but substantial swings up and down are presumably to be avoided. A second closely related principle holds that we are to consider the advantages of rising, or at least of not significantly declining, expectations. There are various stages of life, each ideally with its own characteristic tasks and enjoyments. Other things equal, we should arrange things at the earlier stages so as to permit a happy life at the later ones. It would seem that for the most part rising expectations over time are to be preferred. If the value of an activity is assessed relative to its own period, assuming that this is possible, we might try to explain this preference by the relatively greater intensity of the pleasures of anticipation over those of memory. Even though the total sum of enjoyment is the same when enjoyments are

17. This name is taken from Jan Tinbergen, "Optimum Savings and Utility Maximization over Time," *Econometrica*, vol. 28 (1960).

estimated locally, increasing expectations provide a measure of contentment that makes the difference. But even leaving this element aside, the rising or at least the nondeclining plan appears preferable since later activities can often incorporate and bind together the results and enjoyments of an entire life into one coherent structure as those of a declining plan cannot.

In these remarks about the devices of deliberation and time-related principles I have tried to fill in Sidgwick's notion of a person's good. In brief, our good is determined by the plan of life that we would adopt with full deliberative rationality if the future were accurately foreseen and adequately realized in the imagination. The matters we have just discussed are connected with being rational in this sense. Here it is worth stressing that a rational plan is one that would be selected if certain conditions were fulfilled. The criterion of the good is hypothetical in a way similar to the criterion of justice. When the question arises as to whether doing something accords with our good, the answer depends upon how well it fits the plan that would be chosen with deliberative rationality.

Now one feature of a rational plan is that in carrying it out the individual does not change his mind and wish that he had done something else instead. A rational person does not come to feel an aversion for the foreseen consequences so great that he regrets following the plan he has adopted. The absence of this sort of regret is not however sufficient to insure that a plan is rational. There may be another plan open to us such that were we to consider it we would find it much better. Nevertheless, if our information is accurate and our understanding of the consequences complete in relevant respects, we do not regret following a rational plan, even if it is not a good one judged absolutely. In this instance the plan is objectively rational. We may, of course, regret something else, for example, that we have to live under such unfortunate circumstances that a happy life is impossible. Conceivably we may wish that we had never been born. But we do not regret that, having been born, we followed the best plan as bad as it may be when judged by some ideal standard. A rational person may regret his pursuing a subjectively rational plan, but not because he thinks his choice is in any way open to criticism. For he does what seems best at the time, and if his beliefs later prove to be mistaken with untoward results, it is through no fault of his own. There is no cause for self-reproach. There was no way of knowing which was the best or even a better plan.

Putting these reflections together, we have the guiding principle that a

rational individual is always to act so that he need never blame himself no matter how his plans finally work out. Viewing himself as one continuing being over time, he can say that at each moment of his life he has done what the balance of reasons required, or at least permitted.¹⁸ Therefore any risks he assumes must be worthwhile, so that should the worst happen that he had any reason to foresee, he can still affirm that what he did was above criticism. He does not regret his choice, at least not in the sense that he later believes that at the time it would have been more rational to have done otherwise. This principle will not certainly prevent us from taking steps that lead to misadventure. Nothing can protect us from the ambiguities and limitations of our knowledge, or guarantee that we find the best alternative open to us. Acting with deliberative rationality can only insure that our conduct is above reproach, and that we are responsible to ourselves as one person over time. We should indeed be surprised if someone said that he did not care about how he will view his present actions later any more than he cares about the affairs of other people (which is not much, let us suppose). One who rejects equally the claims of his future self and the interests of others is not only irresponsible with respect to them but in regard to his own person as well. He does not see himself as one enduring individual.

Now looked at in this way, the principle of responsibility to self resembles a principle of right: the claims of the self at different times are to be so adjusted that the self at each time can affirm the plan that has been and is being followed. The person at one time, so to speak, must not be able to complain about actions of the person at another time. This principle does not, of course, exclude the willing endurance of hardship and suffering; but it must be presently acceptable in view of the expected or achieved good. From the standpoint of the original position the relevance of responsibility to self seems clear enough. Since the notion of deliberative rationality applies there, it means that the parties cannot agree to a conception of justice if the consequences of applying it may lead to self-reproach should the least happy possibilities be realized. They should strive to be free from such regrets. And the principles of justice as fairness seem to meet this requirement better than other conceptions, as we can see from the earlier discussion of the strains of commitment (§29).

A final observation about goodness as rationality. It may be objected that this conception implies that one should be continually planning and

18. For this and other points in this paragraph see Charles Fried, *An Anatomy of Values* (Cambridge, Harvard University Press, 1970), pp. 158–169, and Thomas Nagel, *The Possibility of Altruism* (Oxford, The Clarendon Press, 1970), esp. ch. VIII.

calculating. But this interpretation rests upon a misunderstanding. The first aim of the theory is to provide a criterion for the good of the person. This criterion is defined chiefly by reference to the rational plan that would be chosen with full deliberative rationality. The hypothetical nature of the definition must be kept in mind. A happy life is not one taken up with deciding whether to do this or that. From the definition alone very little can be said about the content of a rational plan, or the particular activities that comprise it. It is not inconceivable that an individual, or even a whole society, should achieve happiness moved entirely by spontaneous inclination. With great luck and good fortune some men might by nature just happen to hit upon the way of living that they would adopt with deliberative rationality. For the most part, though, we are not so blessed, and without taking thought and seeing ourselves as one person with a life over time, we shall almost certainly regret our course of action. Even when a person does succeed in relying on his natural impulses without misadventure, we still require a conception of his good in order to assess whether he has really been fortunate or not. He may think so, but he may be deluded; and to settle this matter, we have to examine the hypothetical choices that it would have been rational for him to make, granting due allowance for whatever benefits he may have obtained from not worrying about these things. As I noted before, the value of the activity of deciding is itself subject to rational appraisal. The efforts we should expend making decisions will depend like so much else on circumstances. Goodness as rationality leaves this question to the person and the contingencies of his situation.

65. THE ARISTOTELIAN PRINCIPLE

The definition of the good is purely formal. It simply states that a person's good is determined by the rational plan of life that he would choose with deliberative rationality from the maximal class of plans. Although the notion of deliberative rationality and the principles of rational choice rely upon concepts of considerable complexity, we still cannot derive from the definition of rational plans alone what sorts of ends these plans are likely to encourage. In order to draw conclusions about these ends, it is necessary to take note of certain general facts.

First of all, there are the broad features of human desires and needs, their relative urgency and cycles of recurrence, and their phases of development as affected by physiological and other circumstances. Second,

plans must fit the requirements of human capacities and abilities, their trends of maturation and growth, and how they are best trained and educated for this or that purpose. Moreover, I shall postulate a basic principle of motivation which I shall refer to as the Aristotelian Principle. Finally, the general facts of social interdependency must be reckoned with. The basic structure of society is bound to encourage and support certain kinds of plans more than others by rewarding its members for contributing to the common good in ways consistent with justice. Taking account of these contingencies narrows down the alternative plans so that the problem of decision becomes, in some cases anyway, reasonably definite. To be sure, as we shall see, a certain arbitrariness still remains, but the priority of right limits it in such a way that it is no longer a problem from the standpoint of justice (§68).

The general facts about human needs and abilities are perhaps clear enough and I shall assume that common sense knowledge suffices for our purposes here. Before taking up the Aristotelian Principle, however, I should comment briefly on the human goods (as I shall call them) and the constraints of justice. Given the definition of a rational plan, we may think of these goods as those activities and ends that have the features whatever they are that suit them for an important if not a central place in our life.¹⁹ Since in the full theory rational plans must be consistent with the principles of justice, the human goods are similarly constrained. Thus the familiar values of personal affection and friendship, meaningful work and social cooperation, the pursuit of knowledge and the fashioning and contemplation of beautiful objects, are not only prominent in our rational plans but they can for the most part be advanced in a manner which justice permits. Admittedly to attain and to preserve these values, we are often tempted to act unjustly; but achieving these ends involves no inherent injustice. In contrast with the desire to cheat and to degrade others, doing something unjust is not included in the description of the human goods (§66).

The social interdependency of these values is shown in the fact that not only are they good for those who enjoy them but they are likely to enhance the good of others. In achieving these ends we generally contribute to the rational plans of our associates. In this sense, they are complementary goods, and this accounts for their being singled out for special commendation. For to commend something is to praise it, to recount the

19. For the explanation of these goods I have drawn from C. A. Campbell, "Moral and Non-Moral Values," *Mind*, vol. 44 (1935); see pp. 279–291.

properties that make it good (rational to want) with emphasis and expressions of approval. These facts of interdependency are further reasons for including the recognized values in long-term plans. For assuming that we desire the respect and good will of other persons, or at least to avoid their hostility and contempt, those plans of life will tend to be preferable which further their aims as well as our own.

Turning now to our present topic, it will be recalled that the Aristotelian Principle runs as follows: other things equal, human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and this enjoyment increases the more the capacity is realized, or the greater its complexity.²⁰ The intuitive idea here is that human beings take more pleasure in doing something as they become more proficient at it, and of two activities they do equally well, they prefer the one calling on a larger repertoire of more intricate and subtle discriminations. For example, chess is a more complicated and subtle game than checkers, and algebra is more intricate than elementary arithmetic. Thus the principle says that someone who can do both generally prefers playing chess to playing checkers, and that he would rather study algebra than arithmetic. We need not explain here why the Aristotelian Principle is true. Presumably complex activities are more enjoyable because they satisfy the desire for variety and novelty of experience, and leave room for feats of ingenuity and invention. They also evoke the pleasures of anticipation and surprise, and often the overall form of the activity, its structural develop-

20. The name "Aristotelian Principle" seems to me appropriate in view of what Aristotle says about the relations between happiness, activity, and enjoyment in the *Nicomachean Ethics*, bk. VII, chs. 11–14, and bk. X, chs. 1–5. Yet since he does not state such a principle explicitly, and some of it is at best only implied, I have not called it "Aristotle's Principle." Nevertheless, Aristotle certainly affirms two points that the principle conveys: (1) that enjoyment and pleasure are not always by any means the result of returning to a healthy or normal state, or of making up deficiencies; rather many kinds of pleasure and enjoyment arise when we exercise our faculties; and (2) that the exercise of our natural powers is a leading human good. Further, (3) the idea that the more enjoyable activities and the more desirable and enduring pleasures spring from the exercise of greater abilities involving more complex discriminations is not only compatible with Aristotle's conception of the natural order, but something like it usually fits the judgments of value he makes, even when it does not express his reasons. For a discussion of Aristotle's account of enjoyment and pleasure, see W. F. R. Hardie, *Aristotle's Ethical Theory* (Oxford, The Clarendon Press, 1968), ch. XIV. The interpretation of Aristotle's doctrine given by G. C. Field, *Moral Theory* (London, Methuen, 1932), pp. 76–78, strongly suggests what I have called the Aristotelian Principle. Mill comes very close to stating it in *Utilitarianism*, ch. II, pars. 4–8. Important here is the concept of effectance motivation introduced by R. W. White, "Ego and Reality in Psychoanalytic Theory," *Psychological Issues*, vol. III, (1963), ch. III, upon which I have relied. See also pp. 173–175, 180f. I am indebted to J. M. Cooper for discussion on the interpretation of this principle and the propriety of its name.

ment, is fascinating and beautiful. Moreover, simpler activities exclude the possibility of individual style and personal expression which complex activities permit or even require, for how could everyone do them in the same way? That we should follow our natural bent and the lessons of our past experience seems inevitable if we are to find our way at all. Each of these features is well illustrated by chess, even to the point where grand masters have their characteristic style of play. Whether these considerations are explanations of the Aristotelian Principle or elaboration of its meaning, I shall leave aside. I believe that nothing essential for the theory of the good depends upon this question.

It is evident that the Aristotelian Principle contains a variant of the principle of inclusiveness. Or at least the clearest cases of greater complexity are those in which one of the activities to be compared includes all the skills and discriminations of the other activity and some further ones in addition. Once again, we can establish but a partial order, since each of several activities may require abilities not used in the others. Such an ordering is the best that we can have until we possess some relatively precise theory and measure of complexity that enables us to analyze and compare seemingly disparate activities. I shall not, however, discuss this problem here, but assume instead that our intuitive notion of complexity will suffice for our purposes.

The Aristotelian Principle is a principle of motivation. It accounts for many of our major desires, and explains why we prefer to do some things and not others by constantly exerting an influence over the flow of our activity. Moreover, it expresses a psychological law governing changes in the pattern of our desires. Thus the principle implies that as a person's capacities increase over time (brought about by physiological and biological maturation, for example, the development of the nervous system in a young child), and as he trains these capacities and learns how to exercise them, he will in due course come to prefer the more complex activities that he can now engage in which call upon his newly realized abilities. The simpler things he enjoyed before are no longer sufficiently interesting or attractive. If we ask why we are willing to undergo the stresses of practice and learning, the reason may be (if we leave out of account external rewards and penalties) that having had some success at learning things in the past, and experiencing the present enjoyments of the activity, we are led to expect even greater satisfaction once we acquire a greater repertoire of skills. There is also a companion effect to the Aristotelian Principle. As we witness the exercise of well-trained abilities

by others, these displays are enjoyed by us and arouse a desire that we should be able to do the same things ourselves. We want to be like those persons who can exercise the abilities that we find latent in our nature.

Thus it would appear that how much we learn and how far we educate our innate capacities depends upon how great these capacities are and how difficult is the effort of realizing them. There is a race, so to speak, between the increasing satisfaction of exercising greater realized ability and the increasing strains of learning as the activity becomes more strenuous and difficult. Assuming that natural talents have an upper bound, whereas the hardships of training can be made more severe without limit, there must be some level of achieved ability beyond which the gains from a further increase in this level are just offset by the burdens of the further practice and study necessary to bring it about and to maintain it. Equilibrium is reached when these two forces balance one another, and at this point the effort to achieve greater realized capacity ceases. It follows that if the pleasures of the activity increase too slowly with rising ability (an index let us suppose of a lower level of innate ability), then the correspondingly greater efforts of learning will lead us to give up sooner. In this case we will never engage in certain more complex activities nor acquire the desires evoked by taking part in them.

Now accepting the Aristotelian Principle as a natural fact, it will generally be rational, in view of the other assumptions, to realize and train mature capacities. Maximal or satisfactory plans are almost certainly plans that provide for doing this in significant measure. Not only is there a tendency in this direction postulated by the Aristotelian Principle, but the plain facts of social interdependency and the nature of our interests more narrowly construed incline us in the same way. A rational plan—constrained as always by the principles of right—allows a person to flourish, so far as circumstances permit, and to exercise his realized abilities as much as he can. Moreover, his fellow associates are likely to support these activities as promoting the common interest and also to take pleasure in them as displays of human excellence. To the degree, then, that the esteem and admiration of others is desired, the activities favored by the Aristotelian Principle are good for other persons as well.

There are several points to keep in mind in order to prevent misunderstandings of this principle. For one thing, it formulates a tendency and not an invariable pattern of choice, and like all tendencies it may be overridden. Countervailing inclinations can inhibit the development of realized capacity and the preference for more complex activities. Various hazards and risks, both psychological and social, are involved in training and

prospective accomplishment, and apprehensions about these may outweigh the original propensity. We must interpret the principle so as to allow for these facts. Yet if it is a useful theoretical notion, the tendency postulated should be relatively strong and not easily counterbalanced. I believe that this is indeed the case, and that in the design of social institutions a large place has to be made for it, otherwise human beings will find their culture and form of life dull and empty. Their vitality and zest will fail as their life becomes a tiresome routine. And this seems borne out by the fact that the forms of life which absorb men's energies, whether they be religious devotions or purely practical matters or even games and pastimes, tend to develop their intricacies and subtleties almost without end. As social practices and cooperative activities are built up through the imagination of many individuals, they increasingly call forth a more complex array of abilities and new ways of doing things. That this process is carried along by the enjoyment of natural and free activity seems to be verified by the spontaneous play of children and animals which shows all the same features.

A further consideration is that the principle does not assert that any particular kind of activity will be preferred. It says only that we prefer, other things equal, activities that depend upon a larger repertoire of realized capacities and that are more complex. More precisely, suppose that we can order a certain number of activities in a chain by the inclusion relation. This means that the n th activity exercises all the skills of the $n-1$ th activity and some further ones in addition. Now there are indefinitely many such chains with no elements in common, let us say; and moreover, numerous chains may start from the same activity representing different ways in which this activity can be built upon and enriched. What the Aristotelian Principle says is that whenever a person engages in an activity belonging to some chain (and perhaps to several chains) he tends to move up the chain. In general, he will prefer doing the n th to doing the $n-1$ th activity; and this tendency will be stronger the more his capacity is yet to be realized and the less onerous he finds the strains of learning and training. Presumably there is a preference for ascending the chain or chains which offer the greatest prospects of exercising the higher abilities with the least stress. The actual course that a person follows, the combination of activities that he finds most appealing, is decided by his inclinations and talents and by his social circumstances, by what his associates appreciate and are likely to encourage. Thus natural assets and social opportunities obviously influence the chains that individuals eventually prefer. By itself the principle simply asserts a propensity to ascend what-

ever chains are chosen. It does not entail that a rational plan includes any particular aims, nor does it imply any special form of society.

Again, we may suppose, although it is probably not essential, that every activity belongs to some chain. The reason for this is that human ingenuity can and normally will discover for each activity a continuing chain that elicits a growing inventory of skills and discriminations. We stop moving up a chain, however, when going higher will use up resources required for raising or for maintaining the level of a preferred chain. And resources here is to be taken broadly, so that among the most important ones are time and energy. This is the reason why, for example, we are content to lace our shoes or to tie our tie in a straightforward way, and do not ordinarily make complex rituals of these daily actions. There are only so many hours in a day, and this prevents our ascending to the upper limits of our capacity all the chains that are open to us. But then a prisoner in a cell might take time with daily routines and invent ways of doing them that he would not otherwise bother with. The formal criterion is that a rational individual selects a preferred pattern of activities (compatible with the principle of justice) and proceeds along each of its chains up to the point where no further improvement results from any feasible change in the schedule. This overall standard does not, of course, tell us how to decide; rather it emphasizes the limited resources of time and energy, and explains why some activities are slighted in favor of others even though, in the form in which we engage in them, they allow for further elaboration.

Now it may be objected that there is no reason to suppose that the Aristotelian Principle is true. Like the idealist notion of self-realization, to which it bears a certain resemblance, it may have the ring of a philosopher's principle with little to support it. But it seems to be borne out by many facts of everyday life, and by the behavior of children and some of the higher animals. Moreover, it appears to be susceptible to an evolutionary explanation. Natural selection must have favored creatures of whom this principle is true. Aristotle says that men desire to know. Presumably we have acquired this desire by a natural development, and indeed, if the principle is sound, a desire to engage in more complex and demanding activities of any kind as long as they are within our reach.²¹ Human beings enjoy the greater variety of experience, they take pleasure in the

21. See B. G. Campbell, *Human Evolution* (Chicago, Aldine Publishing Co., 1966), pp. 49–53; and W. H. Thorpe, *Science, Man, and Morals* (London, Methuen, 1965), pp. 87–92. For animals see Irenäus Eibl-Eibesfeldt, *Ethology*, trans. Erich Klinghammer (New York, Holt, Rinehart, and Winston, 1970), pp. 217–248.

novelty and surprises and the occasions for ingenuity and invention that such activities provide. The multiplicity of spontaneous activities is an expression of the delight that we take in imagination and creative fantasy. Thus the Aristotelian Principle characterizes human beings as importantly moved not only by the pressure of bodily needs, but also by the desire to do things enjoyed simply for their own sakes, at least when the urgent and pressing wants are satisfied. The marks of such enjoyed activities are many, varying from the manner and way in which they are done to the persistence with which they are returned to at a later time. Indeed, we do them without the incentive of evident reward, and allowing us to engage in them can itself act often as a reward for doing other things.²² Since the Aristotelian Principle is a feature of human desires as they now exist, rational plans must take it into account. The evolutionary explanation, even if it is correct, is not of course a justification for this aspect of our nature. In fact, the question of justification does not arise. The question is rather: granted that this principle characterizes human nature as we know it, to what extent is it to be encouraged and supported, and how is it to be reckoned with in framing rational plans of life?

The role of the Aristotelian Principle in the theory of the good is that it states a deep psychological fact which, in conjunction with other general facts and the conception of a rational plan, accounts for our considered judgments of value. The things that are commonly thought of as human goods should turn out to be the ends and activities that have a major place in rational plans. The principle is part of the background that regulates these judgments. Provided that it is true, and leads to conclusions matching our convictions about what is good and bad (in reflective equilibrium), it has a proper place in moral theory. Even if this conception should not be true of some persons, the idea of a rational long-term plan still applies. We can work out what is good for them in much the same way as before. Thus imagine someone whose only pleasure is to count blades of grass in various geometrically shaped areas such as park squares and well-trimmed lawns. He is otherwise intelligent and actually possesses unusual skills, since he manages to survive by solving difficult mathematical problems for a fee. The definition of the good forces us to admit that the good for this man is indeed counting blades of grass, or more accurately, his good is determined by a plan that gives an especially prominent place to this activity. Naturally we would be surprised that such a person should exist. Faced with his case, we would try out other

22. This seems also to be true of monkeys. See Eibl-Eibesfeldt, *ibid.*, p. 239.

hypotheses. Perhaps he is peculiarly neurotic and in early life acquired an aversion to human fellowship, and so he counts blades of grass to avoid having to deal with other people. But if we allow that his nature is to enjoy this activity and not to enjoy any other, and that there is no feasible way to alter his condition, then surely a rational plan for him will center around this activity. It will be for him the end that regulates the schedule of his actions, and this establishes that it is good for him. I mention this fanciful case only to show that the correctness of the definition of a person's good in terms of the rational plan for him does not require the truth of the Aristotelian Principle. The definition is satisfactory, I believe, even if this principle should prove inaccurate, or fail altogether. But by assuming the principle we seem able to account for what things are recognized as good for human beings taking them as they are. Moreover, since this principle ties in with the primary good of self-respect, it turns out to have a central position in the moral psychology underlying justice as fairness (§67).

66. THE DEFINITION OF GOOD APPLIED TO PERSONS

Having defined a person's good as the successful execution of a rational plan of life, and his lesser goods as parts thereof, we are in a position to introduce further definitions. In this way the concept of goodness is applied to other subjects that have an important place in moral philosophy. But before doing this we should note the assumption that the primary goods can be accounted for by the thin theory of the good. That is, I suppose that it is rational to want these goods whatever else is wanted, since they are in general necessary for the framing and the execution of a rational plan of life. The persons in the original position are assumed to accept this conception of the good, and therefore they take for granted that they desire greater liberty and opportunity, and more extensive means for achieving their ends. With these objectives in mind, as well as that of securing the primary good of self-respect (§67), they evaluate the conceptions of justice available to them in the original position.

That liberty and opportunity, income and wealth, and above all self-respect are primary goods must indeed be explained by the thin theory. The constraints of the principles of justice cannot be used to draw up the list of primary goods that serves as part of the description of the initial situation. The reason is, of course, that this list is one of the premises from which the choice of the principles of right is derived. To cite these

principles in explaining the list would be a circular argument. We must assume, then, that the list of primary goods can be accounted for by the conception of goodness as rationality in conjunction with the general facts about human wants and abilities, their characteristic phases and requirements of nurture, the Aristotelian Principle, and the necessities of social interdependence. At no point can we appeal to the constraints of justice. But once we are satisfied that the list of primary goods can be arrived at in this way, then in all further applications of the definition of good the constraints of right may be freely invoked. I shall not argue the case for the list of primary goods here, since their claims seem evident enough. I shall, however, come back to this point from time to time, especially in connection with the primary good of self-respect. In what follows I take the list as established and apply the full theory of the good. The test of this theory is that it should fit our considered judgments of value in reflective equilibrium.

Two fundamental cases for the theory of the good remain to be considered. We must see whether the definition holds for both persons and societies. In this section I discuss the case of persons, leaving the question of a good society for the last chapter when all parts of justice as fairness can be brought to bear. Now many philosophers have been willing to accept some variant of goodness as rationality for artifacts and roles, and for such nonmoral values as friendship and affection, the pursuit of knowledge and the enjoyment of beauty, and the like. Indeed, I have emphasized that the main elements of goodness as rationality are extremely common, being shared by philosophers of markedly different persuasions. Nevertheless, it is often thought that this conception of the good expresses an instrumental or economic theory of value that does not hold for the case of moral worth. When we speak of the just or the benevolent person as morally good, a different concept of goodness is said to be involved.²³ I wish to argue, however, that once the principles of right and justice are on hand, the full theory of goodness as rationality can in fact cover these judgments. The reason why the so-called instrumental or economic theory fails is that what is in effect the thin theory is applied directly to the problem of moral worth. What we must do instead is to use this theory only as a part of the description of the original position from which the principles of right and justice are derived. We can then apply the full theory of the good without restrictions and are free to use it for

23. See C. A. Campbell, "Moral and Non-Moral Values," *Mind*, vol. 44 (1935); and R. M. Hare, "Geach on Good and Evil," *Analysis*, vol. 18 (1957).

the two basic cases of a good person and a good society. Developing the thin into the full theory via the original position is the essential step.

Several ways suggest themselves for extending the definition to the problem of moral worth, and I believe that at least one of these will serve well enough. First of all, we might identify some basic role or position, say that of citizen, and then say that a good person is one who has to a higher degree than the average the properties which it is rational for citizens to want in one another. Here the relevant point of view is that of a citizen judging other citizens in the same role. Second, the notion of a good person could be interpreted as requiring some general or average assessment so that a good person is one who performs well in his various roles, especially those that are considered more important. Finally, there may exist properties which it is rational to want in persons when they are viewed with respect to almost any of their social roles. Let us say that such properties, if they exist, are broadly based.²⁴ To illustrate this idea in the case of tools, the broadly based properties are efficiency, durability, ease of maintenance, and so on. These features are desirable in tools of almost any kind. Much less broadly based properties are properties such as keeps its cutting edge, does not rust, and so on. The question whether some tools have these would not even arise. By analogy, a good person, in contrast to a good doctor or a good farmer, and the like, is one who has to a higher degree than the average person the broadly based properties (yet to be specified) that it is rational for persons to want in one another.

Offhand it seems that the last suggestion is the most plausible one. It can be made to include the first as a special case and to capture the intuitive idea of the second. There are, however, certain complications in working it out. The first thing is to identify the point of view from which the broadly based properties are rationally preferred and the assumptions upon which this preference is founded. I note straightway that the fundamental moral virtues, that is, the strong and normally effective desires to act on the basic principles of right, are undoubtedly among the broadly based properties. At any rate, this seems bound to be true so long as we suppose that we are considering a well-ordered society, or one in a state of near justice, as I shall indeed take to be the case. Now since the basic structure of such a society is just, and these arrangements are stable with respect to the society's public conception of justice, its members will in general have the appropriate sense of justice and a desire to see their institutions affirmed. But it is also true that it is rational for each person to

24. For the notion of broadly based properties and its use here, I am indebted to T. M. Scanlon.

act on the principles of justice only on the assumption that for the most part these principles are recognized and similarly acted upon by others. Therefore the representative member of a well-ordered society will find that he wants others to have the basic virtues, and in particular a sense of justice. His rational plan of life is consistent with the constraints of right, and he will surely want others to acknowledge the same restrictions. In order to make this conclusion absolutely firm, we should also like to be sure that it is rational for those belonging to a well-ordered society who have already acquired a sense of justice to maintain and even to strengthen this moral sentiment. I shall discuss this question later (§86); for the present I suppose that it is the case. Thus with all these presumptions on hand, it seems clear that the fundamental virtues are among the broadly based properties that it is rational for members of a well-ordered society to want in one another.

A further complication must be considered. There are other properties that are presumably as broadly based as the virtues, for example, intelligence and imagination, strength and endurance. Indeed, a certain minimum of these attributes is necessary for right conduct, since without judgment and imagination, say, benevolent intentions may easily lead to harm. On the other hand, unless intellect and vigor are regulated by a sense of justice and obligation, they may only enhance one's capacity to override the legitimate claims of others. Certainly it would not be rational to want some to be so superior in these respects that just institutions would be jeopardized. Yet the possession of these natural assets in the appropriate degree is clearly desirable from a social point of view; and therefore within limits these attributes are also broadly based. Thus while the moral virtues are included in the broadly based properties, they are not the only ones in this class.

It is necessary, then, to distinguish the moral virtues from the natural assets. The latter we may think of as natural powers developed by education and training, and often exercised in accordance with certain characteristic intellectual or other standards by reference to which they can be roughly measured. The virtues on the other hand are sentiments and habitual attitudes leading us to act on certain principles of right. We can distinguish the virtues from each other by means of their corresponding principles. I assume, then, that the virtues can be singled out by using the conception of justice already established; once this conception is understood, we can rely on it to define the moral sentiments and to mark them off from the natural assets.

A good person, then, or a person of moral worth, is someone who has

to a higher degree than the average the broadly based features of moral character that it is rational for the persons in the original position to want in one another. Since the principles of justice have been chosen, and we are assuming strict compliance, each knows that in society he will want the others to have the moral sentiments that support adherence to these standards. Thus we could say alternatively that a good person has the features of moral character that it is rational for members of a well-ordered society to want in their associates. Neither of these interpretations introduces any new ethical notions, and so the definition of goodness as rationality has been extended to persons. In conjunction with the theory of justice which has the thin account of the good as a subpart, the full theory seems to give a satisfactory rendering of moral worth, the third main concept of ethics.

Some philosophers have thought that since a person *qua* person has no definite role or function, and is not to be treated as an instrument or object, a definition along the lines of goodness as rationality must fail.²⁵ But as we have seen, it is possible to develop a definition of this sort without supposing that persons hold some particular role, much less that they are things to be used for some ulterior purpose. It is true, of course, that the extension of the definition to the case of moral worth makes many assumptions. In particular, I assume that being a member of some community and engaging in many forms of cooperation is a condition of human life. But this presumption is sufficiently general so as not to compromise a theory of justice and moral worth. Indeed, it is entirely proper, as I have noted previously, that an account of our considered moral judgments should draw upon the natural circumstances of society. In this sense there is nothing *a priori* about moral philosophy. It suffices to recall by way of summation that what permits this definition of the good to cover the notion of moral worth is the use of the principles of justice already derived. Moreover, the specific content and mode of derivation of these principles is also relevant. The main idea of justice as fairness, that the principles of justice are those that would be agreed to by rational persons in an original position of equality, prepares the way for extending the definition of good to the larger questions of moral goodness.

It seems desirable to indicate the way in which the definition of good might be extended to other cases. Doing this will give us more confidence in its application to persons. Thus let us suppose that for each person

25. See, for example, Hare, "Geach on Good and Evil," pp. 109ff.

there is a rational plan of life that determines his good. We can now define a good act (in the sense of a beneficent act) as one which we are at liberty to do or not to do, that is, no requirements of natural duty or obligation constrain us either to do it or not to do it, and which advances and is intended to advance another's good (his rational plan). Taking a further step, we can define a good action (in the sense of a benevolent action) as a good act performed for the sake of the other person's good. A beneficent act promotes another's good; and a benevolent action is done from the desire that the other should have this good. When the benevolent action is one that brings much good for the other person and when it is undertaken at considerable loss or risk to the agent as estimated by his interests more narrowly construed, then the action is supererogatory. An act which would be very good for another, especially one which protects him from great harm or injury, is a natural duty required by the principle of mutual aid, provided that the sacrifice and hazards to the agent are not very great. Thus a supererogatory act may be thought of as one which a person does for the sake of another's good even though the proviso that nullifies the natural duty is satisfied. In general, supererogatory actions are ones that would be duties were not certain exempting conditions fulfilled which make allowance for reasonable self-interest. Eventually, of course, for a complete contractarian account of right, we would have to work out from the standpoint of the original position what is to count as reasonable self-interest. But I shall not pursue this question here.

Finally, the full theory of the good enables us to distinguish different sorts of moral worth, or the lack of it. Thus we can distinguish between the unjust, the bad, and the evil man. To illustrate, consider the fact that some men strive for excessive power, that is, authority over others which goes beyond what is allowed by the principles of justice and which can be exercised arbitrarily. In each of these cases there is a willingness to do what is wrong and unjust in order to achieve one's ends. But the unjust man seeks dominion for the sake of aims such as wealth and security which when appropriately limited are legitimate. The bad man desires arbitrary power because he enjoys the sense of mastery which its exercise gives to him and he seeks social acclaim. He too has an inordinate desire for things which when duly circumscribed are good, namely, the esteem of others and the sense of self-command. It is his way of satisfying these ambitions that makes him dangerous. By contrast, the evil man aspires to unjust rule precisely because it violates what independent persons would consent to in an original position of equality, and therefore its possession and display manifest his superiority and affront the self-

respect of others. It is this display and affront which is sought after. What moves the evil man is the love of injustice: he delights in the impotence and humiliation of those subject to him and he relishes being recognized by them as the willful author of their degradation. Once the theory of justice is joined to the theory of the good in what I have called the full theory, we can make these and other distinctions. There seems to be no reason to fear that the numerous variations of moral worth cannot be accounted for.

67. SELF-RESPECT, EXCELLENCES, AND SHAME

On several occasions I have mentioned that perhaps the most important primary good is that of self-respect. We must make sure that the conception of goodness as rationality explains why this should be so. We may define self-respect (or self-esteem) as having two aspects. First of all, as we noted earlier (§29), it includes a person's sense of his own value, his secure conviction that his conception of his good, his plan of life, is worth carrying out. And second, self-respect implies a confidence in one's ability, so far as it is within one's power, to fulfill one's intentions. When we feel that our plans are of little value, we cannot pursue them with pleasure or take delight in their execution. Nor plagued by failure and self-doubt can we continue in our endeavors. It is clear then why self-respect is a primary good. Without it nothing may seem worth doing, or if some things have value for us, we lack the will to strive for them. All desire and activity becomes empty and vain, and we sink into apathy and cynicism. Therefore the parties in the original position would wish to avoid at almost any cost the social conditions that undermine self-respect. The fact that justice as fairness gives more support to self-esteem than other principles is a strong reason for them to adopt it.

The conception of goodness as rationality allows us to characterize more fully the circumstances that support the first aspect of self-esteem, the sense of our own worth. These are essentially two: (1) having a rational plan of life, and in particular one that satisfies the Aristotelian Principle; and (2) finding our person and deeds appreciated and confirmed by others who are likewise esteemed and their association enjoyed. I assume then that someone's plan of life will lack a certain attraction for him if it fails to call upon his natural capacities in an interesting fashion. When activities fail to satisfy the Aristotelian Principle, they are

likely to seem dull and flat, and to give us no feeling of competence or a sense that they are worth doing. A person tends to be more confident of his value when his abilities are both fully realized and organized in ways of suitable complexity and refinement.

But the companion effect of the Aristotelian Principle influences the extent to which others confirm and take pleasure in what we do. For while it is true that unless our endeavors are appreciated by our associates it is impossible for us to maintain the conviction that they are worthwhile, it is also true that others tend to value them only if what we do elicits their admiration or gives them pleasure. Thus activities that display intricate and subtle talents, and manifest discrimination and refinement, are valued by both the person himself and those around him. Moreover the more someone experiences his own way of life as worth fulfilling, the more likely he is to welcome our attainments. One who is confident in himself is not grudging in the appreciation of others. Putting these remarks together, the conditions for persons respecting themselves and one another would seem to require that their common plans be both rational and complementary: they call upon their educated endowments and arouse in each a sense of mastery, and they fit together into one scheme of activity that all can appreciate and enjoy.

Now it may be thought that these stipulations cannot be generally satisfied. One might suppose that only in a limited association of highly gifted individuals united in the pursuit of common artistic, scientific, or social ends is anything of this sort possible. There would seem to be no way to establish an enduring basis of self-respect throughout society. Yet this surmise is mistaken. The application of the Aristotelian Principle is always relative to the individual and therefore to his natural assets and particular situation. It normally suffices that for each person there is some association (one or more) to which he belongs and within which the activities that are rational for him are publicly affirmed by others. In this way we acquire a sense that what we do in everyday life is worthwhile. Moreover, associative ties strengthen the second aspect of self-esteem, since they tend to reduce the likelihood of failure and to provide support against the sense of self-doubt when mishaps occur. To be sure, men have varying capacities and abilities, and what seems interesting and challenging to some will not seem so to others. Yet in a well-ordered society anyway, there are a variety of communities and associations, and the members of each have their own ideals appropriately matched to their aspirations and talents. Judged by the doctrine of perfectionism, the ac-

tivities of many groups may not display a high degree of excellence. But no matter. What counts is that the internal life of these associations is suitably adjusted to the abilities and wants of those belonging to them, and provides a secure basis for the sense of worth of their members. The absolute level of achievement, even if it could be defined, is irrelevant. But in any case, as citizens we are to reject the standard of perfection as a political principle, and for the purposes of justice avoid any assessment of the relative value of one another's way of life (§50). Thus what is necessary is that there should be for each person at least one community of shared interests to which he belongs and where he finds his endeavors confirmed by his associates. And for the most part this assurance is sufficient whenever in public life citizens respect one another's ends and adjudicate their political claims in ways that also support their self-esteem. It is precisely this background condition that is maintained by the principles of justice. The parties in the original position do not adopt the principle of perfection, for rejecting this criterion prepares the way to recognize the good of all activities that fulfill the Aristotelian Principle (and are compatible with the principles of justice). This democracy in judging each other's aims is the foundation of self-respect in a well-ordered society.

Later on I shall relate these matters to the idea of social union and the place of the principles of justice in human good (§§79–82). Here I wish to discuss the connections between the primary good of self-respect, the excellences, and shame, and consider when shame is a moral as opposed to a natural emotion. Now we may characterize shame as the feeling that someone has when he experiences an injury to his self-respect or suffers a blow to his self-esteem. Shame is painful since it is the loss of a prized good. There is a distinction however between shame and regret that should be noted. The latter is a feeling occasioned by the loss of most any sort of good, as when we regret having done something either imprudently or inadvertently that resulted in harm to ourselves. In explaining regret we focus say on the opportunities missed or the means squandered. Yet we may also regret having done something that put us to shame, or even having failed to carry out a plan of life that established a basis for our self-esteem. Thus we may regret the lack of a sense of our own worth. Regret is the general feeling aroused by the loss or absence of what we think good for us, whereas shame is the emotion evoked by shocks to our self-respect, a special kind of good.

Now both regret and shame are self-regarding, but shame implies an

especially intimate connection with our person and with those upon whom we depend to confirm the sense of our own worth.²⁶ Also, shame is sometimes a moral feeling, a principle of right being cited to account for it. We must find an explanation of these facts. Let us distinguish between things that are good primarily for us (for the one who possesses them) and attributes of our person that are good both for us and for others as well. These two classes are not exhaustive but they indicate the relevant contrast. Thus commodities and items of property (exclusive goods) are goods mainly for those who own them and have use of them, and for others only indirectly. On the other hand, imagination and wit, beauty and grace, and other natural assets and abilities of the person are goods for others too: they are enjoyed by our associates as well as ourselves when properly displayed and rightly exercised. They form the human means for complementary activities in which persons join together and take pleasure in their own and one another's realization of their nature. This class of goods constitutes the excellences: they are the characteristics and abilities of the person that it is rational for everyone (including ourselves) to want us to have. From our standpoint, the excellences are goods since they enable us to carry out a more satisfying plan of life enhancing our sense of mastery. At the same time these attributes are appreciated by those with whom we associate, and the pleasure they take in our person and in what we do supports our self-esteem. Thus the excellences are a condition of human flourishing; they are goods from everyone's point of view. These facts relate them to the conditions of self-respect, and account for their connection with our confidence in our own value.

Considering first natural shame, it arises not from a loss or absence of exclusive goods, or at least not directly, but from the injury to our self-esteem owing to our not having or failing to exercise certain excellences. The lack of things primarily good for us would be an occasion for regret but not for shame. Thus one may be ashamed of his appearance or slow-wittedness. Normally these attributes are not voluntary and so they

26. My definition of shame is close to William McDougall, *An Introduction to Social Psychology* (London, Methuen, 1908), pp. 124–128. On the connection between self-esteem and what I have called the Aristotelian Principle, I have followed White, "Ego and Reality in Psychoanalytic Theory," ch.7. On the relation of shame to guilt, I am indebted to Gerhart Piers and Milton Singer, *Shame and Guilt* (Springfield, Ill., Charles C. Thomas, 1953), though the setting of my discussion is quite different. See also Erik Erikson, "Identity and the Life Cycle," *Psychological Issues*, vol. 1 (1959), pp. 39–41, 65–70. For the intimacy of shame, see Stanley Cavell, "The Avoidance of Love," in *Must We Mean What We Say?* (New York, Charles Scribner's Sons, 1969), pp. 278, 286f.

do not render us blameworthy; yet given the tie between shame and self-respect, the reason for being downcast by them is straightforward. With these defects our way of life is often less fulfilling and we receive less appreciative support from others. Thus natural shame is aroused by blemishes in our person, or by acts and attributes indicative thereof, that manifest the loss or lack of properties that others as well as ourselves would find it rational for us to have. However, a qualification is necessary. It is our plan of life that determines what we feel ashamed of, and so feelings of shame are relative to our aspirations, to what we try to do and with whom we wish to associate.²⁷ Those with no musical ability do not strive to be musicians and feel no shame for this lack. Indeed it is no lack at all, not at least if satisfying associations can be formed by doing other things. Thus we should say that given our plan of life, we tend to be ashamed of those defects in our person and failures in our actions that indicate a loss or absence of the excellences essential to our carrying out our more important associative aims.

Turning now to moral shame, we have only to put together the account of the notion of a good person (in the previous section) and the remarks above concerning the nature of shame. Thus someone is liable to moral shame when he prizes as excellences of his person those virtues that his plan of life requires and is framed to encourage. He regards the virtues, or some of them anyway, as properties that his associates want in him and that he wants in himself. To possess these excellences and to express them in his actions are among his regulative aims and are felt to be a condition of his being valued and esteemed by those with whom he cares to associate. Actions and traits that manifest or betray the absence of these attributes in his person are likely then to occasion shame, and so is the awareness or recollection of these defects. Since shame springs from a feeling of the diminishment of self, we must explain how moral shame can be so regarded. First of all, the Kantian interpretation of the original position means that the desire to do what is right and just is the main way for persons to express their nature as free and equal rational beings. And from the Aristotelian Principle it follows that this expression of their nature is a fundamental element of their good. Combined with the account of moral worth, we have, then, that the virtues are excellences. They are good from the standpoint of ourselves as well as from that of others. The lack of them will tend to undermine both our self-esteem

27. See William James, *The Principles of Psychology*, vol. I (New York, 1890), pp. 309f.

and the esteem that our associates have for us. Therefore indications of these faults will wound one's self-respect with accompanying feelings of shame.

It is instructive to observe the differences between the feelings of moral shame and guilt. Although both may be occasioned by the same action, they do not have the same explanation (§73). Imagine for example someone who cheats or gives in to cowardice and then feels both guilty and ashamed. He feels guilty because he has acted contrary to his sense of right and justice. By wrongly advancing his interests he has transgressed the rights of others, and his feelings of guilt will be more intense if he has ties of friendship and association to the injured parties. He expects others to be resentful and indignant at his conduct, and he fears their righteous anger and the possibility of reprisal. Yet he also feels ashamed because his conduct shows that he has failed to achieve the good of self-command, and he has been found unworthy of his associates upon whom he depends to confirm his sense of his own worth. He is apprehensive lest they reject him and find him contemptible, an object of ridicule. In his behavior he has betrayed a lack of the moral excellences he prizes and to which he aspires.

We see, then, that being excellences of our person which we bring to the affairs of social life, all of the virtues may be sought and their absence may render us liable to shame. But some virtues are joined to shame in a special way, since they are peculiarly indicative of the failure to achieve self-command and its attendant excellences of strength, courage, and self-control. Wrongs manifesting the absence of these qualities are especially likely to subject us to painful feelings of shame. Thus while the principles of right and justice are used to describe the actions disposing us to feel both moral shame and guilt, the perspective is different in each case. In the one we focus on the infringement of the just claims of others and the injury we have done to them, and on their probable resentment or indignation should they discover our deed. Whereas in the other we are struck by the loss to our self-esteem and our inability to carry out our aims: we sense the diminishment of self from our anxiety about the lesser respect that others may have for us and from our disappointment with ourselves for failing to live up to our ideals. Moral shame and guilt, it is clear, both involve our relations to others, and each is an expression of our acceptance of the first principles of right and justice. Nevertheless, these emotions occur within different points of view, our circumstances being seen in contrasting ways.

68. SEVERAL CONTRASTS BETWEEN THE RIGHT AND THE GOOD

In order to bring out the structural features of the contract view, I shall now mention several contrasts between the concepts of the right and the good. Since these concepts enable us to explain moral worth, they are the two fundamental concepts of the theory. The structure of an ethical doctrine depends upon how it relates these two notions and defines their differences. The distinctive features of justice as fairness can be shown by noting these points.

One difference is that whereas the principles of justice (and the principles of right generally) are those that would be chosen in the original position, the principles of rational choice and the criteria of deliberative rationality are not chosen at all. The first task in the theory of justice is to define the initial situation so that the principles that result express the correct conception of justice from a philosophical point of view. This means that the typical features of this situation should represent reasonable constraints on arguments for accepting principles and that the principles agreed to should match our considered convictions of justice in reflective equilibrium. Now, the analogous problem for the theory of the good does not arise. There is, to begin with, no necessity for an agreement upon the principles of rational choice. Since each person is free to plan his life as he pleases (so long as his intentions are consistent with the principles of justice), unanimity concerning the standards of rationality is not required. All the theory of justice assumes is that, in the thin account of the good, the evident criteria of rational choice are sufficient to explain the preference for the primary goods, and that such variations as exist in conceptions of rationality do not affect the principles of justice adopted in the original position.

Nevertheless, I have assumed that human beings do recognize certain principles and that these standards may be taken by enumeration to replace the notion of rationality. We can, if we wish, allow certain variations in the list. Thus there is disagreement as to the best way to deal with uncertainty.²⁸ There is no reason, though, why individuals in making their plans should not be thought of as following their inclinations in this case. Therefore any principle of choice under uncertainty which seems plausible can be added to the list, so long as decisive arguments against it

28. See the discussion in R. D. Luce and Howard Raiffa, *Games and Decisions* (New York, John Wiley and Sons, 1957), pp. 278–306.

are not forthcoming. It is only in the thin theory of the good that we have to worry about these matters. Here the notion of rationality must be interpreted so that the general desire for the primary goods can be established and the choice of the principles of justice demonstrated. But even in this case, I have suggested that the conception of justice adopted is insensitive with respect to conflicting interpretations of rationality. But in any event, once the principles of justice are chosen, and we are working within the full theory, there is no need to set up the account of the good so as to force unanimity on all the standards of the rational choice. In fact, it would contradict the freedom of choice that justice as fairness assures to individuals and groups within the framework of just institutions.

A second contrast between the right and the good is that it is, in general, a good thing that individuals' conceptions of their good should differ in significant ways, whereas this is not so for conceptions of right. In a well-ordered society citizens hold the same principles of right and they try to reach the same judgment in particular cases. These principles are to establish a final ordering among the conflicting claims that persons make upon one another and it is essential that this ordering be identifiable from everyone's point of view, however difficult it may be in practice for everyone to accept it. On the other hand, individuals find their good in different ways, and many things may be good for one person that would not be good for another. Moreover, there is no urgency to reach a publicly accepted judgment as to what is the good of particular individuals. The reasons that make such an agreement necessary in questions of justice do not obtain for judgments of value. Even when we take up another's point of view and attempt to estimate what would be to his advantage, we do so as an adviser, so to speak. We try to put ourselves in the other's place, and imagining that we have his aims and wants, we attempt to see things from his standpoint. Cases of paternalism aside, our judgment is offered when it is asked for, but there is no conflict of right if our advice is disputed and our opinion is not acted upon.

In a well-ordered society, then, the plans of life of individuals are different in the sense that these plans give prominence to different aims, and persons are left free to determine their good, the views of others being counted as merely advisory. Now this variety in conceptions of the good is itself a good thing, that is, it is rational for members of a well-ordered society to want their plans to be different. The reasons for this are obvious. Human beings have various talents and abilities the totality of which is unrealizable by any one person or group of persons. Thus we not only benefit from the complementary nature of our developed inclinations

but we take pleasure in one another's activities. It is as if others were bringing forth a part of ourselves that we have not been able to cultivate. We have had to devote ourselves to other things, to only a small part of what we might have done (§79). But the situation is quite otherwise with justice: here we require not only common principles but sufficiently similar ways of applying them in particular cases so that a final ordering of conflicting claims can be defined. Judgments of justice are advisory only in special circumstances.

The third difference is that many applications of the principles of justice are restricted by the veil of ignorance, whereas evaluations of a person's good may rely upon a full knowledge of the facts. Thus, as we have seen, not only must the principles of justice be chosen in the absence of certain kinds of particular information, but when these principles are used in designing constitutions and basic social arrangements, and in deciding between laws and policies, we are subject to similar although not as strict limitations. The delegates to a constitutional convention, and ideal legislators and voters, are also required to take up a point of view in which they know only the appropriate general facts. An individual's conception of his good, on the other hand, is to be adjusted from the start to his particular situation. A rational plan of life takes into account our special abilities, interests, and circumstances, and therefore it quite properly depends upon our social position and natural assets. There is no objection to fitting rational plans to these contingencies, since the principles of justice have already been chosen and constrain the content of these plans, the ends that they encourage and the means that they use. But in judgments of justice, it is only at the judicial and administrative stage that all restrictions on information are dropped, and particular cases are to be decided in view of all the relevant facts.

In the light of these contrasts we may further clarify an important difference between the contract doctrine and utilitarianism. Since the principle of utility is to maximize the good understood as the satisfaction of rational desire, we are to take as given existing preferences and the possibilities of their continuation into the future, and then to strive for the greatest net balance of satisfaction. But as we have seen, the determination of rational plans is indeterminate in important ways (§64). The more evident and easily applied principles of rational choice do not specify the best plan; a great deal remains to be decided. This indeterminacy is no difficulty for justice as fairness, since the details of plans do not affect in any way what is right or just. Our way of life, whatever our particular circumstances, must always conform to the principles of justice that are

arrived at independently. Thus the arbitrary features of plans of life do not affect these principles, or how the basic structure is to be arranged. The indeterminacy in the notion of rationality does not translate itself into legitimate claims that men can impose on one another. The priority of the right prevents this.

The utilitarian, on the other hand, must concede the theoretical possibility that configurations of preferences allowed by this indeterminacy may lead to injustice as ordinarily understood. For example, assume that the larger part of society has an abhorrence for certain religious or sexual practices, and regards them as an abomination. This feeling is so intense that it is not enough that these practices be kept from the public view; the very thought that these things are going on is enough to arouse the majority to anger and hatred. Even when these attitudes are unsupportable on moral grounds, there appears to be no sure way to exclude them as irrational. Seeking the greatest satisfaction of desire may, then, justify harsh repressive measures against actions that cause no social injury. To defend individual liberty in this case the utilitarian has to show that given the circumstances the real balance of advantages in the long run still lies on the side of freedom; and this argument may or may not be successful.

In justice as fairness, however, this problem never arises. The intense convictions of the majority, if they are indeed mere preferences without any foundation in the principles of justice antecedently established, have no weight to begin with. The satisfaction of these feelings has no value that can be put in the scales against the claims of equal liberty. To have a complaint against the conduct and belief of others we must show that their actions injure us, or that the institutions that authorize what they do treat us unjustly. And this means that we must appeal to the principles that we would acknowledge in the original position. Against these principles neither the intensity of feeling nor its being shared by the majority counts for anything. On the contract view, then, the grounds of liberty are completely separate from existing preferences. Indeed, we may think of the principles of justice as an agreement not to take into account certain feelings when assessing the conduct of others. As I noted before (§50), these points are familiar elements of the classical liberal doctrine. I have mentioned them again in order to show that the indeterminacy in the full theory of the good is no cause for objection. It may leave a person unsettled as to what to do, since it cannot provide him with instructions as to how to decide. But since the aim of justice is not to maximize the fulfillment of rational plans, the content of justice is not in any way affected. Of course, it cannot be denied that prevailing social attitudes tie

the statesman's hands. The convictions and passions of the majority may make liberty impossible to maintain. But bowing to these practical necessities is a different thing from accepting the justification that if these feelings are strong enough and outweigh in intensity any feelings that might replace them, they should carry the decision. By contrast, the contract view requires that we move toward just institutions as speedily as the circumstances permit irrespective of existing sentiments. A definite scheme of ideal institutions is embedded in its principles of justice (§41).

It is evident from these contrasts that in justice as fairness the concepts of the right and the good have markedly distinct features. These differences arise from the structure of contract theory and the priority of right and justice that results. I do not suggest, however, that the terms "right" and "good" (and their relatives) are normally used in ways that reflect these distinctions. Although our ordinary speech may tend to support the account of these concepts, this correspondence is not needed for the correctness of the contract doctrine. Rather, two things suffice. First, there is a way of mapping our considered judgments into the theory of justice such that in reflective equilibrium the counterparts of these convictions turn out to be true, to express judgments that we can accept. And second, once we understand the theory, we can acknowledge these interpretations as suitable renderings of what on reflection we now wish to maintain. Even though we would not normally use these replacements, perhaps because they are too cumbersome, or would be misunderstood, or whatever, we are prepared to grant that they cover in substance all that wants to be said. Certainly these substitutes may not mean the same as the ordinary judgments with which they are paired. How far this is the case is a question that I shall not examine. Moreover, the replacements may indicate a shift more or less drastic from our initial moral judgments as they existed prior to philosophical reflection. Some changes anyway are bound to have taken place as philosophical criticism and construction lead us to revise and extend our views. But what counts is whether the conception of justice as fairness, better than any other theory presently known to us, turns out to lead to true interpretations of our considered judgments, and provides a mode of expression for what we want to affirm.

CHAPTER VIII. THE SENSE OF JUSTICE

Having presented an account of the good, I now turn to the problem of stability. I shall treat it in two stages. In this chapter I discuss the acquisition of the sense of justice by the members of a well-ordered society, and I consider briefly the relative strength of this sentiment when defined by different moral conceptions. The final chapter examines the question of congruence, that is, whether the sense of justice coheres with the conception of our good so that both work together to uphold a just scheme. It is well to keep in mind that much of this chapter is preparation and that various topics are touched upon only to indicate the more basic points that are relevant for the philosophical theory. I begin with a definition of a well-ordered society and with some brief remarks about the meaning of stability. Then I sketch the development of the sense of justice as it presumably would take place once just institutions are firmly established and recognized to be just. The principles of moral psychology also receive some discussion; I emphasize the fact that they are reciprocity principles and connect this with the question of relative stability. The chapter concludes with an examination of the natural attributes in virtue of which human beings are owed the guarantees of equal justice, and which define the natural basis of equality.

69. THE CONCEPT OF A WELL-ORDERED SOCIETY

At the beginning (§1) I characterized a well-ordered society as one designed to advance the good of its members and effectively regulated by a public conception of justice. Thus it is a society in which everyone accepts and knows that the others accept the same principles of justice, and the basic social institutions satisfy and are known to satisfy these principles. Now justice as fairness is framed to accord with this idea of society. The persons in the original position are to assume that the principles

chosen are public, and so they must assess conceptions of justice in view of their probable effects as the generally recognized standards (§23). Conceptions that might work out well enough if understood and followed by a few or even by all, so long as this fact were not widely known, are excluded by the publicity condition. We should also note that since principles are consented to in the light of true general beliefs about men and their place in society, the conception of justice adopted is acceptable on the basis of these facts. There is no necessity to invoke theological or metaphysical doctrines to support its principles, nor to imagine another world that compensates for and corrects the inequalities which the two principles permit in this one. Conceptions of justice must be justified by the conditions of our life as we know it or not at all.¹

Now a well-ordered society is also regulated by its public conception of justice. This fact implies that its members have a strong and normally effective desire to act as the principles of justice require. Since a well-ordered society endures over time, its conception of justice is presumably stable: that is, when institutions are just (as defined by this conception), those taking part in these arrangements acquire the corresponding sense of justice and desire to do their part in maintaining them. One conception of justice is more stable than another if the sense of justice that it tends to generate is stronger and more likely to override disruptive inclinations and if the institutions it allows foster weaker impulses and temptations to act unjustly. The stability of a conception depends upon a balance of motives: the sense of justice that it cultivates and the aims that it encourages must normally win out against propensities toward injustice. To estimate the stability of a conception of justice (and the well-ordered society that it defines), one must examine the relative strength of these opposing tendencies.

It is evident that stability is a desirable feature of moral conceptions. Other things equal, the persons in the original position will adopt the more stable scheme of principles. However attractive a conception of justice might be on other grounds, it is seriously defective if the principles of moral psychology are such that it fails to engender in human beings the requisite desire to act upon it. Thus in arguing further for the principles of justice as fairness, I should like to show that this conception is more stable than other alternatives. This argument from stability is for the most part in addition to the reasons so far adduced (except for consid-

1. It follows that such devices as Plato's Noble Lie in the *Republic*, bk. III, 414–415, are ruled out, as well as the advocacy of religion (when not believed) to buttress a social system that could not otherwise survive, as by the Grand Inquisitor in Dostoevsky's *The Brothers Karamazov*.

erations presented in §29). I wish to consider this notion in more detail both for its own sake and to prepare the way for the discussion of other matters such as the basis of equality and the priority of liberty.

To be sure, the criterion of stability is not decisive. In fact, some ethical theories have flouted it entirely, at least on some interpretations. Thus Bentham is occasionally said to have held both the classical principle of utility and the doctrine of psychological egoism. But if it is a psychological law that individuals pursue only interests in themselves, it is impossible for them to have an effective sense of justice (as defined by the principle of utility). The best that the ideal legislator can do is to design social arrangements so that from self- or group-interested motives citizens are persuaded to act in ways that maximize the sum of well-being. In this conception the identification of interests that results is truly artificial: it rests upon the artifice of reason, and individuals comply with the institutional scheme solely as a means to their separate concerns.²

This sort of divergence between principles of right and justice and human motives is unusual, although instructive as a limiting case. Most traditional doctrines hold that to some degree at least human nature is such that we acquire a desire to act justly when we have lived under and benefited from just institutions. To the extent that this is true, a conception of justice is psychologically suited to human inclinations. Moreover, should it turn out that the desire to act justly is also regulative of a rational plan of life, then acting justly is part of our good. In this event the conceptions of justice and goodness are compatible and the theory as a whole is congruent. The task of this chapter is to explain how justice as fairness generates its own support and to show that it is likely to have greater stability than the traditional alternatives, since it is more in line with the principles of moral psychology. To this end, I shall describe briefly how human beings in a well-ordered society might acquire a sense of justice and the other moral sentiments. Inevitably we shall have to take up some rather speculative psychological questions; but all along I have assumed that general facts about the world, including basic psychological principles, are known to the persons in the original position and relied upon by them in making their decisions. By reflecting on these problems here we survey these facts as they affect the initial agreement.

It may prevent misunderstanding if I make a few remarks about the

2. While Bentham is sometimes interpreted as a psychological egoist, he is not by Jacob Viner, "Bentham and J. S. Mill: The Utilitarian Background" (1949), reprinted in *The Long View and the Short* (Glencoe, Ill., Free Press, 1958); see pp. 312–314. Viner also gives what must be the correct rendering of Bentham's conception of the role of the legislator, pp. 316–319.

concepts of equilibrium and stability. Both of these ideas admit of considerable theoretical and mathematical refinement but I shall use them in an intuitive way.³ The first thing to note perhaps is that they are applied to systems of some kind. Thus it is a system that is in equilibrium, and it is so when it has reached a state that persists indefinitely over time so long as no external forces impinge upon it. In order to define an equilibrium state precisely, the boundaries of the system have to be carefully drawn and its determining characteristics clearly set out. Three things are essential: first, to identify the system and to distinguish between internal and external forces; second, to define the states of the system, a state being a certain configuration of its determining characteristics; and third, to specify the laws connecting the states.

Some systems have no equilibrium states, while others have many. These matters depend upon the nature of the system. Now an equilibrium is stable whenever departures from it, caused say by external disturbances, call into play forces within the system that tend to bring it back to this equilibrium state, unless of course the outside shocks are too great. By contrast, an equilibrium is unstable when a movement away from it arouses forces within the system that lead to even greater changes. Systems are more or less stable depending upon the strength of the internal forces that are available to return them to equilibrium. Since in practice all social systems are subject to disturbances of some kind, they are practically stable, let us say, if the departures from their preferred equilibrium positions caused by normal disturbances elicit forces sufficiently strong to restore these equilibria after a decent length of time, or else to stay sufficiently close to them. These definitions are unhappily vague but they should serve our purposes.

The relevant systems here, of course, are the basic structures of the well-ordered societies corresponding to the different conceptions of justice. We are concerned with this complex of political, economic, and social institutions when it satisfies, and is publicly known by those en-

3. For the notions of equilibrium and stability applied to systems, see, for example, W. R. Ashby, *Design for a Brain*, 2nd ed. revised (London, Chapman and Hall, 1960), chs. 2–4, 19–20. The concept of stability I use is actually that of quasi-stability: if an equilibrium is stable, then all the variables return to their equilibrium values after a disturbance has moved the system away from equilibrium; a quasi-stable equilibrium is one in which only some of the variables return to their equilibrium configuration. For this definition, see Harvey Leibenstein, *Economic Backwardness and Economic Growth* (New York, John Wiley and Sons, 1957), p. 18. A well-ordered society is quasi-stable with respect to the justice of its institutions and the sense of justice needed to maintain this condition. While a shift in social circumstances may render its institutions no longer just, in due course they are reformed as the situation requires, and justice is restored.

gaged in it to satisfy, the appropriate principles of justice. We must try to assess the relative stability of these systems. Now I assume that the boundaries of these schemes are given by the notion of a self-contained national community. This supposition is not relaxed until the derivation of the principles of justice for the law of nations (§58), but the wider problems of international law I shall not further discuss. It is also essential to note that in the present case equilibrium and stability are to be defined with respect to the justice of the basic structure and the moral conduct of individuals. The stability of a conception of justice does not imply that the institutions and practices of the well-ordered society do not alter. In fact, such a society will presumably contain great diversity and adopt different arrangements from time to time. In this context stability means that however institutions are changed, they still remain just or approximately so, as adjustments are made in view of new social circumstances. The inevitable deviations from justice are effectively corrected or held within tolerable bounds by forces within the system. Among these forces I assume that the sense of justice shared by the members of the community has a fundamental role. To some degree, then, moral sentiments are necessary to insure that the basic structure is stable with respect to justice.

I now turn to how these sentiments are formed, and on this question there are, broadly speaking, two main traditions. The first stems historically from the doctrine of empiricism and is found in the utilitarians from Hume to Sidgwick. In its most recent and developed form it is represented by social learning theory. One main contention is that the aim of moral training is to supply missing motives: the desire to do what is right for its own sake, and the desire not to do what is wrong. Right conduct is conduct generally beneficial to others and to society (as defined by the principle of utility) for the doing of which we commonly lack an effective motive, whereas wrong conduct is behavior generally injurious to others and to society for the doing of which we often have a sufficient motive. Society must somehow make good these defects. This is achieved by the approbation and disapprobation of parents and of others in authority, who when necessary use rewards and punishments ranging from bestowal and withdrawal of affection to the administration of pleasures and pains. Eventually by various psychological processes we acquire a desire to do what is right and an aversion to doing what is wrong. A second thesis is that the desire to conform to moral standards is normally aroused early in life before we achieve an adequate understanding of the reasons for these norms. Indeed some persons may never grasp the grounds for them in the

utilitarian principle.⁴ The consequence is that our subsequent moral sentiments are likely to bear the scars of this early training which shapes more or less roughly our original nature.

Freud's theory is similar in important respects to this view. He holds that the processes by which the child comes to have moral attitudes center around the oedipal situation and the deep conflicts to which it gives rise. The moral precepts insisted upon by those in authority (in this case the parents) are accepted by the child as the best way to resolve his anxieties, and the resulting attitudes represented by the superego are likely to be harsh and punitive reflecting the stresses of the oedipal phase.⁵ Thus Freud's account supports the two points that an essential part of moral learning occurs early in life before a reasoned basis for morality can be understood, and that it involves the acquisition of new motives by psychological processes marked by conflict and stress. Indeed, his doctrine is a dramatic illustration of these features. It follows that since parents and others in authority are bound to be in various ways misguided and self-seeking in their use of praise and blame, and rewards and punishments generally, our earlier and unexamined moral attitudes are likely to be in important respects irrational and without justification. Moral advance in later life consists partly in correcting these attitudes in the light of whatever principles we finally acknowledge to be sound.

The other tradition of moral learning derives from rationalist thought and is illustrated by Rousseau and Kant, and sometimes by J. S. Mill, and more recently by the theory of Piaget. Moral learning is not so much a matter of supplying missing motives as one of the free development of our innate intellectual and emotional capacities according to their natural bent. Once the powers of understanding mature and persons come to recognize their place in society and are able to take up the standpoint of others, they appreciate the mutual benefits of establishing fair terms of social cooperation. We have a natural sympathy with other persons and an

4. This sketch of moral learning draws from James Mill, the section of the *Fragment on Mackintosh*, which J. S. Mill included in a footnote to ch. XXIII of his father's *Analysis of the Phenomena of the Human Mind* (1869). The passage is in [J. S.] *Mill's Ethical Writings*, ed. J. B. Schneewind (New York, Collier Books, 1965), pp. 259–270. For an account of social learning theory, see Albert Bandura, *Principles of Behavior Modification* (New York, Holt, Rinehart, and Winston, 1969). For a recent survey of moral learning, see Roger Brown, *Social Psychology* (New York, The Free Press, 1965), ch. VIII; and Martin L. Hoffman, "Moral Development," in *Carmichael's Manual of Psychology*, ed. Paul H. Mussen, 3rd ed. (New York, John Wiley and Sons, 1970), vol. 2, ch. 23; pp. 282–332 is on social learning theory.

5. For accounts of Freud's theory of moral learning, see Roger Brown, *Social Psychology*, pp. 350–381; and Ronald Fletcher, *Instinct in Man* (New York, International Universities Press, 1957), ch. VI, esp. pp. 226–234.

innate susceptibility to the pleasures of fellow feeling and self-mastery, and these provide the affective basis for the moral sentiments once we have a clear grasp of our relations to our associates from an appropriately general perspective. Thus this tradition regards the moral feelings as a natural outgrowth of a full appreciation of our social nature.⁶

Mill expresses the view as follows: the arrangements of a just society are so suited to us that anything which is obviously necessary for it is accepted much like a physical necessity. An indispensable condition of such a society is that all shall have consideration for the others on the basis of mutually acceptable principles of reciprocity. It is painful for us when our feelings are not in union with those of our fellows; and this tendency to sociality provides in due course a firm basis for the moral sentiments. Moreover, Mill adds, to be held accountable to the principles of justice in one's dealings with others does not stunt our nature. Instead it realizes our social sensibilities and by exposing us to a larger good enables us to control our narrower impulses. It is only when we are restrained not because we injure the good of others but by their mere displeasure, or what seems to us their arbitrary authority, that our nature is blunted. If the reasons for moral injunctions are made plain in terms of the just claims of others, these constraints do us no injury but are seen to be compatible with our good.⁷ Moral learning is not so much a matter of acquiring new motives, for these will come about of themselves once the requisite developments in our intellectual and emotional capacities have taken place. It follows that a full grasp of moral conceptions must await maturity; the child's understanding is always primitive and the characteristic features of his morality fall away in later stages. The rationalist tradition presents a happier picture, since it holds that the principles of right and justice spring from our nature and are not at odds with our good, whereas the other account would seem to include no such guarantee.

I shall not try to assess the relative merits of these two conceptions of

6. For Rousseau, see *Emile*, trans. Barbara Foxley (London, J.M. Dent and Sons, 1908), esp. pp. 46–66 (in bk. II), 172–196, 244–258 (in bk. IV); for Kant, *The Critique of Practical Reason*, pt. II, with the misleading name: *The Methodology of Pure Practical Reason*; and J. S. Mill as cited below, note 7. For Jean Piaget, see *The Moral Judgment of the Child*, trans. Marjorie Gabain (London, Kegan Paul, Trench, Trubner, 1932). Further development of this approach is found in Lawrence Kohlberg; see “The Development of Children's Orientation toward a Moral Order: 1. Sequence in the Development of Moral Thought,” *Vita Humana*, vol. 6 (1963); and “Stage and Sequence: The Cognitive Developmental Approach to Socialization,” in *Handbook of Socialization Theory and Research*, ed. D. A. Goslin (Chicago, Rand McNally, 1969), ch. VI. For a critique, see Hoffman, “Moral Development,” pp. 264–275 (on Piaget), pp. 276–281 (on Kohlberg).

7. For Mill's view, see *Utilitarianism*, chs. III and V, pars. 16–25; *On Liberty*, ch. III, par. 10; and *Mill's Ethical Writings*, ed. J. B. Schneewind, pp. 257–259.

moral learning. Surely there is much that is sound in both and it seems preferable to try to combine them in a natural way. It must be emphasized that a moral view is an extremely complex structure of principles, ideals, and precepts, and involves all the elements of thought, conduct, and feeling. Certainly many kinds of learning ranging from reinforcement and classical conditioning to highly abstract reasoning and the refined perception of exemplars enter into its development. Presumably at some time or other each has a necessary role. In the next several sections (§§70–72) I sketch the course of moral development as it might occur in a well-ordered society realizing the principles of justice as fairness. I am concerned solely with this special case. Thus my aim is to indicate the major steps whereby a person would acquire an understanding of and an attachment to the principles of justice as he grows up in this particular form of well-ordered society. These steps I take to be identified by the main structural features of the complete scheme of principles, ideals, and precepts, as these are applied to social arrangements. As I shall explain, we are led to distinguish between the moralities of authority, of association, and of principles. The account of moral development is tied throughout to the conception of justice which is to be learned, and therefore presupposes the plausibility if not the correctness of this theory.⁸

A caveat is apropos here similar to that I made before in regard to the remarks on economic theory (§42). We want the psychological account of moral learning to be true and in accordance with existing knowledge. But of course it is impossible to take the details into account; I sketch at best only the main outlines. One must keep in mind that the purpose of the following discussion is to examine the question of stability and to

8. While the view of moral development to follow in §§70–72 is designed to fit the theory of justice, I have borrowed from several sources. The idea of three stages the content of which is given by precepts, role ideals, and principles is similar to William McDougall, *An Introduction to Social Psychology* (London, Methuen, 1908), chs. VII–VIII. Piaget's *The Moral Judgment of the Child* suggested to me the contrast between the morality of authority and the moralities of association and principles, and much of the description of these stages. See also Kohlberg's further elaboration of this type of theory in the references cited in note 6 above, esp. pp. 369–389, on his six stages. In the last several paragraphs of §75 I note some differences between the view I present and these writers. Concerning Kohlberg's theory, I should add here that I believe the morality of association is parallel to his stages three to five. Development within this stage is being able to assume more complex, demanding, and comprehensive roles. But more important, I assume that the final stage, the morality of principles, may have different contents given by any of the traditional philosophical doctrines we have discussed. It is true that I argue for the theory of justice as superior, and work out the psychological theory on this presumption; but this superiority is a philosophical question and cannot, I believe, be established by the psychological theory of development alone.

contrast the psychological roots of the various conceptions of justice. The crucial point is how the general facts of moral psychology affect the choice of principles in the original position. Unless the psychological account is defective in a way that would call into question the acknowledgment of the principles of justice rather than the standard of utility, say, no irreparable difficulty should ensue. I also hope that none of the further uses of psychological theory will prove too wide of the mark. Particularly important among these is the account of the basis of equality.

70. THE MORALITY OF AUTHORITY

The first stage in the sequence of moral development I shall refer to as the morality of authority. While certain aspects of this morality are preserved at later stages for special occasions, we can regard the morality of authority in its primitive form as that of the child. I assume that the sense of justice is acquired gradually by the younger members of society as they grow up. The succession of generations and the necessity to teach moral attitudes (however simple) to children is one of the conditions of human life.

Now I shall assume that the basic structure of a well-ordered society includes the family in some form, and therefore that children are at first subject to the legitimate authority of their parents. Of course, in a broader inquiry the institution of the family might be questioned, and other arrangements might indeed prove to be preferable. But presumably the account of the morality of authority could, if necessary, be adjusted to fit these different schemes. In any event, it is characteristic of the child's situation that he is not in a position to assess the validity of the precepts and injunctions addressed to him by those in authority, in this case his parents. He lacks both the knowledge and the understanding on the basis of which their guidance can be challenged. Indeed, the child lacks the concept of justification altogether, this being acquired much later. Therefore he cannot with reason doubt the propriety of parental injunctions. But since we are assuming that the society is well-ordered we may suppose, so as to avoid needless complications, that these precepts are on the whole justified. They accord with a reasonable interpretation of familial duties as defined by the principles of justice.

The parents, we may suppose, love the child and in time the child comes to love and to trust his parents. How does this change in the child

come about? To answer this question I assume the following psychological principle: the child comes to love the parents only if they manifestly first love him.⁹ Thus the child's actions are motivated initially by certain instincts and desires, and his aims are regulated (if at all) by rational self-interest (in a suitably restricted sense). Although the child has the potentiality for love, his love of the parents is a new desire brought about by his recognizing their evident love of him and his benefiting from the actions in which their love is expressed.

The parents' love of the child is expressed in their evident intention to care for him, to do for him as his rational self-love would incline, and in the fulfillment of these intentions. Their love is displayed by their taking pleasure in his presence and supporting his sense of competence and self-esteem. They encourage his efforts to master the tasks of growing up and they welcome his assuming his own place. In general, to love another means not only to be concerned for his wants and needs, but to affirm his sense of the worth of his own person. Eventually, then, the love of the parents for the child gives rise to his love in return. The child's love does not have a rational instrumental explanation: he does not love them as a means to achieve his initial self-interested ends. With this aim in view he could conceivably act as if he loved them, but his doing so would not constitute a transformation of his original desires. By the stated psychological principle, a new affection is in time called into being by the evident love of the parents.

There are several ways in which this psychological law may be analyzed into further elements. Thus it is unlikely that the child's recognition of parental affection causes directly a returning sentiment. We may conjecture several other steps as follows: when the parents' love of the child is recognized by him on the basis of their evident intentions, the child is assured of his worth as a person. He is made aware that he is appreciated for his own sake by what are to him the imposing and powerful persons in his world. He experiences parental affection as unconditional; they care for his presence and spontaneous acts, and the pleasure they take in him is not dependent upon disciplined performances that contribute to the well-being of others. In due course, the child comes to trust his parents and to have confidence in his surroundings; and this leads him to launch out and to test his maturing abilities, all the while supported by their affection and

9. The formulation of this psychological law is drawn from Rousseau's *Emile*, p. 174. Rousseau says that while we like from the start what contributes to our preservation, this attachment is quite unconscious and instinctive. "Ce que transforme cet instinct en sentiment, l'attachement en amour, l'aversion en haine, c'est l'intention manifestée de nous nuire ou de nous être utile."

encouragement. Gradually he acquires various skills and develops a sense of competence that affirms his self-esteem. It is in the course of this whole process that the child's affection for his parents develops. He connects them with the success and enjoyment that he has had in sustaining his world, and with his sense of his own worth. And this brings about his love for them.

We must now consider how the child's love and trust will show itself. At this point it is necessary to keep in mind the peculiar features of the authority situation. The child does not have his own standards of criticism, since he is not in a position to reject precepts on rational grounds. If he loves and trusts his parents, he will tend to accept their injunctions. He will also strive to be like them, assuming that they are indeed worthy of esteem and adhere to the precepts which they enjoin. They exemplify, let us suppose, superior knowledge and power, and set forth appealing examples of what is demanded. The child, therefore, accepts their judgment of him and he will be inclined to judge himself as they do when he violates their injunctions. At the same time, of course, his desires exceed the bonds of what is permitted, for otherwise there would be no need for these precepts. Thus parental norms are experienced as constraints and the child may rebel against them. After all, he may see no reason why he should comply with them; they are in themselves arbitrary prohibitions and he has no original tendency to do the things he is told to do. Yet if he does love and trust his parents, then, once he has given in to temptation, he is disposed to share their attitude toward his misdemeanors. He will be inclined to confess his transgression and to seek reconciliation. In these various inclinations are manifested the feelings of (authority) guilt. Without these and related inclinations, feelings of guilt would not exist. But it is also true that the absence of these feelings would indicate a lack of love and trust. For given the nature of the authority situation and the principles of moral psychology connecting the ethical and the natural attitudes, love and trust will give rise to feelings of guilt once the parental injunctions are disobeyed. Admittedly in the case of the child it is sometimes difficult to distinguish feelings of guilt from the fear of punishment, and especially from the dread of the loss of parental love and affection. The child lacks the concepts for understanding moral distinctions and this will reflect itself in his behavior. I have supposed, however, that even in the child's case we can separate (authority) guilt feelings from fear and anxiety.

In the light of this sketch of the development of the morality of authority, it seems that the conditions favoring its being learned by the child are

these.¹⁰ First, the parents must love the child and be worthy objects of his admiration. In this way they arouse in him a sense of his own value and the desire to become the sort of person that they are. Secondly, they must enunciate clear and intelligible (and of course justifiable) rules adapted to the child's level of comprehension. In addition they should set out the reasons for these injunctions so far as these can be understood, and they must also follow these precepts insofar as they apply to them as well. The parents should exemplify the morality which they enjoin, and make explicit its underlying principles as time goes on. Doing this is required not only to arouse the child's inclination to accept these principles at a later time, but also to convey how they are to be interpreted in particular cases. Presumably moral development fails to take place to the extent that these conditions are absent, and especially if parental injunctions are not only harsh and unjustified, but enforced by punitive and even physical sanctions. The child's having a morality of authority consists in his being disposed without the prospect of reward or punishment to follow certain precepts that not only may appear to him largely arbitrary but which in no way appeal to his original inclinations. If he acquires the desire to abide by these prohibitions, it is because he sees them as addressed to him by powerful persons who have his love and trust, and who also act in conformity with them. He then concludes that they express forms of action that characterize the sort of person he should want to be. In the absence of affection, example, and guidance, none of these processes can take place, and certainly not in loveless relationships maintained by coercive threats and reprisals.

The child's morality of authority is primitive because for the most part it consists of a collection of precepts, and he cannot comprehend the larger scheme of right and justice within which the rules addressed to him are justified. But even a developed morality of authority in which the basis of the rules can be understood shows many of the same features, and contains similar virtues and vices. There is typically an authoritative person who is loved and trusted, or at least who is accepted as worthy of his position, and whose precepts it is one's duty to follow implicitly. It is not for us to consider the consequences, this being left for those in authority. The prized virtues are obedience, humility, and fidelity to authoritative persons; the leading vices are disobedience, self-will, and temerity. We are to do what is expected without questioning, for not so to act

10. Here I borrow and adapt from E. E. Maccoby, "Moral Values and Behavior in Childhood," in *Socialization and Society*, ed. J. A. Clausen (Boston, Little, Brown, 1968), and Hoffman, "Moral Development," pp. 282-319.

expresses doubt and distrust, and a certain arrogance and tendency to suspicion. Clearly the morality of authority must be subordinate to the principles of right and justice which alone can determine when these extreme requirements, or analogous constraints, are justified. The child's morality of authority is temporary, a necessity arising from his peculiar situation and limited understanding. Moreover, the theological parallel is a special case which, in view of the principle of equal liberty, does not apply to the basic structure of society (§33). Thus the morality of authority has but a restricted role in fundamental social arrangements and can be justified only when the unusual demands of the practice in question make it essential to give certain individuals the prerogatives of leadership and command. In all cases, the scope of this morality is governed by the principles of justice.

71. THE MORALITY OF ASSOCIATION

The second stage of moral development is that of the morality of association. This stage covers a wide range of cases depending on the association in question and it may even include the national community as a whole. Whereas the child's morality of authority consists largely of a collection of precepts, the content of the morality of association is given by the moral standards appropriate to the individual's role in the various associations to which he belongs. These standards include the common sense rules of morality along with the adjustments required to fit them to a person's particular position; and they are impressed upon him by the approval and disapproval of those in authority, or by the other members of the group. Thus at this stage the family itself is viewed as a small association, normally characterized by a definite hierarchy, in which each member has certain rights and duties. As the child becomes older he is taught the standards of conduct suitable for one in his station. The virtues of a good son or a good daughter are explained, or at least conveyed by parental expectations as shown in their approvals and disapprovals. Similarly there is the association of the school and the neighborhood, and also such short-term forms of cooperation, though not less important for this, as games and play with peers. Corresponding to these arrangements one learns the virtues of a good student and classmate, and the ideals of a good sport and companion. This type of moral view extends to the ideals adopted in later life, and so to one's various adult statuses and occupations, one's family position, and even to one's place as a member of

society. The content of these ideals is given by the various conceptions of a good wife and husband, a good friend and citizen, and so on. Thus the morality of association includes a large number of ideals each defined in ways suitable for the respective status or role. Our moral understanding increases as we move in the course of life through a sequence of positions. The corresponding sequence of ideals requires increasingly greater intellectual judgment and finer moral discriminations. Clearly some of these ideals are also more comprehensive than others and make quite different demands upon the individual. As we shall see, having to follow certain ideals quite naturally leads up to a morality of principles.

Now each particular ideal is presumably explained in the context of the aims and purposes of the association to which the role or position in question belongs. In due course a person works out a conception of the whole system of cooperation that defines the association and the ends which it serves. He knows that others have different things to do depending upon their place in the cooperative scheme. Thus he eventually learns to take up their point of view and to see things from their perspective. It seems plausible, then, that acquiring a morality of association (represented by some structure of ideals) rests upon the development of the intellectual skills required to regard things from a variety of points of view and to think of these together as aspects of one system of cooperation. In fact, when we consider it, the requisite array of abilities is quite complex.¹¹ First of all, we must recognize that these different points of view exist, that the perspectives of others are not the same as ours. But we must not only learn that things look different to them, but that they have different wants and ends, and different plans and motives; and we must learn how to gather these facts from their speech, conduct, and countenance. Next, we need to identify the definitive features of these perspectives, what it is that others largely want and desire, what are their controlling beliefs and opinions. Only in this way can we understand and assess their actions, intentions, and motives. Unless we can identify these leading elements, we cannot put ourselves into another's place and find out what we would do in his position. To work out these things, we must, of course, know what the other person's perspective really is. But finally, having understood another's situation, it still remains for us to regulate our own conduct in the appropriate way by reference to it.

Doing these things to a certain minimum degree at least comes easily

11. For the following remarks, I am indebted to John Flavell, *The Development of Role-Taking and Communication Skills in Children* (New York, John Wiley and Sons, 1968), pp. 208–211. See also G. H. Mead, *Mind, Self and Society* (Chicago, University of Chicago Press, 1934), pp. 135–164.

to adults, but it is difficult for children. No doubt this explains in part why the precepts of the child's primitive morality of authority are usually expressed in terms referring to external behavior, and why motives and intentions are largely neglected by children in their appraisal of actions. The child has not yet mastered the art of perceiving the person of others, that is, the art of discerning their beliefs, intentions, and feelings, so that an awareness of these things cannot inform his interpretation of their behavior. Moreover, his ability to put himself in their place is still untutored and likely to lead him astray. It is no surprise, then, that these elements, so important from the final moral point of view, are left out of account at the earliest stage.¹² But this lack is gradually overcome as we assume a succession of more demanding roles with their more complex schemes of rights and duties. The corresponding ideals require us to view things from a greater multiplicity of perspectives as the conception of the basic structure implies.

I have touched upon these aspects of intellectual development for the sake of completeness. I cannot consider them in any detail, but we should note that they obviously have a central place in the acquisition of moral views. How well the art of perceiving the person is learned is bound to affect one's moral sensibility; and it is equally important to understand the intricacies of social cooperation. But these abilities are not sufficient. Someone whose designs are purely manipulative, and who wishes to exploit others for his own advantage, must likewise, if he lacks overwhelming force, possess these skills. The tricks of persuasion and gamesmanship call upon the same intellectual accomplishments. We must, then, examine how we become attached to our fellow associates and later to social arrangements generally. Consider the case of an association the public rules of which are known by all to be just. Now how does it come about that those taking part in the arrangement are bound by ties of friendship and mutual trust and that they rely on one another to do their part? We may suppose that these feelings and attitudes have been generated by participation in the association. Thus once a person's capacity for fellow feeling has been realized by his acquiring attachments in accordance with the first psychological law, then as his associates with evident intention live up to their duties and obligations, he develops friendly feelings toward them, together with feelings of trust and confidence. And this principle is a second psychological law. As individuals enter the association one by one over a period of time, or group by group (suitably

12. For a discussion of these points, see Roger Brown, *Social Psychology*, pp. 239–244.

limited in size), they acquire these attachments when others of longer standing membership do their part and live up to the ideals of their station. Thus if those engaged in a system of social cooperation regularly act with evident intention to uphold its just (or fair) rules, bonds of friendship and mutual trust tend to develop among them, thereby holding them ever more securely to the scheme.

Once these ties are established, a person tends to experience feelings of (association) guilt when he fails to do his part. These feelings show themselves in various ways, for example, in the inclination to make good the harms caused to others (reparation), if such harms have occurred, as well as in a willingness to admit that what one has done is unfair (wrong) and to apologize for it. Feelings of guilt are also manifest in conceding the propriety of punishment and censure, and in finding it more difficult to be angry and indignant with others when they likewise fail to do their share. The absence of these inclinations would betray an absence of ties of friendship and mutual trust. It would indicate a readiness to associate with others in disregard of the standards and criteria of legitimate expectations that are publicly recognized and used by all to adjudicate their disagreements. A person without these feelings of guilt has no qualms about the burdens that fall on others, nor is he troubled by the breaches of confidence by which they are deceived. But when relations of friendship and trust exist, such inhibitions and reactions tend to be aroused by the failure to fulfill one's duties and obligations. If these emotional constraints are missing, there is at best only a show of fellow feeling and mutual trust. Thus just as in the first stage certain natural attitudes develop toward the parents, so here ties of friendship and confidence grow up among associates. In each case certain natural attitudes underlie the corresponding moral feelings: a lack of these feelings would manifest the absence of these attitudes.

The second psychological law presumably takes hold in ways similar to the first. Since the arrangements of an association are recognized to be just (and in the more complex roles the principles of justice are understood and serve to define the ideal appropriate), thereby insuring that all of its members benefit and know that they benefit from its activities, the conduct of others in doing their part is taken to be to the advantage of each. Here the evident intention to honor one's obligations and duties is seen as a form of good will, and this recognition arouses feelings of friendship and trust in return. In due course the reciprocal effects of everyone's doing his share strengthen one another until a kind of equilib-

rium is reached. But we may also suppose that the newer members of the association recognize moral exemplars, that is, persons who are in various ways admired and who exhibit to a high degree the ideal corresponding to their position. These individuals display skills and abilities, and virtues of character and temperament, that attract our fancy and arouse in us the desire that we should be like them, and able to do the same things. Partly this desire to emulate springs from viewing their attributes as prerequisites for their more privileged positions, but it is also a companion effect to the Aristotelian Principle, since we enjoy the display of more complex and subtle activities and these displays tend to elicit a desire in us to do these things ourselves. Thus when the moral ideals belonging to the various roles of a just association are lived up to with evident intention by attractive and admirable persons, these ideals are likely to be adopted by those who witness their realization. These conceptions are perceived as a form of good will and the activity in which they are exemplified is shown to be a human excellence that others likewise can appreciate. The same two psychological processes are present as before: other persons act with evident intention to affirm our well-being and at the same time they exhibit qualities and ways of doing things that appeal to us and arouse the desire to model ourselves after them.

The morality of association takes many forms depending upon the association and role in question, and these forms represent many levels of complexity. But if we consider the more demanding offices that are defined by the major institutions of society, the principles of justice will be recognized as regulating the basic structure and as belonging to the content of a number of important ideals. Indeed, these principles apply to the role of citizen held by all, since everyone, and not only those in public life, is meant to have political views concerning the common good. Thus we may suppose that there is a morality of association in which the members of society view one another as equals, as friends and associates, joined together in a system of cooperation known to be for the advantage of all and governed by a common conception of justice. The content of this morality is characterized by the cooperative virtues: those of justice and fairness, fidelity and trust, integrity and impartiality. The typical vices are graspingness and unfairness, dishonesty and deceit, prejudice and bias. Among associates, giving into these faults tends to arouse feelings of (association) guilt on the one side and resentment and indignation on the other. These moral attitudes are bound to exist once we become attached to those cooperating with us in a just (or fair) scheme.

72. THE MORALITY OF PRINCIPLES

Someone attaining to the more complex forms of the morality of association, as expressed say by the ideal of equal citizen, has an understanding certainly of the principles of justice. He has also developed an attachment to many particular individuals and communities, and he is disposed to follow the moral standards that apply to him in his various positions and which are upheld by social approval and disapproval. Having become affiliated with others and aspiring to live up to these ethical conceptions, he is concerned to win acceptance for his conduct and aims. It would seem that while the individual understands the principles of justice, his motive for complying with them, for some time at least, springs largely from his ties of friendship and fellow feeling for others, and his concern for the approbation of the wider society. I should now like to consider the process whereby a person becomes attached to these highest-order principles themselves, so that just as during the earlier phase of the morality of association he may want to be a good sport, say, he now wishes to be a just person. The conception of acting justly, and of advancing just institutions, comes to have for him an attraction analogous to that possessed before by subordinate ideals.

In conjecturing how this morality of principles might come about (principles here meaning first principles such as those considered in the original position), we should note that the morality of association quite naturally leads up to a knowledge of the standards of justice. In a well-ordered society anyway not only do those standards define the public conception of justice, but citizens who take an interest in political affairs, and those holding legislative and judicial and other similar offices, are constantly required to apply and to interpret them. They often have to take up the point of view of others, not simply with the aim of working out what they will want and probably do, but for the purpose of striking a reasonable balance between competing claims and for adjusting the various subordinate ideals of the morality of association. To put the principles of justice into practice requires that we adopt the standpoints defined by the four-stage sequence (§31). As the situation dictates, we take up the perspective of a constitutional convention, or of a legislature, or whatever. Eventually one achieves a mastery of these principles and understands the values they secure and the way in which they are to everyone's advantage. Now this leads to an acceptance of these principles by a third psychological law. This law states that once the attitudes of love and trust, and of friendly feelings and mutual confidence, have been generated in accord-

ance with the two preceding psychological laws, then the recognition that we and those for whom we care are the beneficiaries of an established and enduring just institution tends to engender in us the corresponding sense of justice. We develop a desire to apply and to act upon the principles of justice once we realize how social arrangements answering to them have promoted our good and that of those with whom we are affiliated. In due course we come to appreciate the ideal of just human cooperation.

Now a sense of justice shows itself in at least two ways. First, it leads us to accept the just institutions that apply to us and from which we and our associates have benefited. We want to do our part in maintaining these arrangements. We tend to feel guilty when we do not honor our duties and obligations, even though we are not bound to those of whom we take advantage by any ties of particular fellow feeling. It may be that they have not yet had sufficient opportunity to display an evident intention to do their share, and are not therefore the objects of such feelings by the second law. Or, again, the institutional scheme in question may be so large that particular bonds never get widely built up. In any case, the citizen body as a whole is not generally bound together by ties of fellow feeling between individuals, but by the acceptance of public principles of justice. While every citizen is a friend to some citizens, no citizen is a friend to all. But their common allegiance to justice provides a unified perspective from which they can adjudicate their differences. Secondly, a sense of justice gives rise to a willingness to work for (or at least not to oppose) the setting up of just institutions, and for the reform of existing ones when justice requires it. We desire to act on the natural duty to advance just arrangements. And this inclination goes beyond the support of those particular schemes that have affirmed our good. It seeks to extend the conception they embody to further situations for the good of the larger community.

When we go against our sense of justice we explain our feelings of guilt by reference to the principles of justice. These feelings, then, are accounted for quite differently than the emotions of authority and association guilt. The complete moral development has now taken place and for the first time we experience feelings of guilt in the strict sense; and the same is true of the other moral emotions. In the child's case, the notion of a moral ideal, and the relevance of intentions and motives, are not understood, and so the appropriate setting for feelings of (principle) guilt does not exist. And in the morality of association, moral feelings depend essentially on ties of friendship and trust to particular individuals or communities, and moral conduct is based in large part on wanting the ap-

proval of one's associates. This may still be true even in the more demanding phases of this morality. Individuals in their role as citizens with a full understanding of the content of the principles of justice may be moved to act upon them largely because of their bonds to particular persons and an attachment to their own society. Once a morality of principles is accepted, however, moral attitudes are no longer connected solely with the well-being and approval of particular individuals and groups, but are shaped by a conception of right chosen irrespective of these contingencies. Our moral sentiments display an independence from the accidental circumstances of our world, the meaning of this independence being given by the description of the original position and its Kantian interpretation.

But even though moral sentiments are in this sense independent from contingencies, our natural attachments to particular persons and groups still have an appropriate place. For within the morality of principles the infractions which earlier gave rise to (association) guilt and resentment, and to the other moral feelings, now occasion these feelings in the strict sense. A reference to the relevant principle is made in explaining one's emotions. When the natural ties of friendship and mutual trust are present, however, these moral feelings are more intense than if they are absent. Existing attachments heighten the feeling of guilt and indignation, or whatever feeling is called for, even at the stage of the morality of principles. Now granting that this heightening is appropriate, it follows that violations of these natural ties are wrongs. For if we suppose that, say, a rational feeling of guilt (that is, a feeling of guilt arising from applying the correct moral principles in the light of true or reasonable beliefs) implies a fault on our part, and that a greater feeling of guilt implies a greater fault, then indeed breach of trust and the betrayal of friendships, and the like, are especially forbidden. The violation of these ties to particular individuals and groups arouses more intense moral feelings, and this entails that these offenses are worse. To be sure, deceit and infidelity are always wrong, being contrary to natural duties and obligations. But they are not always equally wrong. They are worse whenever bonds of affection and good faith have been formed, and this consideration is relevant in working out the appropriate priority rules.

It may seem strange at first that we should come to have the desire to act from a conception of right and justice. How is it possible that moral principles can engage our affections? In justice as fairness there are several answers to this question. First of all, as we have seen (§25), moral principles are bound to have a certain content. Since they are chosen by

rational persons to adjudicate competing claims, they define agreed ways of advancing human interests. Institutions and actions are appraised from the standpoint of securing these ends; and therefore pointless principles, for example, that one is not to look up at the sky on Tuesdays, are rejected as burdensome and irrational constraints. In the original position rational persons have no reason for acknowledging standards of this kind. But secondly, it is also the case that the sense of justice is continuous with the love of mankind. I noted earlier (§30) that benevolence is at a loss when the many objects of its love oppose one another. The principles of justice are needed to guide it. The difference between the sense of justice and the love of mankind is that the latter is supererogatory, going beyond the moral requirements and not invoking the exemptions which the principles of natural duty and obligation allow. Yet clearly the objects of these two sentiments are closely related, being defined in large part by the same conception of justice. If one of them seems natural and intelligible, so is the other. Moreover, feelings of guilt and indignation are aroused by the injuries and deprivations of others unjustifiably brought about either by ourselves or third parties, and our sense of justice is offended in the same way. The content of the principles of justice accounts for this. Finally, the Kantian interpretation of these principles shows that by acting upon them men express their nature as free and equal rational beings (§40). Since doing this belongs to their good, the sense of justice aims at their well-being even more directly. It supports those arrangements that enable everyone to express his common nature. Indeed, without a common or overlapping sense of justice civic friendship cannot exist. The desire to act justly is not, then, a form of blind obedience to arbitrary principles unrelated to rational aims.

I should not, of course, contend that justice as fairness is the only doctrine that can interpret the sense of justice in a natural way. As Sidgwick notes, a utilitarian never regards himself as acting merely for the sake of an impersonal law, but always for the welfare of some being or beings for whom he has some degree of fellow feeling.¹³ The utilitarian view, and no doubt perfectionism as well, meets the condition that the sentiment of justice can be characterized so that it is psychologically understandable. Best of all, a theory should present a description of an ideally just state of affairs, a conception of a well-ordered society such that the aspiration to realize this state of affairs, and to maintain it in being, answers to our good and is continuous with our natural sentiments.

13. *Methods of Ethics*, 7th ed. (London, Macmillan, 1907), p. 501.

A perfectly just society should be part of an ideal that rational human beings could desire more than anything else once they had full knowledge and experience of what it was.¹⁴ The content of the principles of justice, the way in which they are derived, and the stages of moral development show how in justice as fairness such an interpretation is possible.

It would seem, then, that the doctrine of the purely conscientious act is irrational. This doctrine holds, first, that the highest moral motive is the desire to do what is right and just simply because it is right and just, no other description being appropriate; and second, that while other motives certainly have moral value, for example the desire to do what is right because doing this increases human happiness, or because it tends to promote equality, these desires are less morally worthy than that to do what is right solely in virtue of its being right. Ross holds that the sense of right is a desire for a distinct (and unanalyzable) object, since a specific (and unanalyzable) property characterizes actions that are our duty. The other morally worthy desires, while indeed desires for things necessarily connected with what is right, are not desires for the right as such.¹⁵ But on this interpretation the sense of right lacks any apparent reason; it resembles a preference for tea rather than coffee. Although such a preference might exist, to make it regulative of the basic structure of society is utterly capricious; and no less so because it is masked by a fortunate necessary connection with reasonable grounds for judgments of right.

But for one who understands and accepts the contract doctrine, the sentiment of justice is not a different desire from that to act on principles that rational individuals would consent to in an initial situation which gives everyone equal representation as a moral person. Nor is it different from wanting to act in accordance with principles that express men's nature as free and equal rational beings. The principles of justice answer to these descriptions and this fact allows us to give an acceptable interpretation to the sense of justice. In the light of the theory of justice we understand how the moral sentiments can be regulative in our life and have the role attributed to them by the formal conditions on moral principles. Being governed by these principles means that we want to live with

14. On this point, see G. C. Field, *Moral Theory*, 2nd ed. (London, Methuen, 1932), pp. 135f, 141f.

15. For the notion of the purely conscientious act, see W. D. Ross, *The Right and the Good* (Oxford, The Clarendon Press, 1930), pp. 157–160, and *The Foundations of Ethics* (Oxford, The Clarendon Press, 1939), pp. 205f. That this notion makes the right an arbitrary preference, I borrow from J. N. Findlay, *Values and Intentions* (London, George Allen and Unwin, 1961), pp. 213f.

others on terms that everyone would recognize as fair from a perspective that all would accept as reasonable. The ideal of persons cooperating on this basis exercises a natural attraction upon our affections.

Finally, we may observe that the morality of principles takes two forms, one corresponding to the sense of right and justice, the other to the love of mankind and to self-command. As we have noted, the latter is supererogatory, while the former is not. In its normal form of right and justice the morality of principles includes the virtues of the moralities of authority and association. It defines the last stage at which all the subordinate ideals are finally understood and organized into a coherent system by suitably general principles. The virtues of the other moralities receive their explanation and justification within the larger scheme; and their respective claims are adjusted by the priorities assigned by the more comprehensive conception. The morality of supererogation has two aspects depending upon the direction in which the requirements of the morality of principles are willingly surpassed. On the one hand, the love of mankind shows itself in advancing the common good in ways that go well beyond our natural duties and obligations. This morality is not one for ordinary persons, and its peculiar virtues are those of benevolence, a heightened sensitivity to the feelings and wants of others, and a proper humility and unconcern with self. The morality of self-command, on the other hand, in its simplest form is manifest in fulfilling with complete ease and grace the requirements of right and justice. It becomes truly supererogatory when the individual displays its characteristic virtues of courage, magnanimity, and self-control in actions presupposing great discipline and training. And this he may do either by freely assuming offices and positions which call upon these virtues if their duties are to be well performed; or else by seeking superior ends in a manner consistent with justice but surpassing the demands of duty and obligation. Thus the moralities of supererogation, those of the saint and the hero, do not contradict the norms of right and justice; they are marked by the willing adoption by the self of aims continuous with these principles but extending beyond what they enjoin.¹⁶

16. In this account of the aspects of the morality of supererogation I have drawn upon J. O. Urmson, "Saints and Heroes," in *Essays in Moral Philosophy*, ed. A. I. Melden (Seattle, University of Washington Press, 1958). The notion of self-command is taken from Adam Smith, *The Theory of the Moral Sentiments*, pt. VI, sec. III, which may be found in *Adam Smith's Moral and Political Philosophy*, ed. H. W. Schneider (New York, Hafner, 1948), pp. 251–277.

73. FEATURES OF THE MORAL SENTIMENTS

In the next sections I discuss several aspects of the three stages of morality in more detail. The concept of a moral sentiment, the nature of the three psychological laws, and the process whereby they take hold call for further comment. Turning to the first of these matters, I should explain that I shall use the older term "sentiment" for permanent ordered families of governing dispositions, such as the sense of justice and the love of mankind (§30), and for lasting attachments to particular individuals or associations that have a central place in a person's life. Thus there are both moral and natural sentiments. The term "attitude" I use more broadly. Like sentiments, attitudes are ordered families of dispositions either moral or natural, but in their case the tendencies need not be so regulative or enduring. Finally, I shall use the phrases "moral feeling" and "moral emotion" for the feelings and emotions that we experience on particular occasions. I wish to clarify the connection between moral sentiments, attitudes, and feelings, and the relevant moral principles.

The main features of moral sentiments can perhaps be best elucidated by considering the various questions that arise in trying to characterize them and the various feelings in which they are manifested.¹⁷ It is worthwhile to observe the ways in which they are distinguished both from each other and from those natural attitudes and feelings with which they are likely to be confused. Thus, first of all, there are such questions as the following: (a) What are the linguistic expressions that are used to give voice to having a particular moral feeling, and the significant variations, if any, in these expressions? (b) What are the characteristic behavioral indications of a given feeling, and what are the ways in which a person typically betrays how he feels? (c) What are the characteristic sensations and kinesthetic feelings, if any, that are connected with moral emotions? When a person is angry, for example, he may feel hot; he may tremble and experience a tightening of the stomach. He may be unable to speak

17. These questions are suggested by applying to the concept of the moral feelings the kind of inquiry carried out by Wittgenstein in the *Philosophical Investigations* (Oxford, Basil Blackwell, 1953). See also, for example, G. E. M. Anscombe, "Pretending," *Proceedings of the Aristotelian Society*, supp. vol. 32 (1958), pp. 285–289; Phillipa Foot, "Moral Beliefs," *Proceedings of the Aristotelian Society*, vol. 59 (1958–1959), pp. 86–89; and George Pitcher, "On Approval," *Philosophical Review*, vol. 67 (1958). See also B. A. O. Williams, "Morality and the Emotions," *Inaugural Lecture*, Bedford College, University of London, 1965. It may be a difficulty with the emotive theory of ethics as presented by C. L. Stevenson in *Ethics and Language* (New Haven, Yale University Press, 1944) that it cannot identify and distinguish the moral from the nonmoral feelings. For a discussion of this question, see W. P. Alston, "Moral Attitudes and Moral Judgments," *Nous*, vol. 2 (1968).

without his voice shaking; and perhaps he cannot suppress certain gestures. If there are such characteristic sensations and behavioral manifestations for a moral feeling, these do not constitute the feeling of guilt, shame, indignation, or whatever. Such characteristic sensations and manifestations are neither necessary nor sufficient in particular instances for someone to feel guilty, ashamed, or indignant. This is not to deny that some characteristic sensations and behavioral manifestations of disturbance may be necessary if one is to be overwhelmed by feelings of guilt, shame, or indignation. But to have these feelings it is often sufficient that a person sincerely say that he feels guilty, ashamed or indignant, and that he is prepared to give an appropriate explanation of why he feels as he does (assuming of course that he accepts this explanation as correct).

This last consideration introduces the main question in distinguishing the moral feelings from other emotions and from each other, namely: (d) What is the definitive type of explanation required for having a moral feeling, and how do these explanations differ from one feeling to another? Thus when we ask someone why he feels guilty, what sort of answer do we want? Certainly not any reply is acceptable. A reference merely to expected punishment is not enough; this might account for fear or anxiety, but not for guilt feelings. Similarly, mention of harms or misadventures that have fallen upon oneself as a consequence of one's past actions explains feelings of regret but not those of guilt, and much less those of remorse. To be sure, fear and anxiety often accompany feelings of guilt for obvious reasons, but these emotions must not be confused with the moral feelings. We should not suppose, then, that the experience of guilt is somehow a mixture of fear, anxiety, and regret. Anxiety and fear are not moral feelings at all, and regret is connected with some view of our own good, being occasioned, say, by failures to further our interests in sensible ways. Even such phenomena as neurotic guilt feelings, and other deviations from the standard case, are accepted as feelings of guilt and not simply as irrational fears and anxieties because of the special type of explanation for the departure from the norm. It is always supposed in such cases that a deeper psychological investigation will uncover (or has uncovered) the relevant similarity to other guilt feelings.

In general, it is a necessary feature of moral feelings, and part of what distinguishes them from the natural attitudes, that the person's explanation of his experience invokes a moral concept and its associated principles. His account of his feeling makes reference to an acknowledged right or wrong. When we question this, we are likely to offer various forms of guilt feelings as counterexamples. This is easy to understand since the

earliest forms of guilt feelings are those of authority guilt, and we are unlikely to grow up without having what one may call residue guilt feelings. For example, a person raised in a strict religious sect may have been taught that going to the theater is wrong. While he no longer believes this, he tells us that he still feels guilty when attending the theater. But these are not proper guilt feelings, since he is not about to apologize to anyone, or to resolve not to see another play, and so on. Indeed, he should say rather that he has certain sensations and feelings of uneasiness, and the like, which resemble those which he has when he feels guilty. Assuming, then, the soundness of the contract view, the explanation of some moral feelings relies on principles of right that would be chosen in the original position, while the other moral feelings are related to the concept of goodness. For example, a person feels guilty because he knows that he has taken more than his share (as defined by some just scheme), or has treated others unfairly. Or a person feels ashamed because he has been cowardly and not spoken out. He has failed to live up to a conception of moral worth which he has set himself to achieve (§68). What distinguishes the moral feelings from one another are the principles and faults which their explanations typically invoke. For the most part, the characteristic sensations and behavioral manifestations are the same, being psychological disturbances and having the common features of these.

It is worthwhile to note that the same action may give rise to several moral feelings at once provided that, as is often the case, the appropriate explanation for each one can be given (§67). For example, a person who cheats may feel both guilty and ashamed: guilty because he has violated a trust and unfairly advanced himself, his guilt being in answer to the injuries done to others; ashamed because by resorting to such means he has convicted himself in his own eyes (and in those of others) as weak and untrustworthy, as someone who resorts to unfair and covert means to further his ends. These explanations appeal to different principles and values, thus distinguishing the corresponding feelings; but both explanations frequently apply. We may add here that for a person to have a moral feeling, it is not necessary that everything asserted in his explanation be true; it is sufficient that he accepts the explanation. Someone may be in error, then, in thinking that he has taken more than his share. He may not be guilty. Nevertheless, he feels guilty since his explanation is of the right sort, and although mistaken, the beliefs he expresses are sincere.

Next, there is a group of questions concerning the relation of moral attitudes to action: (e) What are the characteristic intentions, endeavors,

and inclinations of a person experiencing a given feeling? What sorts of things does he want to do, or find himself unable to do? An angry man characteristically tries to strike back, or to block the purposes of the person at whom he is angry. When plagued by feelings of guilt, say, a person wishes to act properly in the future and strives to modify his conduct accordingly. He is inclined to admit what he has done and to ask for reinstatement, and to acknowledge and accept reproofs and penalties; and he finds himself less able to condemn others when they behave wrongly. The particular situation will determine which of these dispositions are realized; and we may also suppose that the family of dispositions which may be elicited varies according to the morality of the individual. It is clear, for example, that the typical expressions of guilt and the appropriate explanations will be quite different as the ideals and roles of the morality of association become more complex and demanding; and these feelings in turn will be distinct from the emotions connected with the morality of principles. In justice as fairness, these variations are accounted for in the first instance by the content of the corresponding moral view. The structure of precepts, ideals, and principles shows what sorts of explanations are required.

Further, we can ask: (f) What emotions and responses does a person having a particular feeling expect on the part of other persons? How does he anticipate that they will react toward him, as this is shown, say, in various characteristic distortions in his interpretation of others' conduct toward him? Thus, one who feels guilty, recognizing his action as a transgression of the legitimate claims of others, expects them to resent his conduct and to penalize him in various ways. He also assumes that third parties will be indignant with him. Someone who feels guilty, then, is apprehensive about the resentment and indignation of others, and the uncertainties which thereby arise. By contrast, someone who feels ashamed anticipates derision and contempt. He has fallen short of a standard of excellence, given in to weakness, and shown himself unworthy of association with others who share his ideals. He is apprehensive lest he be cut off and rejected, made an object of scorn and ridicule. Just as the feelings of guilt and shame have different principles in their explanations, they lead us to anticipate different attitudes in other persons. In general, guilt, resentment, and indignation invoke the concept of right, whereas shame, contempt, and derision appeal to the concept of goodness. And these remarks extend in the obvious way to feelings of duty and obligation (if there are such), and to proper pride and a sense of one's own worth.

Finally, we can ask: (g) What are the characteristic temptations to

actions that give rise to the moral feeling and how is the feeling typically resolved? Here again there are marked differences between the moral emotions. Feelings of guilt and shame have different settings and are overcome in distinct ways, and these variations reflect the defining principles with which they are connected and their peculiar psychological bases. Thus, for example, guilt is relieved by reparation and the forgiveness that permits reconciliation; whereas shame is undone by proofs of defects made good, by a renewed confidence in the excellence of one's person. It is also clear, for example, that resentment and indignation have their characteristic resolutions, since the first is aroused by what we regard as wrongs done to ourselves, the second is concerned with wrongs done to others.

Yet the contrasts between the feelings of guilt and shame are so striking that it is helpful to note how they fit in with the distinctions made between different aspects of morality. As we have seen, a breach of any virtue may give rise to shame; it suffices that one prizes the form of action among one's excellences (§67). Analogously, a wrong can always occasion guilt whenever others are in some way harmed, or their rights violated. Thus guilt and shame reflect the concern with others and with one's person that must be present in all moral conduct. Nevertheless, some virtues, and so those moralities that emphasize them, are more typical of the standpoint of one feeling than the other, and therefore are more closely connected with it. Thus in particular, the moralities of supererogation provide the stage for shame; for they represent the higher forms of moral excellence, the love of humankind and self-command, and in choosing them one risks failure from their very nature. It would be a mistake, however, to emphasize the perspective of one feeling more than the other in the complete moral conception. For the theory of right and justice is founded on the notion of reciprocity which reconciles the points of view of the self and of others as equal moral persons. This reciprocity has the consequence that both perspectives characterize moral thought and feeling, usually in roughly even measure. Neither concern for others nor for self has priority, for all are equal; and the balance between persons is given by the principles of justice. And where this balance moves to one side, as with the moralities of supererogation, it does so from the election of self, which freely takes on the larger part. Thus while we may think of the points of view of the self and of others as characteristic of some moralities historically, or of certain perspectives within a full conception, a complete moral doctrine includes both. All by themselves, a morality of shame or of guilt is but a part of a moral view.

In these remarks I have stressed two main points. First of all, the moral attitudes are not to be identified with characteristic sensations and behavioral manifestations, even if these exist. Moral feelings require certain types of explanations. Thus, second, the moral attitudes involve the acceptance of specific moral virtues; and the principles which define these virtues are used to account for the corresponding feelings. The judgments that elucidate different emotions are distinguished from one another by the standards cited in their explanation. Guilt and shame, remorse and regret, indignation and resentment, either appeal to principles belonging to different parts of morality or invoke them from contrasting points of view. An ethical theory must explain and find a place for these distinctions, although presumably each theory will try to do so in its own way.

74. THE CONNECTION BETWEEN MORAL AND NATURAL ATTITUDES

There is a further aspect of moral attitudes that I have noted in the sketch of the development of the sense of justice, namely, their connection with certain natural attitudes.¹⁸ Thus in examining a moral feeling we should ask: what if any are the natural attitudes to which it is related? Now there are two questions here, one the converse of the other. The first asks about the natural attitudes that are shown to be absent when a person fails to have certain moral feelings. Whereas the second asks which natural attitudes are evidenced to be present when someone experiences a moral emotion. In sketching the three stages of morality I have been concerned only with the first question, since its converse raises other and more difficult problems. I have held that, in the context of the authority situation, the child's natural attitudes of love and trust for those in authority lead to feelings of (authority) guilt when he violates the injunctions addressed to him. The absence of these moral feelings would evidence a lack of these natural ties. Similarly, within the framework of the morality of association, the natural attitudes of friendship and mutual trust give rise to feelings of guilt for not fulfilling the duties and obligations recognized by the group. The absence of these feelings would imply the absence of these attachments. These propositions must not be mistaken for their converses, for while feelings of indignation and guilt, say, can often

18. Throughout this section, and indeed on the subject of the moral emotions generally, I am very much indebted to David Sachs.

be taken as evidence for such affections, there may be other explanations. In general, moral principles are affirmed for various reasons and their acceptance is normally sufficient for the moral feelings. To be sure, on the contract theory principles of right and justice have a certain content, and as we have just seen, there is a sense in which acting in accordance with them can be interpreted as acting from a concern for mankind, or for the good of other persons. Whether this fact shows that one acts in part from certain natural attitudes, especially as these involve attachments to particular individuals, and not simply from the general forms of sympathy and benevolence, is a question that I shall leave aside here. Certainly the preceding account of the development of morality supposes that affection for particular persons plays an essential part in the acquisition of morality. But how far these attitudes are required for later moral motivation can be left open, although it would, I think, be surprising if these attachments were not to some degree necessary.

Now the connection between the natural attitudes and the moral sentiments may be expressed as follows: these sentiments and attitudes are both ordered families of characteristic dispositions, and these families overlap in such a manner that the absence of certain moral feelings evidences the absence of certain natural ties. Or alternatively, the presence of certain natural attachments gives rise to a liability to certain moral emotions once the requisite moral development has taken place. We can see how this is so by an example. If A cares for B, then failing a special explanation A is afraid for B when B is in danger and tries to come to B's assistance. Again, if C plans to treat B unjustly, A is indignant with C and attempts to prevent his plans from succeeding. In both cases, A is disposed to protect B's interests. Further, unless there are special circumstances, A is joyful when together with B, and when B suffers injury or dies, A is stricken with grief. If the injury to B is A's responsibility, A will feel remorse. Love is a sentiment, a hierarchy of dispositions to experience and to manifest these primary emotions as the occasion elicits and to act in the appropriate way.¹⁹ To confirm the connection between the natural attitudes and the moral sentiments one simply notes that the disposition on A's part to feel remorse when he injures B, or guilt when he violates B's legitimate claims, or A's disposition to feel indignation when C seeks to deny B's right, are as closely related psychologically with the natural attitudes of love as the disposition to be

19. On this point, see A. F. Shand, *The Foundations of Character*, 2nd ed. (London, Macmillan, 1920), pp. 55f.

joyful in the other's presence, or to feel sorrow when he suffers. The moral sentiments are in some ways more complex. In their complete form they presuppose an understanding and an acceptance of certain principles and an ability to judge in accordance with them. But assuming these things, the liability to moral feelings seems to be as much a part of the natural sentiments as the tendency to be joyful, or the liability to grief. Love sometimes expresses itself in sorrow, at other times in indignation. Either one without the other would be equally unusual. The content of rational moral principles is such as to render these connections intelligible.

Now one main consequence of this doctrine is that the moral feelings are a normal feature of human life. We could not do away with them without at the same time eliminating certain natural attitudes. Among persons who never acted in accordance with their duty of justice except as reasons of self-interest and expediency dictated there would be no bonds of friendship and mutual trust. For when these attachments exist, other reasons are acknowledged for acting fairly. This much seems reasonably obvious. But it also follows from what has been said that, barring self-deception, egoists are incapable of feeling resentment and indignation. If either of two egoists deceives the other and this is found out, neither of them has a ground for complaint. They do not accept the principles of justice, or any other conception that is reasonable from the standpoint of the original position; nor do they experience any inhibition from guilt feelings for breaches of their duties. As we have seen, resentment and indignation are moral feelings and therefore they presuppose an explanation by reference to an acceptance of the principles of right and justice. But by hypothesis the appropriate explanations cannot be given. To deny that self-interested persons are incapable of resentment and indignation is not of course to say that they cannot be angry and annoyed with one another. A person without a sense of justice may be enraged at someone who fails to act fairly. But anger and annoyance are distinct from indignation and resentment; they are not, as the latter are, moral emotions. Nor should it be denied that egoists may want others to recognize the bonds of friendship and to treat them in a friendly way. But these desires are not to be mistaken for ties of affection that lead one to make sacrifices for one's friends. No doubt there are difficulties in distinguishing between resentment and anger, and between apparent and real friendship. Certainly the overt manifestations and actions may seem the same when viewing a limited span of conduct. Yet in the longer run the difference can usually be made out.

One may say, then, that a person who lacks a sense of justice, and who would never act as justice requires except as self-interest and expediency prompt, not only is without ties of friendship, affection, and mutual trust, but is incapable of experiencing resentment and indignation. He lacks certain natural attitudes and moral feelings of a particularly elementary kind. Put another way, one who lacks a sense of justice lacks certain fundamental attitudes and capacities included under the notion of humanity. Now the moral feelings are admittedly unpleasant, in some extended sense of unpleasant; but there is no way for us to avoid a liability to them without disfiguring ourselves. This liability is the price of love and trust, of friendship and affection, and of a devotion to institutions and traditions from which we have benefited and which serve the general interests of mankind. Further, assuming that persons are possessed of interests and aspirations of their own, and that they are prepared in the pursuit of their own ends and ideals to press their claims on one another—that is, so long as the conditions giving rise to questions of justice obtain among them—it is inevitable that, given temptation and passion, this liability will be realized. And since being moved by ends and ideals of excellence implies a liability to humiliation and shame, and an absence of a liability to humiliation and shame implies a lack of such ends and ideals, one can say of shame and humiliation also that they are a part of the notion of humanity. Now the fact that one who lacks a sense of justice, and thereby a liability to guilt, lacks certain fundamental attitudes and capacities is not to be taken as a reason for acting as justice dictates. But it has this significance: by understanding what it would be like not to have a sense of justice—that it would be to lack part of our humanity too—we are led to accept our having this sentiment.

It follows that the moral sentiments are a normal part of human life. One cannot do away with them without at the same time dismantling the natural attitudes as well. And we have also seen (§§30, 72) that the moral sentiments are continuous with these attitudes in the sense that the love of mankind and the desire to uphold the common good include the principles of right and justice as necessary to define their object. None of this is to deny that our existing moral feelings may be in many respects irrational and injurious to our good. Freud is right in his view that these attitudes are often punitive and blind, incorporating many of the harsher aspects of the authority situation in which they were first acquired. Resentment and indignation, feelings of guilt and remorse, a sense of duty and the censure of others, often take perverse and destructive forms, and blunt without reason human spontaneity and enjoyment. When I say that

moral attitudes are part of our humanity, I mean those attitudes that appeal to sound principles of right and justice in their explanation. The reasonableness of the underlying ethical conception is a necessary condition; and so the appropriateness of moral sentiments to our nature is determined by the principles that would be consented to in the original position.²⁰ These principles regulate moral education and the expression of moral approval and disapproval, just as they govern the design of institutions. Yet even if the sense of justice is the normal outgrowth of natural human attitudes within a well-ordered society, it is still true that our present moral feelings are liable to be unreasonable and capricious. However, one of the virtues of a well-ordered society is that, since arbitrary authority has disappeared, its members suffer much less from the burdens of oppressive conscience.

75. THE PRINCIPLES OF MORAL PSYCHOLOGY

We must soon examine the relative stability of justice as fairness in the light of the sketch of moral development. But before doing this I should like to make a few remarks about the three psychological laws. It will help to have a statement of them before us. Taking for granted that they represent tendencies and are effective other things being equal, they can be rendered as follows.

First law: given that family institutions are just and that the parents love the child and manifestly express their love by caring for his good, then the child, recognizing their evident love of him, comes to love them.

Second law: given that a person's capacity for fellow feeling has been realized by acquiring attachments in accordance with the first law, and given that a social arrangement is just and publicly known by all to be just, then this person develops ties of friendly feeling and trust toward others in the association as they with evident intention comply with their duties and obligations, and live up to the ideals of their station.

Third law: given that a person's capacity for fellow feeling has been realized by his forming attachments in accordance with the first two laws, and given that a society's institutions are just and are publicly known by all to be just, then this person acquires the corresponding

20. Mill observes in *On Liberty*, ch. III, par. 10, that while being held to rigid rules of justice for the sake of others develops the social part of our nature, and therefore is compatible with our well-being, being restrained in ways not for their good but because of their mere displeasure blunts our nature if acquiesced in.

sense of justice as he recognizes that he and those for whom he cares are the beneficiaries of these arrangements.

Perhaps the most striking feature of these laws (or tendencies) is that their formulation refers to an institutional setting as being just, and in the last two, as being publicly known to be such. The principles of moral psychology have a place for a conception of justice; and different formulations of these principles result when different conceptions are used. Thus some view of justice enters into the explanation of the development of the corresponding sentiment; hypotheses about this psychological process incorporate moral notions even if these are understood only as part of the psychological theory. This much seems straightforward, and assuming that ethical ideas can be stated clearly, there is no difficulty in seeing how there can be laws of this kind. The preceding outline of moral development indicates how these matters can be worked out. After all, the sense of justice is a settled disposition to adopt and to want to act from the moral point of view insofar at least as the principles of justice define it. It is hardly surprising that these principles should be involved in the formation of this regulative sentiment. Indeed, it seems likely that our understanding of moral learning cannot far exceed our grasp of the moral conceptions that are to be learned. Analogously, our understanding of how we learn our language is limited by what we know about its grammatical and semantic structure. Just as psycholinguistics depends upon linguistics, so the theory of moral learning depends upon an account of the nature of morality and its various forms. Our common sense ideas about these matters do not suffice for the aims of theory.

No doubt some prefer that social theories avoid the use of moral notions. For instance, they may wish to explain the formation of affective ties by laws referring to the frequency of interaction among those engaged in some common task, or to the regularity with which some persons take the initiative or exercise authoritative guidance. Thus one law may state that among equals cooperating together, where equality is defined by the accepted rules, the more often individuals interact with one another, the more likely it is that friendly feelings develop between them. Another law may assert that the more someone in a position of authority uses his powers and leads those subject to him, the more they come to respect him.²¹ But since these laws (or tendencies) do not mention the

21. For examples of suggested laws (or tendencies) of this type, see G. C. Homans, *The Human Group* (New York, Harcourt, Brace, 1950), pp. 243, 247, 249, 251. In a later book, however, the notion of justice is explicitly brought in. See *Social Behavior: Its Elementary Forms* (New York, Harcourt, Brace and World, 1961), pp. 295f, which applies the theory developed at pp. 232–264.

justice (or fairness) of the arrangement in question, they are bound to be very limited in scope. Those subject to another exercising authority will surely regard him differently depending upon whether the whole arrangement is just and well designed to advance what they take to be their legitimate interests. And the same is true of cooperation among equals. Institutions are patterns of human conduct defined by public systems of rules, and the very holding of the offices and positions which they define normally indicates certain intentions and aims. The justice or injustice of society's arrangements and men's beliefs about these questions profoundly influence the social feelings; to a large extent they determine how we regard another's accepting or rejecting an institution, or his attempt to reform or defend it.

It may be objected that much social theory does well enough without using any moral ideas. The obvious example is economics. However, the situation in economic theory is peculiar in that one can often assume a fixed structure of rules and constraints that define the actions open to individuals and firms, and certain simplifying motivational assumptions are highly plausible. The theory of price (its more elementary parts anyway) is an illustration. One does not consider why buyers and sellers behave in accordance with the rules of law governing economic activity; or how preferences get formed or legal norms established. For the most part, these matters are taken as given, and at a certain level there is no objection to this. On the other hand, the so-called economic theory of democracy, the view that extends the basic ideas and methods of price theory to the political process, must for all its merits be regarded with caution.²² For a theory of a constitutional regime cannot take the rules as given, nor simply assume that they will be followed. Clearly the political process is importantly one of enacting and revising rules and of trying to control the legislative and executive branches of government. Even if everything is done in accordance with constitutional procedures, we need to explain why these are accepted. Nothing analogous to the constraints of a competitive market holds for this case; and there are no legal sanctions in the ordinary sense for many sorts of unconstitutional actions by parliaments and chief executives, and the political forces they represent. The leading political actors are guided therefore in part by what they regard as morally permissible; and since no system of constitutional checks and balances succeeds in setting up an invisible hand that can be

22. For references to this theory of democracy, see §31, note 2, and §54, note 18. Of course, those who have developed the theory are aware of this limitation. See, for example, Anthony Downs, "The Public Interest: Its Meaning in a Democracy," *Social Research*, vol. 29 (1962).

relied upon to guide the process to a just outcome, a public sense of justice is to some degree necessary. It would appear, then, that a correct theory of politics in a just constitutional regime presupposes a theory of justice which explains how moral sentiments influence the conduct of public affairs. I touched upon this question in connection with the role of civil disobedience; it suffices to add here that one test of the contract doctrine is how well it serves this purpose.

A second point about the psychological laws is that they govern changes in the affective ties which belong to our final ends. To clarify this, we may observe that to explain an intentional action is to show how, given our beliefs and the available alternatives, it accords with our plan of life, or with that subpart of it relevant in the circumstances. Often this is done by a series of explanations saying that a first thing is done in order to achieve a second; that the second thing is done in order to achieve a third, and so on, the series being finite and ending at an aim for the sake of which the previous things are done. In accounting for our various actions, we may cite many different chains of reasons, and these normally stop at different points given the complexity of a plan of life and its plurality of ends. Moreover, a chain of reasons may have several branches, since an action may be done to advance more than one end. How activities furthering the many ends are scheduled and balanced against each other is settled by the plan itself and the principles upon which it is based.

Now among our final ends are the attachments we have for persons, the interests we take in the realization of their interests, and the sense of justice. The three laws describe how our system of desires comes to have new final ends as we acquire affective ties. These changes are to be distinguished from our forming derivative desires as a consequence of additional knowledge or further opportunities, or from our determining our existing wants in a more specific way. For example, someone wishing to travel to a certain place is informed that a certain route is the best. Upon accepting this advice, he has a desire to proceed in a particular direction. Derivative desires of this sort have a rational explanation. They are desires to do what in view of the evidence on hand will most effectively realize our present aims, and they shift along with knowledge and belief, and the available opportunities. The three psychological laws do not provide rational explanations of desires in this sense; rather they characterize transformations of our pattern of final ends that arise from our recognizing the manner in which institutions and the actions of others affect our good. Of course, whether an aim is final or derivative is not always easy to ascertain. The distinction is made on the basis of a person's

rational plan of life and the structure of this plan is not generally obvious, even to him. Yet for our purposes here, the distinction is clear enough.

A third observation is that the three laws are not merely principles of association or of reinforcement. While they have a certain resemblance to these learning principles, they assert that the active sentiments of love and friendship, and even the sense of justice, arise from the manifest intention of other persons to act for our good. Because we recognize that they wish us well, we care for their well-being in return. Thus we acquire attachments to persons and institutions according to how we perceive our good to be affected by them. The basic idea is one of reciprocity, a tendency to answer in kind. Now this tendency is a deep psychological fact. Without it our nature would be very different and fruitful social cooperation fragile if not impossible. For surely a rational person is not indifferent to things that significantly affect his good; and supposing that he develops some attitude toward them, he acquires either a new attachment or a new aversion. If we answered love with hate, or came to dislike those who acted fairly toward us, or were averse to activities that furthered our good, a community would soon dissolve. Beings with a different psychology either have never existed or must soon have disappeared in the course of evolution. A capacity for a sense of justice built up by responses in kind would appear to be a condition of human sociability. The most stable conceptions of justice are presumably those for which the corresponding sense of justice is most firmly based on these tendencies (§76).

Finally, several comments about the account of moral development as a whole. The reliance upon the three principles of moral psychology is of course a simplification. A fuller account would distinguish between different kinds of learning and therefore between instrumental conditioning (reinforcement) and classical conditioning, so likely to shape our emotions and feelings. A consideration of modeling and imitation, and the learning of concepts and principles, would also be necessary.²³ There is no reason to deny the significance of these forms of learning. For our purposes, though, the three-stage schema may suffice. Insofar as it stresses the forming of attachments as final ends, the sketch of moral learning resembles the empiricist tradition with its emphasis on the importance of acquiring new motives.

There are also ties with what I have called the rationalistic view. For one thing, the acquisition of the sense of justice takes place in stages connected with the growth of knowledge and understanding. One must

23. See Brown, *Social Psychology*, pp. 411f.

develop a conception of the social world and of what is just and unjust if the sentiment of justice is to be acquired. The manifest intentions of others are recognized against a background of public institutions as interpreted by one's view of the self and its situation. I have not maintained, however, that the stages of development are innate or determined by psychological mechanisms. Whether various native propensities influence these stages is a matter I have left aside. Rather a theory of right and justice is used to describe what the expected course of development might be. The manner in which a well-ordered society is arranged, and the full system of principles, ideals, and precepts that govern the complete scheme, provide a way of distinguishing the three levels of morality. It seems plausible that, in a society regulated by the contract doctrine, moral learning would follow the order presented. The stages are determined by the structure of what is to be learned, proceeding from the simpler to the more complex as the requisite capacities are realized.

Last of all, by founding the account of moral learning explicitly upon a particular ethical theory, it is evident in what sense the sequence of stages represents a progressive development and not simply a regular sequence. Just as persons gradually formulate rational plans of life that answer to their deeper interests, so they come to know the derivation of moral precepts and ideals from the principles that they would accept in an initial situation of equality. Ethical norms are no longer experienced merely as constraints, but are tied together into one coherent conception. The connection between these standards and human aspirations is now comprehended, and persons understand their sense of justice as an extension of their natural attachments, and as a way of caring about the collective good. The many chains of reasons with their various stopping points are no longer simply distinct but are seen as elements of a systematic view. These remarks assume, however, a particular theory of justice. Those who espouse a different one will favor another account of these matters. But in any case, some conception of justice surely has a place in explaining moral learning, even if this conception belongs solely to the psychological theory and is not itself accepted as philosophically correct.

76. THE PROBLEM OF RELATIVE STABILITY

I now turn to the comparison between justice as fairness and other conceptions with respect to stability. It may be useful to recall that the problem of stability arises because a just scheme of cooperation may not

be in equilibrium, much less stable. To be sure, from the standpoint of the original position, the principles of justice are collectively rational; everyone may expect to improve his situation if all comply with these principles, at least in comparison with what his prospects would be in the absence of any agreement. General egoism represents this no-agreement point. Nevertheless, from the perspective of any one man, both first-person and free-rider egoism would be still better. Of course given the conditions of the original position neither of these options is a serious candidate (§23). Yet in everyday life an individual, if he is so inclined, can sometimes win even greater benefits for himself by taking advantage of the cooperative efforts of others. Sufficiently many persons may be doing their share so that when special circumstances allow him not to contribute (perhaps his omission will not be found out), he gets the best of both worlds: on these occasions anyway things proceed much as if free-rider egoism had been acknowledged.

Just arrangements may not be in equilibrium then because acting fairly is not in general each man's best reply to the just conduct of his associates. To insure stability men must have a sense of justice or a concern for those who would be disadvantaged by their defection, preferably both. When these sentiments are sufficiently strong to overrule the temptations to violate the rules, just schemes are stable. Meeting one's duties and obligations is now regarded by each person as the correct answer to the actions of others. His rational plan of life regulated by his sense of justice leads to this conclusion.

As I remarked earlier, Hobbes connected the question of stability with that of political obligation. One may think of the Hobbesian sovereign as a mechanism added to a system of cooperation which would be unstable without it. The general belief in the sovereign's efficacy removes the two kinds of instability (§42). Now it is evident how relations of friendship and mutual trust, and the public knowledge of a common and normally effective sense of justice, bring about the same result. For given these natural attitudes and the desire to do what is just, no one wishes to advance his interests unfairly to the disadvantage of others; this removes instability of the first kind. And since each recognizes that these inclinations and sentiments are prevalent and effective, there is no reason for anyone to think that he must violate the rules to protect his legitimate interests; so instability of the second kind is likewise absent. Of course, some infractions will presumably occur, but when they do feelings of guilt arising from friendship and mutual trust and the sense of justice tend to restore the arrangement.

Moreover, a society regulated by a public sense of justice is inherently stable: other things equal, the forces making for stability increase (up to some limit) as time passes. This inherent stability is a consequence of the reciprocal relation between the three psychological laws. The more effective operation of one law strengthens that of the other two. For example, when the second law leads to stronger attachments, the sense of justice acquired by the third law is reinforced because of the greater concern for the beneficiaries of just institutions. And going the other way, a more effective sense of justice leads to a more secure intention to do one's share, and the recognition of this fact arouses more intense feelings of friendship and trust. Again, it seems that with a firmer assurance of one's own worth and a livelier capacity for fellow feeling brought about by more favorable conditions for the first law, the effects governed by the other two laws should be similarly enhanced. Conversely, persons who have developed a regulative sense of justice and are confident in their self-esteem are more likely to care for their children with manifest intention. Thus all three psychological principles conspire together to support the institutions of a well-ordered society.

There seems to be no doubt then that justice as fairness is a reasonably stable moral conception. But a decision in the original position depends on a comparison: other things equal, the preferred conception of justice is the most stable one. Ideally we should compare the contract view with all its rivals in this respect, but as so often I shall only consider the principle of utility. In order to do this, it is useful to recall three elements that enter into the operation of the psychological laws: namely, an unconditional caring for our good, a clear awareness of the reasons for moral precepts and ideals (aided by explanation and instruction, and the possibility of giving precise and convincing justifications), and the recognition that those complying with these precepts and ideals, and doing their part in social arrangements, both accept these norms and express in their life and character forms of human good which evoke our admiration and esteem (§70). The resulting sense of justice is stronger the more these three elements are realized. The first enlivens the sense of our own worth strengthening the tendency to answer in kind, the second presents the moral conception so that it can be readily understood, and the third displays the adherence to it as attractive. The most stable conception of justice, therefore, is presumably one that is perspicuous to our reason, congruent with our good, and rooted not in abnegation but in affirmation of the self.

Now several things suggest that the sense of justice corresponding to justice as fairness is stronger than the parallel sentiment inculcated by the

other conceptions. First of all, the unconditional concern of other persons and institutions for our good is far stronger on the contract view. The restrictions contained in the principle of justice guarantee everyone an equal liberty and assure us that our claims will not be neglected or overridden for the sake of a larger sum of benefits, even for the whole society. We have only to keep in mind the various priority rules, and the meaning of the difference principle as rendered by its Kantian interpretation (persons are not to be treated as means at all) and its relation to the idea of fraternity (§§29, 17). The effect of these aspects of justice as fairness is to heighten the operation of the reciprocity principle. As we have noted, a more unconditional caring for our good and a clearer refusal by others to take advantage of accident and happenstance, must strengthen our self-esteem; and this greater good must in turn lead to a closer affiliation with persons and institutions by way of an answer in kind. These effects are more intense than in the case of the utility principle, and so the resulting attachments should be stronger.

We can confirm this suggestion by considering the well-ordered society paired with the principle of utility. In this case, the three psychological laws have to be altered. For example, the second law now holds that persons tend to develop friendly feelings toward those who with evident intention do their part in cooperative schemes publicly known to maximize the sum of advantages, or the average well-being (whichever variant is used). In either case the resulting psychological law is not as plausible as before. For suppose that certain institutions are adopted on the public understanding that the greater advantages of some counterbalance the lesser losses of others. Why should the acceptance of the principle of utility (in either form) by the more fortunate inspire the less advantaged to have friendly feelings toward them? This response would seem in fact to be rather surprising, especially if those in a better situation have pressed their claims by maintaining that a greater sum (or average) of well-being would result from their satisfaction. No reciprocity principle is at work in this case and the appeal to utility may simply arouse suspicion. The concern which is expressed for all persons by counting each as one (by weighing everyone's utility equally) is weak compared to that conveyed by the principles of justice. Thus the attachments generated within a well-ordered society regulated by the utility criterion are likely to vary widely between one sector of society and another. Some groups may acquire little if any desire to act justly (now defined by the utilitarian principle) with a corresponding loss in stability.

To be sure, in any kind of well-ordered society the strength of the sense

of justice will not be the same in all social groups. Yet to insure that mutual ties bind the entire society, each and every member of it, one must adopt something like the two principles of justice. It is evident why the utilitarian stresses the capacity for sympathy. Those who do not benefit from the better situation of others must identify with the greater sum (or average) of satisfaction else they will not desire to follow the utility criterion. Now such altruistic inclinations no doubt exist. Yet they are likely to be less strong than those brought about by the three psychological laws formulated as reciprocity principles; and a marked capacity for sympathetic identification seems relatively rare. Therefore these feelings provide less support for the basic structure of society. In addition, as we have seen, following the utilitarian conception tends to be destructive of the self-esteem of those who lose out, particularly when they are already less fortunate (§29). Now it is characteristic of the morality of authority when conceived as a morality for the social order as a whole to demand self-sacrifice for the sake of a higher good and to deprecate the worth of the individual and lesser associations. The emptiness of the self is to be overcome in the service of larger ends. This doctrine is likely to encourage self-hatred with its destructive consequences. Certainly utilitarianism does not go to this extreme, but there is bound to be a similar effect which further weakens the capacity for sympathy and distorts the development of affective ties.

By contrast, in a social system regulated by justice as fairness, identification with the good of others, and an appreciation of what they do as an element in our own good (§79), might be quite strong. But this is possible only because of the mutuality already implicit in the principles of justice. With the constant assurance expressed by these principles, persons will develop a secure sense of their own worth that forms the basis for the love of humankind. By appealing straightway to the capacity for sympathy as a foundation of just conduct in the absence of reciprocity, the principle of utility not only requires more than justice as fairness but depends upon weaker and less common inclinations. Two other elements affect the strength of the sense of justice: the clarity of the moral conception and the attractiveness of its ideals. I shall consider the latter in the next chapter. There I try to show that the contract view is more congruent with our good than its rivals; and assuming this conclusion here, it lends further support to the preceding considerations. The greater clarity of the principles of justice was considered earlier (§49). I noted that in comparison with teleological doctrines, the principles of justice define a per-

spicuous conception. By contrast, the idea of maximizing the aggregate of well-being, or of attaining the greatest perfection, is vague and amorphous. It is easier to ascertain when the equal liberties are infringed and to establish discrepancies from the difference principle than it is to decide whether unequal treatment increases social welfare. The more definite structure of the two principles (and the various priority rules) offers them with greater sharpness to the intellect and thereby secures their hold on the mind. The explanations and reasons given for them are more easily understood and accepted; the conduct expected of us is more clearly defined by publicly acknowledged criteria. On all three counts, then, the contract view seems to possess greater stability.

It is remarkable that Mill appears to agree with this conclusion. He notes that with the advance of civilization persons come more and more to recognize that society between human beings is manifestly impossible on any other basis than that the interests of all are to be consulted. The improvement in political institutions removes the opposition of interests and the barriers and inequalities that encourage individuals and classes to disregard one another's claims. The natural end of this development is a state of the human mind in which each person has a feeling of unity with others. Mill maintains that when this state of mind is perfected, it leads the individual to desire for himself only those things in the benefits of which others are included. One of a person's natural wants is that there should be harmony between his feelings and those of his fellow citizens. He desires to know that his aims and theirs are not in opposition, that he is not setting himself against their good but is furthering what they really wish for.²⁴

Now the desire Mill characterizes here is the desire to act upon the difference principle (or some similar criterion), and not a desire to act on the principle of utility. Mill does not notice the discrepancy; but he seems intuitively to recognize that a perfectly just society in which men's aims are reconciled in ways acceptable to them all would be one that follows the notion of reciprocity expressed by the principles of justice. His remarks accord with the idea that a stable conception of justice which elicits men's natural sentiments of unity and fellow feeling is more likely to incorporate these principles than the utilitarian standard. And this conclusion is borne out by Mill's account of the roots of the sense of justice, for he believes that this sentiment arises not only from sympathy

24. *Utilitarianism*, ch. III, pars. 10–11.

but also from the natural instinct of self-protection and the desire for security.²⁵ This double origin suggests that, in his view, justice strikes a balance between altruism and the claims of self and therefore involves a notion of reciprocity. The contract doctrine achieves the same result, but it does so not by an ad hoc weighing of two competing tendencies, but by a theoretical construction which leads to the appropriate reciprocity principles as a conclusion.

In arguing for the greater stability of the principles of justice I have assumed that certain psychological laws are true, or approximately so. I shall not pursue the question of stability beyond this point. We may note however that one might ask how it is that human beings have acquired a nature described by these psychological principles. The theory of evolution would suggest that it is the outcome of natural selection; the capacity for a sense of justice and the moral feelings is an adaptation of mankind to its place in nature. As ethologists maintain, the behavior patterns of a species, and the psychological mechanisms of their acquisition, are just as much its characteristics as are the distinctive features of its bodily structures; and these patterns of behavior have an evolution exactly as organs and bones do.²⁶ It seems clear that for members of a species which lives in stable social groups, the ability to comply with fair cooperative arrangements and to develop the sentiments necessary to support them is highly advantageous, especially when individuals have a long life and are dependent on one another. These conditions guarantee innumerable occasions when mutual justice consistently adhered to is beneficial to all parties.²⁷

The crucial question here, however, is whether the principles of justice are closer to the tendency of evolution than the principle of utility. Off-hand it would seem that if selection is always of individuals and of their genetic lines, and if the capacity for the various forms of moral behavior

25. *Ibid.*, ch. V, pars. 16–25.

26. See Konrad Lorenz, his introduction to Darwin's *The Expression of the Emotions in Man and Animals* (Chicago, University of Chicago Press, 1965), pp. xii–xiii.

27. Biologists do not always distinguish between altruism and other kinds of moral conduct. Frequently behavior is classified as either altruistic or egoistic. Not so, however, R. B. Trivers in "Evolution of Reciprocal Altruism," *Quarterly Review of Biology*, vol. 46 (1971). He draws a distinction between altruism and reciprocal altruism (or what I should prefer to call simply reciprocity). The latter is the biological analogue of the cooperative virtues of fairness and good faith. Trivers discusses the natural conditions and selective advantages of reciprocity and the capacities that sustain it. See also G. C. Williams, *Adaptation and Natural Selection* (Princeton, Princeton University Press, 1966), pp. 93–96, 113, 195–197, 247. For a discussion of mutualism between species, see Irenäus Eibl-Eibesfeldt, *Ethology*, trans. Erich Klinghammer (New York, Holt, Rinehart and Winston, 1970), pp. 146f, 292–302.

has some genetic basis, then altruism in the strict sense would generally be limited to kin and the smaller face-to-face groups. In these cases the willingness to make considerable self-sacrifice would favor one's descendants and tend to be selected. Turning to the other extreme, a society which had a strong propensity to supererogatory conduct in its relations with other societies would jeopardize the existence of its own distinctive culture and its members would risk domination. Therefore one might conjecture that the capacity to act from the more universal forms of rational benevolence is likely to be eliminated, whereas the capacity to follow the principles of justice and natural duty in relations between groups and individuals other than kin would be favored. We can also see how the system of the moral feelings might evolve as inclinations supporting the natural duties and as stabilizing mechanisms for just schemes.²⁸ If this is correct, then once again the principles of justice are more securely based.

These remarks are not intended as justifying reasons for the contract view. The main grounds for the principles of justice have already been presented. At this point we are simply checking whether the conception already adopted is a feasible one and not so unstable that some other choice might be better. We are in the second part of the argument in which we ask if the acknowledgment previously made should be reconsidered (§25). I do not contend then that justice as fairness is the most stable conception of justice. The understanding required to answer this question is far beyond the primitive theory I have sketched. The conception agreed to need only be stable enough.

77. THE BASIS OF EQUALITY

I now turn to the basis of equality, the features of human beings in virtue of which they are to be treated in accordance with the principles of justice. Our conduct toward animals is not regulated by these principles, or so it is generally believed. On what grounds then do we distinguish between mankind and other living things and regard the constraints of justice as holding only in our relations to human persons? We must examine what determines the range of application of conceptions of justice.

To clarify our question, we may distinguish three levels where the concept of equality applies. The first is to the administration of institutions as public systems of rules. In this case equality is essentially justice

28. On this last point, see Trivers, *ibid.*, pp. 47–54.

as regularity. It implies the impartial application and consistent interpretation of rules according to such precepts as to treat similar cases similarly (as defined by statutes and precedents) and the like (§38). Equality at this level is the least controversial element in the common sense idea of justice.²⁹ The second and much more difficult application of equality is to the substantive structure of institutions. Here the meaning of equality is specified by the principles of justice which require that equal basic rights be assigned to all persons. Presumably this excludes animals; they have some protection certainly but their status is not that of human beings. But this outcome is still unexplained. We have yet to consider what sorts of beings are owed the guarantees of justice. This brings us to the third level at which the question of equality arises.

The natural answer seems to be that it is precisely the moral persons who are entitled to equal justice. Moral persons are distinguished by two features: first they are capable of having (and are assumed to have) a conception of their good (as expressed by a rational plan of life); and second they are capable of having (and are assumed to acquire) a sense of justice, a normally effective desire to apply and to act upon the principles of justice, at least to a certain minimum degree. We use the characterization of the persons in the original position to single out the kind of beings to whom the principles chosen apply. After all, the parties are thought of as adopting these criteria to regulate their common institutions and their conduct toward one another; and the description of their nature enters into the reasoning by which these principles are selected. Thus equal justice is owed to those who have the capacity to take part in and to act in accordance with the public understanding of the initial situation. One should observe that moral personality is here defined as a potentiality that is ordinarily realized in due course. It is this potentiality which brings the claims of justice into play. I shall return to this point below.

We see, then, that the capacity for moral personality is a sufficient condition for being entitled to equal justice.³⁰ Nothing beyond the essential minimum is required. Whether moral personality is also a necessary

29. See Sidgwick, *Methods of Ethics*, p. 496.

30. This fact can be used to interpret the concept of natural rights. For one thing, it explains why it is appropriate to call by this name the rights that justice protects. These claims depend solely on certain natural attributes the presence of which can be ascertained by natural reason pursuing common sense methods of inquiry. The existence of these attributes and the claims based upon them is established independently from social conventions and legal norms. The propriety of the term "natural" is that it suggests the contrast between the rights identified by the theory of justice and the rights defined by law and custom. But more than this, the concept of natural rights includes the idea that these rights are assigned in the first instance to persons, and that they are given a special weight.

condition I shall leave aside. I assume that the capacity for a sense of justice is possessed by the overwhelming majority of mankind, and therefore this question does not raise a serious practical problem. That moral personality suffices to make one a subject of claims is the essential thing. We cannot go far wrong in supposing that the sufficient condition is always satisfied. Even if the capacity were necessary, it would be unwise in practice to withhold justice on this ground. The risk to just institutions would be too great.

It should be stressed that the sufficient condition for equal justice, the capacity for moral personality, is not at all stringent. When someone lacks the requisite potentiality either from birth or accident, this is regarded as a defect or deprivation. There is no race or recognized group of human beings that lacks this attribute. Only scattered individuals are without this capacity, or its realization to the minimum degree, and the failure to realize it is the consequence of unjust and impoverished social circumstances, or fortuitous contingencies. Furthermore, while individuals presumably have varying capacities for a sense of justice, this fact is not a reason for depriving those with a lesser capacity of the full protection of justice. Once a certain minimum is met, a person is entitled to equal liberty on a par with everyone else. A greater capacity for a sense of justice, as shown say in a greater skill and facility in applying the principles of justice and in marshaling arguments in particular cases, is a natural asset like any other ability. The special advantages a person receives for its exercise are to be governed by the difference principle. Thus if some have to a preeminent degree the judicial virtues of impartiality and integrity which are needed in certain positions, they may properly have whatever benefits should be attached to these offices. Yet the application of the principle of equal liberty is not affected by these differences. It is sometimes thought that basic rights and liberties should vary with capacity, but justice as fairness denies this: provided the minimum for moral personality is satisfied, a person is owed all the guarantees of justice.

This account of the basis of equality calls for a few comments. First of

Claims easily overridden for other values are not natural rights. Now the rights protected by the first principle have both of these features in view of the priority rules. Thus justice as fairness has the characteristic marks of a natural rights theory. Not only does it ground fundamental rights on natural attributes and distinguish their bases from social norms, but it assigns rights to persons by principles of equal justice, these principles having a special force against which other values cannot normally prevail. Although specific rights are not absolute, the system of equal liberties is absolute practically speaking under favorable conditions.

all, it may be objected that equality cannot rest on natural attributes. There is no natural feature with respect to which all human beings are equal, that is, which everyone has (or which sufficiently many have) to the same degree. It might appear that if we wish to hold a doctrine of equality, we must interpret it in another way, namely as a purely procedural principle. Thus to say that human beings are equal is to say that none has a claim to preferential treatment in the absence of compelling reasons. The burden of proof favors equality: it defines a procedural presumption that persons are to be treated alike. Departures from equal treatment are in each case to be defended and judged impartially by the same system of principles that hold for all; the essential equality is thought to be equality of consideration.

There are several difficulties with this procedural interpretation.³¹ For one thing, it is nothing more than the precept of treating similar cases similarly applied at the highest level, together with an assignment of the burden of proof. Equality of consideration puts no restrictions upon what grounds may be offered to justify inequalities. There is no guarantee of substantive equal treatment, since slave and caste systems (to mention extreme cases) may satisfy this conception. The real assurance of equality lies in the content of the principles of justice and not in these procedural presumptions. The placing of the burden of proof is not sufficient. But further, even if the procedural interpretation imposed some genuine restrictions on institutions, there is still the question why we are to follow the procedure in some instances and not others. Surely it applies to creatures who belong to some class, but which one? We still need a natural basis for equality so that this class can be identified.

Moreover, it is not the case that founding equality on natural capacities is incompatible with an egalitarian view. All we have to do is to select a range property (as I shall say) and to give equal justice to those meeting its conditions. For example, the property of being in the interior of the unit circle is a range property of points in the plane. All points inside this circle have this property although their coordinates vary within a certain range. And they equally have this property, since no point interior to a circle is more or less interior to it than any other interior point. Now whether there is a suitable range property for singling out the respect in which human beings are to be counted equal is settled by the conception

31. For a discussion of these, see S. I. Benn, "Egalitarianism and the Equal Consideration of Interests," *Nomos IX: Equality*, ed. J. R. Pennock and J. W. Chapman (New York, Atherton Press, 1967), pp. 62–64, 66–68; and W. K. Frankena, "Some Beliefs about Justice" (The Lindley Lecture, The University of Kansas, 1966), pp. 16f.

of justice. But the description of the parties in the original position identifies such a property, and the principles of justice assure us that any variations in ability within the range are to be regarded as any other natural asset. There is no obstacle to thinking that a natural capacity constitutes the basis of equality.

How then can it seem plausible that founding equality on natural attributes undermines equal justice? The notion of a range property is too obvious to be overlooked. There must be a deeper explanation. The answer, I think, is that a teleological theory is often taken for granted. Thus, if the right is to maximize the net balance of satisfaction, say, then rights and duties are to be assigned so as to achieve this end. Among the relevant aspects of the problem are men's different productive skills and capacities for satisfaction. It may happen that maximizing aggregate welfare requires adjusting basic rights to variations in these features. Of course, given the standard utilitarian assumptions, there is a tendency to equality. The relevant thing, however, is that in either case the correct natural basis and the appropriate assignment of rights depends upon the principle of utility. It is the content of the ethical doctrine, and the fact that it is a maximizing notion, that allows variations in capacity to justify unequal fundamental rights, and not the idea that equality is founded on natural attributes. An examination of perfectionism would, I believe, lead to the same conclusion. But justice as fairness is not a maximizing theory. We are not directed to look for differences in natural features that affect some maximand and therefore serve as possible grounds for different grades of citizenship. Although agreeing with many teleological theories in the relevance of natural attributes, the contract view needs much weaker assumptions about their distribution to establish equal rights. It is enough that a certain minimum is generally fulfilled.

Several further points should be noted briefly. First, the conception of moral personality and the required minimum may often prove troublesome. While many concepts are vague to some degree, that of moral personality is likely to be especially so. But these matters are, I think, best discussed in the context of definite moral problems. The nature of the specific issue and the structure of the available general facts may suggest a fruitful way to settle them. In any case, one must not confuse the vagueness of a conception of justice with the thesis that basic rights should vary with natural capacity.

I have said that the minimal requirements defining moral personality refer to a capacity and not to the realization of it. A being that has this capacity, whether or not it is yet developed, is to receive the full protec-

tion of the principles of justice. Since infants and children are thought to have basic rights (normally exercised on their behalf by parents and guardians), this interpretation of the requisite conditions seems necessary to match our considered judgments. Moreover, regarding the potentiality as sufficient accords with the hypothetical nature of the original position, and with the idea that as far as possible the choice of principles should not be influenced by arbitrary contingencies. Therefore it is reasonable to say that those who could take part in the initial agreement, were it not for fortuitous circumstances, are assured equal justice.

Now of course none of this is literally argument. I have not set out the premises from which this conclusion follows, as I have tried to do, albeit not very rigorously, with the choice of conceptions of justice in the original position. Nor have I tried to prove that the characterization of the parties must be used as the basis of equality. Rather this interpretation seems to be the natural completion of justice as fairness. A full discussion would take up the various special cases of lack of capacity. That of children I have already commented upon briefly in connection with paternalism (§39). The problem of those who have lost their realized capacity temporarily through misfortune, accident, or mental stress can be regarded in a similar way. But those more or less permanently deprived of moral personality may present a difficulty. I cannot examine this problem here, but I assume that the account of equality would not be materially affected.

I should like to conclude this section with a few general comments. First of all, the simplicity of the contract view of the basis of equality is worth emphasizing. The minimum capacity for the sense of justice insures that everyone has equal rights. The claims of all are to be adjudicated by the principles of justice. Equality is supported by the general facts of nature and not merely by a procedural rule without substantive force. Nor does equality presuppose an assessment of the intrinsic worth of persons, or a comparative evaluation of their conceptions of the good. Those who can give justice are owed justice.

The advantages of these straightforward propositions become more evident when other accounts of equality are examined. For example, one might think that equal justice means that society is to make the same proportionate contribution to each person's realizing the best life which he is capable of.³² Offhand this may seem an attractive suggestion. It

32. For this idea, see W. K. Frankena, "Some Beliefs about Justice," pp. 14ff; and J. N. Findlay, *Values and Intentions*, pp. 301f.

suffers however from serious difficulties. For one thing it not only requires a method of estimating the relative goodness of plans of life, but it also presupposes some way of measuring what counts as an equal proportionate contribution to persons with different conceptions of their good. The problems in applying this standard are obvious. A more important difficulty is that the greater abilities of some may give them a stronger claim on social resources irrespective of compensating advantages to others. One must assume that variations in natural assets will affect what is necessary to provide equal proportionate assistance to those with different plans of life. But in addition to violating the principle of mutual advantage, this conception of equality means that the strength of men's claims is directly influenced by the distribution of natural abilities, and therefore by contingencies that are arbitrary from a moral point of view. The basis of equality in justice as fairness avoids these objections. The only contingency which is decisive is that of having or not having the capacity for a sense of justice. By giving justice to those who can give justice in return, the principle of reciprocity is fulfilled at the highest level.

A further observation is that we can now more fully reconcile two conceptions of equality. Some writers have distinguished between equality as it is invoked in connection with the distribution of certain goods, some of which will almost certainly give higher status or prestige to those who are more favored, and equality as it applies to the respect which is owed to persons irrespective of their social position.³³ Equality of the first kind is defined by the second principle of justice which regulates the structure of organizations and distributive shares so that social cooperation is both efficient and fair. But equality of the second kind is fundamental. It is defined by the first principle of justice and by such natural duties as that of mutual respect; it is owed to human beings as moral persons. The natural basis of equality explains its deeper significance. The priority of the first principle over the second enables us to avoid balancing these conceptions of equality in an ad hoc manner, while the argument from the standpoint of the original position shows how this precedence comes about (§82).

The consistent application of the principle of fair opportunity requires us to view persons independently from the influences of their social

33. See B. A. O. Williams, "The Idea of Equality," *Philosophy, Politics, and Society*, second series, ed. Peter Laslett and W. G. Runciman (Oxford, Basil Blackwell, 1962), pp. 129–131; and W. G. Runciman, *Relative Deprivation and Social Justice* (London, Routledge and Kegan Paul, 1966), pp. 274–284.

position.³⁴ But how far should this tendency be carried? It seems that even when fair opportunity (as it has been defined) is satisfied, the family will lead to unequal chances between individuals (§46). Is the family to be abolished then? Taken by itself and given a certain primacy, the idea of equal opportunity inclines in this direction. But within the context of the theory of justice as a whole, there is much less urgency to take this course. The acknowledgment of the difference principle redefines the grounds for social inequalities as conceived in the system of liberal equality; and when the principles of fraternity and redress are allowed their appropriate weight, the natural distribution of assets and the contingencies of social circumstances can more easily be accepted. We are more ready to dwell upon our good fortune now that these differences are made to work to our advantage, rather than to be downcast by how much better off we might have been had we had an equal chance along with others if only all social barriers had been removed. The conception of justice, should it be truly effective and publicly recognized as such, seems more likely than its rivals to transform our perspective on the social world and to reconcile us to the dispositions of the natural order and the conditions of human life.

Last of all, we should recall here the limits of a theory of justice. Not only are many aspects of morality left aside, but no account is given of right conduct in regard to animals and the rest of nature. A conception of justice is but one part of a moral view. While I have not maintained that the capacity for a sense of justice is necessary in order to be owed the duties of justice, it does seem that we are not required to give strict justice anyway to creatures lacking this capacity. But it does not follow that there are no requirements at all in regard to them, nor in our relations with the natural order. Certainly it is wrong to be cruel to animals and the destruction of a whole species can be a great evil. The capacity for feelings of pleasure and pain and for the forms of life of which animals are capable clearly imposes duties of compassion and humanity in their case. I shall not attempt to explain these considered beliefs. They are outside the scope of the theory of justice, and it does not seem possible to extend the contract doctrine so as to include them in a natural way. A correct conception of our relations to animals and to nature would seem to depend upon a theory of the natural order and our place in it. One of the tasks of metaphysics is to work out a view of the world which is suited for this purpose; it should identify and systematize the truths deci-

34. See Williams, *ibid.*, pp. 125–129.

sive for these questions. How far justice as fairness will have to be revised to fit into this larger theory it is impossible to say. But it seems reasonable to hope that if it is sound as an account of justice among persons, it cannot be too far wrong when these broader relationships are taken into consideration.

CHAPTER IX. THE GOOD OF JUSTICE

In this chapter I take up the second and last part of the problem of stability. This concerns the question whether justice as fairness and goodness as rationality are congruent. It remains to be shown that given the circumstances of a well-ordered society, a person's rational plan of life supports and affirms his sense of justice. I approach this problem by discussing in turn the various desiderata of a well-ordered society and the ways in which its just arrangements contribute to the good of its members. Thus I note first that such a society allows for persons' autonomy and the objectivity of their judgments of right and justice. I indicate next how justice combines with the ideal of social union, mitigates the propensity to envy and spite, and defines an equilibrium in which the priority of liberty obtains. Finally, by an examination of the contrast between justice as fairness and hedonistic utilitarianism, I attempt to show how just institutions provide for the unity of the self and enable human beings to express their nature as free and equal moral persons. Taking these features together, I then argue that in a well-ordered society an effective sense of justice belongs to a person's good, and so tendencies to instability are kept in check if not eliminated.

78. AUTONOMY AND OBJECTIVITY

Before taking up the various features of a well-ordered society, I should emphasize that I am concerned with the problem of congruence only for this social form. We are therefore still limiting ourselves to strict compliance theory. Yet this case is the first one to examine, for if congruence fails for a well-ordered society it seems bound to fail everywhere. On the other hand, it is by no means a foregone conclusion even in this instance that the right and the good are congruent. For this relation implies that the members of a well-ordered society, when they appraise their plan of life

by the principles of rational choice, will decide to maintain their sense of justice as regulative of their conduct toward one another. The requisite match exists between the principles of justice that would be agreed to in the absence of information and the principles of rational choice that are not chosen at all and applied with full knowledge. Principles accounted for in strikingly different ways nevertheless fit together when those of justice are perfectly realized. Of course, this congruence has its explanation in how the contract doctrine is set up. But the relation is not a matter of course and its basis needs to be worked out.

I shall proceed by examining a number of features of a well-ordered society which all told lead rational persons to confirm their sense of justice. The argument is cumulative and depends upon a convergence of observations the force of which is not summed up until later (§86).

I begin by noting that we sometimes doubt the soundness of our moral attitudes when we reflect on their psychological origins. Thinking that these sentiments have arisen in situations marked say by submission to authority, we may wonder whether they should not be rejected altogether. Since the argument for the good of justice depends upon the members of a well-ordered society having an effective desire to act justly, we must allay these uncertainties. Imagine then that someone experiences the promptings of his moral sense as inexplicable inhibitions which for the moment he is unable to justify. Why should he not regard them as simply neurotic compulsions? If it should turn out that these scruples are indeed largely shaped and accounted for by the contingencies of early childhood, perhaps by the course of our family history and class situation, and that there is nothing to add on their behalf, then there is surely no reason why they should govern our lives. But of course to someone in a well-ordered society there are many things to say. One can point out to him the essential features of the development of the sentiment of justice and how eventually the morality of principles is to be understood. Moreover his moral education itself has been regulated by the principles of right and justice to which he would consent in an initial situation in which all have equal representation as moral persons. As we have seen, the moral conception adopted is independent of natural contingencies and accidental social circumstances; and therefore the psychological processes by which his moral sense has been acquired conform to principles that he himself would choose under conditions that he would concede are fair and undistorted by fortune and happenstance.

Nor can someone in a well-ordered society object to the practices of moral instruction that inculcate a sense of justice. For in agreeing to

principles of right the parties in the original position at the same time consent to the arrangements necessary to make these principles effective in their conduct. Indeed, the adaptability of these arrangements to the limitations of human nature is an important consideration in choosing a conception of justice. Thus no one's moral convictions are the result of coercive indoctrination. Instruction is throughout as reasoned as the development of understanding permits, just as the natural duty of mutual respect requires. None of the ideals, principles, and precepts upheld in the society takes unfair advantage of human weakness. A person's sense of justice is not a compulsive psychological mechanism cleverly installed by those in authority in order to insure his unswerving compliance with rules designed to advance their interests. Nor is the process of education simply a causal sequence intended to bring about as an end result the appropriate moral sentiments. As far as possible each stage foreshadows in its teaching and explanations the conception of right and justice at which it aims and by reference to which we will later recognize that the moral standards presented to us are justified.

These observations are evident consequences of the contract doctrine and the fact that its principles regulate the practices of moral instruction in a well-ordered society. Following the Kantian interpretation of justice as fairness, we can say that by acting from these principles persons are acting autonomously: they are acting from principles that they would acknowledge under conditions that best express their nature as free and equal rational beings. To be sure, these conditions also reflect the situation of individuals in the world and their being subject to the circumstances of justice. But this simply means that the conception of autonomy is that fitting for human beings; the notion suited to superior or inferior natures is most likely different (§40). Thus moral education is education for autonomy. In due course everyone will know why he would adopt the principles of justice and how they are derived from the conditions that characterize his being an equal in a society of moral persons. It follows that in accepting these principles on this basis we are not influenced primarily by tradition and authority, or the opinions of others. However necessary these agencies may be in order for us to reach complete understanding, we eventually come to hold a conception of right on reasonable grounds that we can set out independently for ourselves.

Now on the contract view the notions of autonomy and objectivity are compatible: there is no antinomy between freedom and reason.¹ Both au-

1. The question of the compatibility of autonomy and objectivity is discussed by H. D. Aiken in his

tonomy and objectivity are characterized in a consistent way by reference to the original position. The idea of the initial situation is central to the whole theory and other basic notions are defined in terms of it. Thus acting autonomously is acting from principles that we would consent to as free and equal rational beings, and that we are to understand in this way. Also, these principles are objective. They are the principles that we would want everyone (including ourselves) to follow were we to take up together the appropriate general point of view. The original position defines this perspective, and its conditions also embody those of objectivity: its stipulations express the restrictions on arguments that force us to consider the choice of principles unencumbered by the singularities of the circumstances in which we find ourselves. The veil of ignorance prevents us from shaping our moral view to accord with our own particular attachments and interests. We do not look at the social order from our situation but take up a point of view that everyone can adopt on an equal footing. In this sense we look at our society and our place in it objectively: we share a common standpoint along with others and do not make our judgments from a personal slant. Thus our moral principles and convictions are objective to the extent that they have been arrived at and tested by assuming this general standpoint and by assessing the arguments for them by the restrictions expressed by the conception of the original position. The judicial virtues such as impartiality and considerateness are the excellences of intellect and sensibility that enable us to do these things well.

One consequence of trying to be objective, of attempting to frame our moral conceptions and judgments from a shared point of view, is that we are more likely to reach agreement. Indeed, other things equal, the preferred description of the initial situation is that which introduces the greatest convergence of opinion. It is partly for this reason that we accept the constraints of a common standpoint, since we cannot reasonably expect our views to fall into line when they are affected by the contingencies of our different circumstances. But of course our judgments will not coincide on all questions, and in fact many if not most social issues may still be insoluble, especially if viewed in their full complexity. This is why the numerous simplifications of justice as fairness are acknowledged. We have only to recall the reasons for such notions as the veil of ignorance, pure procedural justice (as opposed to allocative justice), lexi-

essay "The Concept of Moral Objectivity," in *Reason and Conduct* (New York, Alfred Knopf, 1962), pp. 134–170. See also Huntington Terrell, "Moral Objectivity and Freedom," *Ethics*, vol. 76 (1965), pp. 117–127, for a discussion to which I am indebted.

cal ordering, the division of the basic structure into two parts, and so on. Taken all together the parties hope that these and other devices will simplify political and social questions so that the resulting balance of justice, made possible by the greater consensus, outweighs what may have been lost by ignoring certain potentially relevant aspects of moral situations. The complexity of problems of justice is up to the persons in the original position to decide. Although ethical differences are bound to remain, seeing the social world from the original position does permit essential understandings to be reached. The acceptance of the principles of right and justice forges the bonds of civic friendship and establishes the basis of comity amidst the disparities that persist. Citizens are able to recognize one another's good faith and desire for justice even though agreement may occasionally break down on constitutional questions and most certainly on many issues of policy. But unless there existed a common perspective, the assumption of which narrowed differences of opinion, reasoning and argument would be pointless and we would have no rational grounds for believing in the soundness of our convictions.

It is clear that this interpretation of autonomy and objectivity depends upon the theory of justice. The idea of the original position is used to give a consistent rendering of both notions. Of course, if it is believed that the principles of justice would not be chosen, the content of these conceptions would have to be suitably altered. One who holds that the principle of utility would be consented to thinks that our autonomy is expressed by following this criterion. Nevertheless, the general idea will be the same, and both autonomy and objectivity are still explicated by reference to the initial situation. But some have characterized autonomy and objectivity in an entirely different way. They have suggested that autonomy is the complete freedom to form our moral opinions and that the conscientious judgment of every moral agent ought absolutely to be respected. Objectivity is then attributed to those judgments which satisfy all the standards that the agent himself has in his liberty decided are relevant.² These standards may or may not have anything to do with taking up a common point of view that others might reasonably be expected to share; nor of course is the corresponding idea of autonomy connected with such a perspective. I mention these other interpretations only to indicate by contrast the nature of the contract doctrine.

From the standpoint of justice as fairness it is not true that the conscientious judgments of each person ought absolutely to be respected; nor is

2. See Aiken, *ibid.*, pp. 162–169.

it true that individuals are completely free to form their moral convictions. These contentions are mistaken if they mean that, having arrived at our moral opinions conscientiously (as we believe), we always have a claim to be allowed to act on them. In discussing conscientious objection, we noted that the problem here is that of deciding how one is to answer those who strive to act as their erring conscience directs them (§56). How do we ascertain that their conscience and not ours is mistaken, and under what circumstances can they be compelled to desist? Now the answer to these questions is found by ascending to the original position: a person's conscience is misguided when he seeks to impose on us conditions that violate the principles to which we would each consent in that situation. And we can resist his plans in those ways that would be authorized when the conflict is viewed from that perspective. We are not literally to respect the conscience of an individual. Rather we are to respect him as a person and we do this by limiting his actions, when this proves necessary, only as the principles we would both acknowledge permit. In the original position the parties agree to be held responsible for the conception of justice that is chosen. There is no violation of our autonomy so long as its principles are properly followed. Moreover, these principles stipulate that on many occasions we cannot shift the responsibility for what we do onto others. Those in authority are accountable for the policies they pursue and the instructions they lay down. And those who acquiesce in carrying out unjust commands or in abetting evil designs cannot in general plead that they did not know better or that the fault rests solely with those in higher positions. The details concerning these matters belong to partial compliance theory. The essential point here is that the principles that best conform to our nature as free and equal rational beings themselves establish our accountability. Otherwise autonomy is likely to lead to a mere collision of self-righteous wills, and objectivity to the adherence to a consistent yet idiosyncratic system.

Here we should note that in times of social doubt and loss of faith in long established values, there is a tendency to fall back on the virtues of integrity: truthfulness and sincerity, lucidity and commitment, or, as some say, authenticity. If no one knows what is true, at least we can make our beliefs our own in our own way and not adopt them as handed to us by others. If the traditional moral rules are no longer relevant and we cannot agree which ones should take their place, we can in any event decide with a clear head how we mean to act and stop pretending that somehow or other it is already decided for us and we must accept this or that authority. Now of course the virtues of integrity are virtues, and among the excel-

lences of free persons. Yet while necessary, they are not sufficient; for their definition allows for most any content: a tyrant might display these attributes to a high degree, and by doing so exhibit a certain charm, not deceiving himself by political pretenses and excuses of fortune. It is impossible to construct a moral view from these virtues alone; being virtues of form they are in a sense secondary. But joined to the appropriate conception of justice, one that allows for autonomy and objectivity correctly understood, they come into their own. The idea of the original position, and the principles chosen there, show how this is achieved.

In conclusion then a well-ordered society affirms the autonomy of persons and encourages the objectivity of their considered judgments of justice. Any doubts that its members may entertain about the soundness of their moral sentiments when they reflect upon how these dispositions were acquired may be dispelled by seeing that their convictions match the principles which would be chosen in the original position or, if they do not, by revising their judgments so that they do.

79. THE IDEA OF SOCIAL UNION

We have already seen that despite the individualistic features of justice as fairness, the two principles of justice provide an Archimedean point for appraising existing institutions as well as the desires and aspirations which they generate. These criteria provide an independent standard for guiding the course of social change without invoking a perfectionist or an organic conception of society (§41). But the question remains whether the contract doctrine is a satisfactory framework for understanding the values of community and for choosing among social arrangements to realize them. It is natural to conjecture that the congruence of the right and the good depends in large part upon whether a well-ordered society achieves the good of community. I shall take up several aspects of this question in this and the three following sections.

We may begin by recalling that one of the conditions of the original position is that the parties know that they are subject to the circumstances of justice. They assume that each has a conception of his good in the light of which he presses claims against the rest. So although they view society as a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by an identity of interests. Now there are two ways of viewing these suppositions. The first is that taken by the theory of justice: the idea is to derive satisfactory principles from the weakest possible

assumptions. The premises of the theory should be simple and reasonable conditions that everyone or most everyone would grant, and for which convincing philosophical arguments can be given. At the same time, the greater the initial collision of claims into which the principles can introduce an acceptable order, the more comprehensive the theory is likely to be. Therefore a deep opposition of interests is presumed to obtain.

The other way to think of these suppositions is to regard them as describing a certain kind of social order, or a certain aspect of the basic structure that is actually realized. Thus we are led to the notion of private society.³ Its chief features are first that the persons comprising it, whether they are human individuals or associations, have their own private ends which are either competing or independent, but not in any case complementary. And second, institutions are not thought to have any value in themselves, the activity of engaging in them not being counted as a good but if anything as a burden. Thus each person assesses social arrangements solely as a means to his private aims. No one takes account of the good of others, or of what they possess; rather everyone prefers the most efficient scheme that gives him the largest share of assets. (Expressed more formally, the only variables in an individual's utility function are commodities and assets held by him, and not items possessed by others nor their level of utility.)

We may suppose also that the actual division of advantages is determined largely by the balance of power and strategic position resulting from existing circumstances. Yet this division may of course be perfectly fair and satisfy the claims of mutuality. By good fortune the situation may happen to lead to this outcome. Public goods consist largely of those instrumentalities and conditions maintained by the state for everyone to use for his own purposes as his means permit, in the same manner that each has his own destination when traveling along the highways. The theory of competitive markets is a paradigm description of this type of society. Since the members of this society are not moved by the desire to act justly, the stability of just and efficient arrangements when they exist normally requires the use of sanctions. Therefore the alignment of private and collective interests is the result of stabilizing institutional devices applied to persons who oppose one another as indifferent if not hostile

3. The notion of private society, or something like it, is found in many places. Well-known examples are in Plato, *The Republic*, 369–372, and Hegel, *Philosophy of Right*, trans. T. M. Knox (Oxford, The Clarendon Press, 1942), §§182–187, under the heading of civil society. The natural habitat of this notion is in economic theory (general equilibrium), and Hegel's discussion reflects his reading of Adam Smith, *The Wealth of Nations*.

powers. Private society is not held together by a public conviction that its basic arrangements are just and good in themselves, but by the calculations of everyone, or of sufficiently many to maintain the scheme, that any practicable changes would reduce the stock of means whereby they pursue their personal ends.

It is sometimes contended that the contract doctrine entails that private society is the ideal, at least when the division of advantages satisfies a suitable standard of reciprocity. But this is not so, as the notion of a well-ordered society shows. And as I have just said, the idea of the original position has another explanation. The account of goodness as rationality and the social nature of mankind also requires a different view. Now the sociability of human beings must not be understood in a trivial fashion. It does not imply merely that society is necessary for human life, or that by living in a community men acquire needs and interests that prompt them to work together for mutual advantage in certain specific ways allowed for and encouraged by their institutions. Nor is it expressed by the truism that social life is a condition for our developing the ability to speak and think, and to take part in the common activities of society and culture. No doubt even the concepts that we use to describe our plans and situation, and even to give voice to our personal wants and purposes, often presuppose a social setting as well as a system of belief and thought that are the outcome of the collective efforts of a long tradition. These facts are certainly not trivial; but to use them to characterize our ties to one another is to give a trivial interpretation of human sociability. For all of these things are equally true of persons who view their relations purely instrumentally.

The social nature of mankind is best seen by contrast with the conception of private society. Thus human beings have in fact shared final ends and they value their common institutions and activities as good in themselves. We need one another as partners in ways of life that are engaged in for their own sake, and the successes and enjoyments of others are necessary for and complementary to our own good. These matters are evident enough, but they call for some elaboration. In the account of goodness as rationality we came to the familiar conclusion that rational plans of life normally provide for the development of at least some of a person's powers. The Aristotelian Principle points in this direction. Yet one basic characteristic of human beings is that no one person can do everything that he might do; nor a fortiori can he do everything that any other person can do. The potentialities of each individual are greater than those he can hope to realize; and they fall far short of the powers among men gener-

ally. Thus everyone must select which of his abilities and possible interests he wishes to encourage; he must plan their training and exercise, and schedule their pursuit in an orderly way. Different persons with similar or complementary capacities may cooperate so to speak in realizing their common or matching nature. When men are secure in the enjoyment of the exercise of their own powers, they are disposed to appreciate the perfections of others, especially when their several excellences have an agreed place in a form of life the aims of which all accept.

Thus we may say following Humboldt that it is through social union founded upon the needs and potentialities of its members that each person can participate in the total sum of the realized natural assets of the others. We are led to the notion of the community of humankind the members of which enjoy one another's excellences and individuality elicited by free institutions, and they recognize the good of each as an element in the complete activity the whole scheme of which is consented to and gives pleasure to all. This community may also be imagined to extend over time, and therefore in the history of a society the joint contributions of successive generations can be similarly conceived.⁴ Our predecessors in achieving certain things leave it up to us to pursue them further; their accomplishments affect our choice of endeavors and define a wider back-

4. This idea must have occurred to many and is surely implicit in numerous writings. Yet I have been able to find but a few definite formulations of it as expressed in this section. See Wilhelm von Humboldt, *The Limits of State Action*, ed. J. W. Burrow (Cambridge, The University Press, 1969), pp. 16f, for a clear statement. He says: "Every human being, then, can act with only one dominant faculty at a time; or rather, our whole nature disposes us at any given time to some single form of spontaneous activity. It would therefore seem to follow from this, that man is inevitably destined to a partial cultivation, since he only enfeebls his energies by directing them to a multiplicity of objects. But man has it in his power to avoid this one-sidedness, by attempting to unite the distinct and generally separately exercised faculties of his nature, by bringing into spontaneous cooperation, at each period of his life, the dying sparks of one activity, and those which the future will kindle, and endeavoring to increase and diversify the powers with which he works, by harmoniously combining them, instead of looking for mere variety of objects for their separate exercise. What is achieved, in the case of the individual, by the union of past and future with the present, is produced in society by the mutual cooperation of its different members; for, in all stages of his life, each individual can achieve only one of those perfections, which represent the possible features of human character. It is through a social union, therefore, based on the internal wants and capacities of its members, that each is enabled to participate in the rich collective resources of all the others." As a pure case to illustrate this notion of social union, we may consider a group of musicians every one of whom could have trained himself to play equally as well as the others any instrument in the orchestra, but who each have by a kind of tacit agreement set out to perfect their skills on the one they have chosen so as to realize the powers of all in their joint performances. This idea also has a central place in Kant's "Idea for a Universal History," in *Kant's Political Writings*, ed. Hans Reiss and trans. H. B. Nisbet (Cambridge, The University Press, 1970). See pp. 42f where he says that every individual man would have to live for a vast length of time if he were to learn how to make complete use of all his natural capacities, and therefore it will require perhaps an incalculable series of generations of men. I have

ground against which our aims can be understood. To say that man is a historical being is to say that the realizations of the powers of human individuals living at any one time takes the cooperation of many generations (or even societies) over a long period of time. It also implies that this cooperation is guided at any moment by an understanding of what has been done in the past as it is interpreted by social tradition. By contrast with humankind, every individual animal can and does do what for the most part it might do, or what any other of its kind might or can do that lives at the same time. The range of realized abilities of a single individual of the species is not in general materially less than the potentialities of others similar to it. The striking exception is the difference of sex. This is perhaps why sexual affinity is the most obvious example of the need of individuals both human and animal for each other. Yet this attraction may take but a purely instrumental form, each individual treating the other as a means to his own pleasure or the continuation of his line. Unless this attachment is fused with elements of affection and friendship, it will not exhibit the characteristic features of social union.

Now many forms of life possess the characteristics of social union, shared final ends and common activities valued for themselves. Science and art provide ready-to-hand illustrations. Likewise families, friendships, and other groups are social unions. There is some advantage though in thinking about the simpler instances of games. Here we can easily distinguish four sorts of ends: the aim of the game as defined by its

not been able to find this idea expressly stated where I would expect to, for example, in Schiller's *Letters on the Aesthetic Education of Man*, ed. and trans. E. M. Wilkinson and L. A. Willoughby (Oxford, The Clarendon Press, 1967), esp. the sixth and twenty-seventh letters. Nor, I think, in Marx's early writings, particularly the *Economic and Philosophical Manuscripts*. See *Karl Marx: Early Writings*, trans. and ed. T. B. Bottomore (London, C. A. Watts, 1963), pp. 126–129, 154, 156–157, 189, 202f. However, Marx is interpreted to hold a notion like this by Shlomo Avineri, *The Social and Political Thought of Karl Marx* (Cambridge, The University Press, 1969), pp. 231f. Yet Marx tends, I think, to view full communist society as one in which each person completely realizes his nature, in which he himself expresses all of his powers. In any event, it is important not to confuse the idea of social union with the high value put upon human diversity and individuality, as found in Mill's *On Liberty*, ch. III, and in German Romanticism—see A. O. Lovejoy, *The Great Chain of Being* (Cambridge, Harvard University Press, 1936), ch. X; or with the conception of the good as the harmonious fulfillment of natural powers by (complete) individuals; nor, finally, with gifted individuals, artists, and statesmen, and so on, achieving this for the rest of mankind. Rather, in the limiting case where the powers of each are similar, the group achieves, by a coordination of activities among peers, the same totality of capacities latent in each. Or when these powers differ and are in suitable ways complementary, they express the sum of potentialities of the membership as a whole in activities that are intrinsically good and not merely cooperation for social or economic gain. (On this last, see Smith, *Wealth of Nations*, bk. I, chs. I–II.) In either case, persons need one another since it is only in active cooperation with others that one's powers reach fruition. Only in a social union is the individual complete.

rules, say to score the most runs; the various motives of the players in playing the game, the excitement they get from it, the desire for exercise, and so on, which may be different for each person; the social purposes served by the game which may be unintended and unknown to the players, or even to anyone in the society, these being matters for the reflective observer to ascertain; and then finally, the shared end, the common desire of all the players that there should be a good play of the game. This shared end can be realized only if the game is played fairly according to the rules, if the sides are more or less evenly matched, and if the players all sense that they are playing well. But when this aim is attained, everyone takes pleasure and satisfaction in the very same thing. A good play of the game is, so to speak, a collective achievement requiring the cooperation of all.

Now the shared end of a social union is clearly not merely a common desire for the same particular thing. Grant and Lee were one in their desire to hold Richmond but this desire did not establish community between them. Persons generally want similar sorts of things, liberty and opportunity, shelter and nourishment, yet these wants may put them at odds. Whether individuals have a shared end depends upon the more detailed features of the activity to which their interests incline them as these are regulated by principles of justice. There must be an agreed scheme of conduct in which the excellences and enjoyments of each are complementary to the good of all. Each can then take pleasure in the actions of the others as they jointly execute a plan acceptable to everyone. Despite their competitive side, many games illustrate this type of end in a clear way: the public desire to execute a good and fair play of the game must be regulative and effective if everyone's zest and pleasure are not to languish.

The development of art and science, of religion and culture of all kinds, high and low, can of course be thought of in much the same way. Learning from one another's efforts and appreciating their several contributions, human beings gradually build up systems of knowledge and belief; they work out recognized techniques for doing things and elaborate styles of feeling and expression. In these cases the common aim is often profound and complex, being defined by the respective artistic, scientific, or religious tradition; and to understand this aim often takes years of discipline and study. The essential thing is that there be a shared final end and accepted ways of advancing it which allow for the public recognition of the attainments of everyone. When this end is achieved, all find satisfaction in the very same thing; and this fact together with the complementary nature of the good of individuals affirms the tie of community.

I do not wish to stress, however, the cases of art and science, and high forms of religion and culture. In line with the rejection of the principle of perfection and the acceptance of democracy in the assessment of one another's excellences, they have no special merit from the standpoint of justice. Indeed the reference to games not only has the virtue of simplicity but in some ways is more appropriate. It helps to show that the primary concern is that there are many types of social union and from the perspective of political justice we are not to try to rank them in value. Moreover these unions have no definite size; they range from families and friendships to much larger associations. Nor are there limits of time and space, for those widely separated by history and circumstance can nevertheless cooperate in realizing their common nature. A well-ordered society, and indeed most societies, will presumably contain countless social unions of many different kinds.

With these remarks as a preface, we can now see how the principles of justice are related to human sociability. The main idea is simply that a well-ordered society (corresponding to justice as fairness) is itself a form of social union. Indeed, it is a social union of social unions. Both characteristic features are present: the successful carrying out of just institutions is the shared final end of all the members of society, and these institutional forms are prized as good in themselves. Let us consider these features in turn. The first is quite straightforward. In much the same way that players have the shared end to execute a good and fair play of the game, so the members of a well-ordered society have the common aim of cooperating together to realize their own and another's nature in ways allowed by the principles of justice. This collective intention is the consequence of everyone's having an effective sense of justice. Each citizen wants everyone (including himself) to act from principles to which all would agree in an initial situation of equality. This desire is regulative, as the condition of finality on moral principles requires; and when everyone acts justly, all find satisfaction in the very same thing.

The explanation of the second feature is more involved, yet clear enough from what has been said. We have only to note the various ways in which the fundamental institutions of society, the just constitution and the main parts of the legal order, can be found good in themselves once the idea of social union is applied to the basic structure as a whole. Thus first of all, the Kantian interpretation enables us to say that everyone's acting to uphold just institutions is for the good of each. Human beings have a desire to express their nature as free and equal moral persons, and this they do most adequately by acting from the principles that they

would acknowledge in the original position. When all strive to comply with these principles and each succeeds, then individually and collectively their nature as moral persons is most fully realized, and with it their individual and collective good.

But further, the Aristotelian Principle holds for institutional forms as well as for any other human activity. Seen in this light, a just constitutional order, when adjoined to the smaller social unions of everyday life, provides a framework for these many associations and sets up the most complex and diverse activity of all. In a well-ordered society each person understands the first principles that govern the whole scheme as it is to be carried out over many generations; and all have a settled intention to adhere to these principles in their plan of life. Thus the plan of each person is given a more ample and rich structure than it would otherwise have; it is adjusted to the plans of others by mutually acceptable principles. Everyone's more private life is so to speak a plan within a plan, this superordinate plan being realized in the public institutions of society. But this larger plan does not establish a dominant end, such as that of religious unity or the greatest excellence of culture, much less national power and prestige, to which the aims of all individuals and associations are subordinate. The regulative public intention is rather that the constitutional order should realize the principles of justice. And this collective activity, if the Aristotelian Principle is sound, must be experienced as a good.

We have seen that the moral virtues are excellences, attributes of the person that it is rational for persons to want in themselves and in one another as things appreciated for their own sake, or else as exhibited in activities so enjoyed (§§66–67). Now it is clear that these excellences are displayed in the public life of a well-ordered society. Therefore the companion principle to the Aristotelian Principle implies that men appreciate and enjoy these attributes in one another as they are manifested in cooperating to affirm just institutions. It follows that the collective activity of justice is the preeminent form of human flourishing. For given favorable conditions, it is by maintaining these public arrangements that persons best express their nature and achieve the widest regulative excellences of which each is capable. At the same time just institutions allow for and encourage the diverse internal life of associations in which individuals realize their more particular aims. Thus the public realization of justice is a value of community.

As a final comment, I should note that a well-ordered society does not do away with the division of labor in the most general sense. To be sure,

the worst aspects of this division can be surmounted: no one need be servilely dependent on others and made to choose between monotonous and routine occupations which are deadening to human thought and sensibility. Each can be offered a variety of tasks so that the different elements of his nature find a suitable expression. But even when work is meaningful for all, we cannot overcome, nor should we wish to, our dependence on others. In a fully just society persons seek their good in ways peculiar to themselves, and they rely upon their associates to do things they could not have done, as well as things they might have done but did not. It is tempting to suppose that everyone might fully realize his powers and that some at least can become complete exemplars of humanity. But this is impossible. It is a feature of human sociability that we are by ourselves but parts of what we might be. We must look to others to attain the excellences that we must leave aside, or lack altogether. The collective activity of society, the many associations and the public life of the largest community that regulates them, sustains our efforts and elicits our contribution. Yet the good attained from the common culture far exceeds our work in the sense that we cease to be mere fragments: that part of ourselves that we directly realize is joined to a wider and just arrangement the aims of which we affirm. The division of labor is overcome not by each becoming complete in himself, but by willing and meaningful work within a just social union of social unions in which all can freely participate as they so incline.

80. THE PROBLEM OF ENVY

Throughout I have assumed that the persons in the original position are not moved by certain psychological propensities (§25). A rational individual is not subject to envy, at least when the differences between himself and others are not thought to be the result of injustice and do not exceed certain limits. Nor are the parties influenced by different attitudes toward risk and uncertainty, or by various tendencies to dominate or to submit, and the like. These special psychologies I have also imagined to be behind the veil of ignorance along with the parties' knowledge of their conception of the good. One explanation for these stipulations is that as far as possible the choice of a conception of justice should not be affected by accidental contingencies. The principles adopted should be invariant with respect to differences in these inclinations for the same reason that

we want them to hold irrespective of individual preferences and social circumstances.

These assumptions tie in with the Kantian interpretation of justice as fairness and greatly simplify the argument from the standpoint of the original position. The parties are not swayed by individual differences in these propensities, thereby avoiding the complications in the bargaining process that would result. Without rather definite information about which configuration of attitudes existed, one might not be able to say what agreement if any would be reached. In each case it would be contingent upon the particular hypothesis laid down. Unless we could show some distinctive merit from a moral point of view in the postulated array of special psychologies, the principles adopted would be arbitrary, no longer the outcome of reasonable conditions. And since envy is generally regarded as something to be avoided and feared, at least when it becomes intense, it seems desirable that, if possible, the choice of principles should not be influenced by this trait. Therefore, for reasons both of simplicity and moral theory, I have assumed an absence of envy and a lack of knowledge of the special psychologies.

Nevertheless these inclinations do exist and in some way they must be reckoned with. Thus I have split the argument for the principles of justice into two parts: the first part proceeds on the presumptions just mentioned, and is illustrated by most of the argument so far; the second part asks whether the well-ordered society corresponding to the conception adopted will actually generate feelings of envy and patterns of psychological attitudes that will undermine the arrangements it counts to be just. At first we reason as if there is no problem of envy and the special psychologies; and then having ascertained which principles would be settled upon, we check to see whether just institutions so defined are likely to arouse and encourage these propensities to such an extent that the social system becomes unworkable and incompatible with human good. If so, the adoption of the conception of justice must be reconsidered. But should the inclinations engendered support just arrangements, or be easily accommodated by them, the first part of the argument is confirmed. The essential advantage of the two-step procedure is that no particular constellation of attitudes is taken as given. We are simply checking the reasonableness of our initial assumptions and the consequences we have drawn from them in the light of the constraints imposed by the general facts of our world.

I shall discuss the problem of envy as an illustration of the way in

which the special psychologies enter into the theory of justice. While each special psychology raises no doubt different questions, the general procedure may be much the same. I begin by noting the reason why envy poses a problem, namely, the fact that the inequalities sanctioned by the difference principle may be so great as to arouse envy to a socially dangerous extent. To clarify this possibility it is useful to distinguish between general and particular envy. The envy experienced by the least advantaged towards those better situated is normally general envy in the sense that they envy the more favored for the kinds of goods and not for the particular objects they possess. The upper classes say are envied for their greater wealth and opportunity; those envying them want similar advantages for themselves. By contrast, particular envy is typical of rivalry and competition. Those who lose out in the quest for office and honor, or for the affections of another, are liable to envy the success of their rivals and to covet the very same thing that they have won. Our problem then is whether the principles of justice, and especially the difference principle with fair equality of opportunity, is likely to engender in practice too much destructive general envy.

I now turn to the definition of envy that seems appropriate for this question. To fix ideas, suppose that the necessary interpersonal comparisons are made in terms of the objective primary goods, liberty and opportunity, income and wealth, which for simplicity I have normally used to define expectations in applying the difference principle. Then we may think of envy as the propensity to view with hostility the greater good of others even though their being more fortunate than we are does not detract from our advantages. We envy persons whose situation is superior to ours (estimated by some agreed index of goods as noted above) and we are willing to deprive them of their greater benefits even if it is necessary to give up something ourselves. When others are aware of our envy, they may become jealous of their better circumstances and anxious to take precautions against the hostile acts to which our envy makes us prone. So understood envy is collectively disadvantageous: the individual who envies another is prepared to do things that make them both worse off, if only the discrepancy between them is sufficiently reduced. Thus Kant, whose definition I have pretty much followed, quite properly discusses envy as one of the vices of hating mankind.⁵

This definition calls for comment. First of all, as Kant observes, there

5. *The Metaphysics of Morals*, pt. II, §36. In the edition trans. M. G. Gregor (New York, Harper and Row, 1964), p. 127. Aristotle notes that envy and spite as passions do not admit of a mean; their names already imply badness. *Nicomachean Ethics*, 1107a11.

are many occasions when we openly speak of the greater good of others as enviable. Thus we may remark upon the enviable harmony and happiness of a marriage or a family. Similarly, one might say to another that one envies his greater opportunities or attainments. In these cases, those of benign envy as I shall refer to them, there is no ill will intended or expressed. We do not wish, for example, that the marriage or family should be less happy or harmonious. By these conventional expressions we are affirming the value of certain things that others have. We are indicating that, although we possess no similar good of equal value, they are indeed worth striving for. Those to whom we address these remarks are expected to receive them as a kind of praise and not as a foretaste of our hostility. A somewhat different case is that of emulative envy which leads us to try to achieve what others have. The sight of their greater good moves us to strive in socially beneficial ways for similar things for ourselves.⁶ Thus envy proper, in contrast with benign envy which we freely express, is a form of rancor that tends to harm both its object and its subject. It is what emulative envy may become under certain conditions of defeat and sense of failure.

A further point is that envy is not a moral feeling. No moral principle need be cited in its explanation. It is sufficient to say that the better situation of others catches our attention. We are downcast by their good fortune and no longer value as highly what we have; and this sense of hurt and loss arouses our rancor and hostility. Thus one must be careful not to conflate envy and resentment. For resentment is a moral feeling. If we resent our having less than others, it must be because we think that their being better off is the result of unjust institutions, or wrongful conduct on their part. Those who express resentment must be prepared to show why certain institutions are unjust or how others have injured them. What marks off envy from the moral feelings is the different way in which it is accounted for, the sort of perspective from which the situation is viewed (§73).

We should note also the nonmoral feelings connected with envy but not to be mistaken for it. In particular, jealousy and grudgingness are reverse, so to speak, to envy. A person who is better off may wish those less fortunate than he to stay in their place. He is jealous of his superior position and begrudges them the greater advantages that would put them on a level with himself. And should this propensity extend to denying

6. For the distinction between emulation and envy, see Bishop Butler, *Sermons*, I, in *British Moralists*, ed. L. A. Selby-Bigge (Oxford, 1897), vol. I, p. 205.

them benefits that he does not need and cannot use himself, then he is moved by spite.⁷ These inclinations are collectively harmful in the way that envy is, since the grudging and spiteful man is willing to give up something to maintain the distance between himself and others.

So far I have considered envy and grudgingness as vices. As we have seen, the moral virtues are among the broadly based traits of character which it is rational for persons to want in one another as associates (§66). Thus vices are broadly based traits that are not wanted, spitefulness and envy being clear cases, since they are to everyone's detriment. The parties will surely prefer conceptions of justice the realization of which does not arouse these propensities. We are normally expected to forbear from the actions to which they prompt us and to take the steps necessary to rid ourselves of them. Yet sometimes the circumstances evoking envy are so compelling that given human beings as they are no one can reasonably be asked to overcome his rancorous feelings. A person's lesser position as measured by the index of objective primary goods may be so great as to wound his self-respect; and given his situation, we may sympathize with his sense of loss. Indeed, we can resent being made envious, for society may permit such large disparities in these goods that under existing social conditions these differences cannot help but cause a loss of self-esteem. For those suffering this hurt, envious feelings are not irrational; the satisfaction of their rancor would make them better off. When envy is a reaction to the loss of self-respect in circumstances where it would be unreasonable to expect someone to feel differently, I shall say that it is excusable. Since self-respect is the main primary good, the parties would not agree, I shall assume, to count this sort of subjective loss as irrelevant. Therefore the question is whether a basic structure which satisfies the principles of justice is likely to arouse so much excusable envy that the choice of these principles should be reconsidered.

81. ENVY AND EQUALITY

We are now ready to examine the likelihood of excusable general envy in a well-ordered society. I shall only discuss this case, since our problem is whether the principles of justice are a reasonable undertaking in view of the propensities of human beings, in particular their aversion to dispari-

7. Aristotle, *Nicomachean Ethics*, 1108b1–6, characterizes spite as being pleased at the bad fortune of others, whether deserved or not. For the idea that jealousy, grudgingness, and spite are the reverse of envy, the feelings of those envied and who possess what is wanted, I am indebted to G. M. Foster.

ties in objective goods. Now I assume that the main psychological root of the liability to envy is a lack of self-confidence in our own worth combined with a sense of impotence. Our way of life is without zest and we feel powerless to alter it or to acquire the means of doing what we still want to do.⁸ By contrast, someone sure of the worth of his plan of life and his ability to carry it out is not given to rancor nor is he jealous of his good fortune. Even if he could, he has no desire to level down the advantages of others at some expense to himself. This hypothesis implies that the least favored tend to be more envious of the better situation of the more favored the less secure their self-respect and the greater their feeling that they cannot improve their prospects. Similarly the particular envy aroused by competition and rivalry is likely to be stronger the worse one's defeat, for the blow to one's self-confidence is more severe and the loss may seem irretrievable. It is general envy, however, that mainly concerns us here.

There are three conditions, I assume, that encourage hostile outbreaks of envy. The first of these is the psychological condition we have just noted: persons lack a sure confidence in their own value and in their ability to do anything worthwhile. Second (and one of two social conditions), many occasions arise when this psychological condition is experienced as painful and humiliating. The discrepancy between oneself and others is made visible by the social structure and style of life of one's society. The less fortunate are therefore often forcibly reminded of their situation, sometimes leading them to an even lower estimation of themselves and their mode of living. And third, they see their social position as allowing no constructive alternative to opposing the favored circumstances of the more advantaged. To alleviate their feelings of anguish and inferiority, they believe they have no choice but to impose a loss on those better placed even at some cost to themselves, unless of course they are to relapse into resignation and apathy.

Now many aspects of a well-ordered society work to mitigate if not to prevent these conditions. In regard to the first condition, it is clear that, although it is a psychological state, social institutions are a basic instigating cause. But I have maintained that the contract conception of justice

8. This sort of hypothesis has been proposed by various writers. See, for example, Nietzsche, *On the Genealogy of Morals*, trans. Walter Kaufmann and R. J. Hollingdale (New York, Random House, 1967), I, secs. 10, 11, 13, 14, 16; II, sec. 11; III, secs. 14–16; and Max Scheler, *Ressentiment*, trans. W. W. Holdheim (Glencoe, Ill., The Free Press, 1961), pp. 45–50. For a discussion of Nietzsche's notion of resentment, see Walter Kaufmann, *Nietzsche* (Princeton, Princeton University Press, 1950), pp. 325–331.

supports the self-esteem of citizens generally more firmly than other political principles. In the public forum each person is treated with the respect due to a sovereign equal; and everyone has the same basic rights that would be acknowledged in an initial situation regarded as fair. The members of the community have a common sense of justice and they are bound by ties of civic friendship. I have already discussed these points in connection with stability (§§75–76). We can add that the greater advantages of some are in return for compensating benefits for the less favored; and no one supposes that those who have a larger share are more deserving from a moral point of view. Happiness according to virtue is rejected as a principle of distribution (§48). And so likewise is the principle of perfection: regardless of the excellences that persons or associations display, their claims to social resources are always adjudicated by principles of mutual justice (§50). For all these reasons the less fortunate have no cause to consider themselves inferior and the public principles generally accepted underwrite their self-assurance. The disparities between themselves and others, whether absolute or relative, should be easier for them to accept than in other forms of polity.

Turning to the second condition, both the absolute and the relative differences allowed in a well-ordered society are probably less than those that have often prevailed. Although in theory the difference principle permits indefinitely large inequalities in return for small gains to the less favored, the spread of income and wealth should not be excessive in practice, given the requisite background institutions (§26). Moreover the plurality of associations in a well-ordered society, each with its secure internal life, tends to reduce the visibility, or at least the painful visibility, of variations in men's prospects. For we tend to compare our circumstances with others in the same or in a similar group as ourselves, or in positions that we regard as relevant to our aspirations. The various associations in society tend to divide it into so many noncomparing groups, the discrepancies between these divisions not attracting the kind of attention which unsettles the lives of those less well placed. And this ignoring of differences in wealth and circumstance is made easier by the fact that when citizens do meet one another, as they must in public affairs at least, the principles of equal justice are acknowledged. Moreover in everyday life the natural duties are honored so that the more advantaged do not make an ostentatious display of their higher estate calculated to demean the condition of those who have less. After all, if the disposing conditions for envy are removed, so probably are those for jealousy, grudgingness, and spite, the converses of envy. When the less fortunate segments of

society lack the one, the more fortunate will lack the other. Taken together these features of a well-ordered regime diminish the number of occasions when the less favored are likely to experience their situation as impoverished and humiliating. Even if they have some liability to envy, it may never be strongly evoked.

Finally, considering the last condition, it would seem that a well-ordered society as much as any other offers constructive alternatives to hostile outbreaks of envy. The problem of general envy anyway does not force us to reconsider the choice of the principles of justice. As for particular envy, to a certain extent it is endemic to human life; being associated with rivalry, it may exist in any society. The more specific problem for political justice is how pervasive are the rancor and jealousy aroused by the quest for office and position, and whether it is likely to distort the justice of institutions. It is difficult to settle this matter in the absence of the more detailed knowledge of social forms available at the legislative stage. But there seems to be no reason why the hazards of particular envy should be worse in a society regulated by justice as fairness than by any other conception.

I conclude, then, that the principles of justice are not likely to arouse excusable general envy (nor particular envy either) to a troublesome extent. By this test, the conception of justice again seems relatively stable. I should now like to examine briefly the possible connections between envy and equality, taking equality to be defined in various ways as specified by the theory of justice in question. While there are many forms of equality, and egalitarianism admits of degrees, there are conceptions of justice that are recognizably egalitarian, even though certain significant disparities are permitted. The two principles of justice fall, I assume, under this heading.

Many conservative writers have contended that the tendency to equality in modern social movements is the expression of envy.⁹ In this way they seek to discredit this trend, attributing it to collectively harmful impulses. Before this thesis can be seriously entertained, however, one must first argue that the form of equality objected to is indeed unjust and bound in the end to make everyone including the less advantaged worse off. Yet to insist upon equality as the two principles of justice define it is not to give voice to envy. This is shown by the content of these principles

9. See, for example, Helmut Schoeck, *Envy: A Theory of Social Behavior*, trans. Michael Glenny and Betty Ross (London, Secker and Warburg, 1969). Chapters XIV–XV contain many references. At one point even Marx thought of the first stage of communism as the expression of envy. See *Early Writings*, pp. 153f.

and the characterization of envy. It is also evident from the nature of the parties in the original position: the conception of justice is chosen under conditions where by hypothesis no one is moved by rancor and spite (§25). Thus the claims to equality supported by the two principles do not spring from these feelings. The claims of those affirming the principles may sometimes express resentment, but this as we have seen is another matter.

In order to show that the principles of justice are based in part on envy it would have to be established that one or more of the conditions of the original position arise from this propensity. Since the question of stability does not force a reconsideration of the choice already made, the case for the influence of envy must be made by reference to the first part of the theory. But each of the stipulations of the original position has a justification which makes no mention of envy. For example, one invokes the function of moral principles as being a suitably general and public way of ordering claims (§23). To be sure, there may be forms of equality that do spring from envy. Strict egalitarianism, the doctrine which insists upon an equal distribution of all primary goods, conceivably derives from this propensity. What this means is that this conception of equality would be adopted in the original position only if the parties are assumed to be sufficiently envious. This possibility in no way affects the two principles of justice. The different conception of equality which they define is acknowledged on the supposition that envy does not exist.¹⁰

The importance of separating envy from the moral feelings can be seen from several examples. Suppose first that envy is held to be pervasive in poor peasant societies. The reason for this, it may be suggested, is the general belief that the aggregate of social wealth is more or less fixed, so that one person's gain is another's loss. The social system is interpreted, it might be said, as a naturally established and unchangeable zero-sum game. Now actually, if this belief were widespread and the stock of goods were generally thought to be given, then a strict opposition of interests would be assumed to obtain. In this case, it would be correct to think that justice requires equal shares. Social wealth is not viewed as the outcome of mutually advantageous cooperation and so there is no fair basis for an unequal division of advantages. What is said to be envy may in fact be resentment which might or might not prove to be justified.

Freud's speculations about the origin of the sense of justice suffer from the same defect. He remarks that this sentiment is the outgrowth of envy

10. In this and the next several paragraphs I am indebted to R. A. Schultz for helpful suggestions.

and jealousy. As some members of the social group jealously strive to protect their advantages, the less favored are moved by envy to take them away. Eventually everyone recognizes that they cannot maintain their hostile attitudes toward one another without injury to themselves. Thus as a compromise they settle upon the demand of equal treatment. The sense of justice is a reaction-formation: what was originally jealousy and envy is transformed into a social feeling, the sense of justice that insists upon equality for all. Freud believes that this process is exemplified in the nursery and in many other social circumstances.¹¹ Yet the plausibility of his account assumes that the initial attitudes are correctly described. With a few changes, the underlying features of the examples he depicts correspond to those of the original position. That persons have opposing interests and seek to advance their own conception of the good is not at all the same thing as their being moved by envy and jealousy. As we have seen, this sort of opposition gives rise to the circumstances of justice. Thus if children compete for the attention and affection of their parents, to which one might say they justly have an equal claim, one cannot assert that their sense of justice springs from jealousy and envy. Certainly children are often envious and jealous; and no doubt their moral notions are so primitive that the necessary distinctions are not grasped by them. But waiving these difficulties, we could equally well say that their social feeling arises from resentment, from a sense that they are unfairly treated.¹² And similarly one could say to conservative writers that it is mere grudgingness when those better circumstanced reject the claims of the less advantaged to greater equality. But this contention also calls for careful argument. None of these charges and countercharges can be given credence without first examining the conceptions of justice sincerely held by individuals and their understanding of the social situation in order to see how far these claims are indeed founded on these motives.

None of these remarks is intended to deny that the appeal to justice is often a mask for envy. What is said to be resentment may really be rancor. But rationalizations of this sort present a further problem. In addition to showing that a person's conception of justice is not itself founded on envy, we must determine whether the principles of justice cited in his explanation are sincerely held as this is shown in their being applied by him to other cases in which he is not involved, or even better, in which he

11. See *Group Psychology and the Analysis of the Ego*, rev. ed., trans. James Strachey (London, The Hogarth Press, 1959), pp. 51f.

12. See Rousseau, *Emile*, trans. Barbara Foxley (London, J. M. Dent and Sons, 1911), pp. 61–63. And also J. N. Shklar, *Men and Citizens* (Cambridge, The University Press, 1969), p. 49.

would suffer a loss from their being followed. Freud means to assert more than the truism that envy often masquerades as resentment. He wants to say that the energy that motivates the sense of justice is borrowed from that of envy and jealousy, and that without this energy, there would be no (or much less) desire to give justice. Conceptions of justice have few attractions for us other than those deriving from these and similar feelings. It is this claim that is supported by erroneously conflating envy and resentment.

Unhappily the problem of the other special psychologies must go untouched. They should in any case be treated in much the same way as envy. One tries to assess the configuration of attitudes toward risk and uncertainty, domination and submission, and the like, that just institutions are likely to generate, and then to estimate whether they are likely to render these institutions unworkable or ineffective. We also need to ask whether, from the point of view of the persons in the original position, the conception chosen is acceptable or at least tolerable whatever our special proclivities may turn out to be. The most favorable alternative is that which allows a place for all these different tendencies insofar as they are likely to be encouraged by a just basic structure. There is a division of labor so to speak between persons with contrary inclinations. Of course some of these attitudes may earn a premium in the way that certain trained abilities do, as for example the willingness to be adventuresome and to take unusual risks. But if so, the problem is on all fours with the return to natural assets and it is covered by the discussion of distributive shares (§47). What a social system must not do clearly is to encourage propensities and aspirations that it is bound to repress and disappoint. So long as the pattern of special psychologies elicited by society either supports its arrangements or can be reasonably accommodated by them, there is no need to reconsider the choice of a conception of justice. I believe, though I have not shown, that the principles of justice as fairness pass this test.

82. THE GROUNDS FOR THE PRIORITY OF LIBERTY

We have already considered the meaning of the priority of liberty and how it is incorporated into various rules of precedence (§§39, 46). Now that all the main elements of the contract view are before us, the main grounds for this priority can be surveyed. I have supposed that if the persons in the original position know that their basic liberties can be ef-

fectively exercised, they will not exchange a lesser liberty for greater economic advantages (§26). It is only when social conditions do not allow the full establishment of these rights that one can acknowledge their restriction. The equal liberties can be denied only when it is necessary to change the quality of civilization so that in due course everyone can enjoy these freedoms. The effective realization of all these liberties in a well-ordered society is the long-run tendency of the two principles and rules of priority when they are consistently followed under reasonably favorable conditions. Our problem here, then, is to summarize and arrange the reasons for the precedence of liberty in a well-ordered society as seen from the point of view of the original position.

Let us begin by recalling the reasons contained in the first part of the argument for the two principles. A well-ordered society is defined as one effectively regulated by a public conception of justice (§69). The members of such a society are, and view themselves as, free and equal moral persons. That is, they each have, and view themselves as having, fundamental aims and interests in the name of which they think it legitimate to make claims on one another; and they each have, and view themselves as having, a right to equal respect and consideration in determining the principles by which the basic structure of their society is to be governed. They also have a sense of justice that normally governs their conduct. The original position is specified to embody the appropriate reciprocity and equality between persons so conceived; and given that their fundamental aims and interests are protected by the liberties covered by the first principle, they give this principle priority. The religious interest as guaranteed by equal liberty of conscience was discussed as an example (§§33–35). In this connection, one should keep in mind that the parties seek to secure some particular fundamental interest, even though, given the veil of ignorance, only the general nature of this interest is known to them, for example, that it is a religious interest. Their aim is not merely to be permitted to practice some religion or other, but to practice some definite religion, that is, their religion, whatever it turns out to be (§28). In order to secure their unknown but particular interests from the original position, they are led, in view of the strains of commitment (§29), to give precedence to the basic liberties.

A well-ordered society also realizes the parties' highest-order interest in how their other interests, including even their fundamental ones, are shaped and regulated by social institutions (§26). The parties conceive of themselves as free persons who can revise and alter their final ends and who give priority to preserving their liberty in this respect. The manner in

which the principles of justice govern the basic structure, as illustrated by the account of autonomy and objectivity (§78), shows that this highest-order interest is achieved in a well-ordered society.

Thus the persons in the original position are moved by a certain hierarchy of interests. They must first secure their highest-order interest and fundamental aims (only the general form of which is known to them), and this fact is reflected in the precedence they give to liberty; the acquisition of means that enable them to advance their other desires and ends has a subordinate place. Even though the fundamental interests in liberty have a definite objective, namely, the effective establishment of the basic liberties, these interests may not always appear to be controlling. The realization of these interests may necessitate certain social conditions and degree of fulfillment of needs and material wants, and this explains why the basic liberties can sometimes be restricted. But once the required social conditions and level of satisfaction of needs and material wants is attained, as they are in a well-ordered society under favorable circumstances, the higher-order interests are regulative from then on. Indeed, as Mill supposed, these interests become more intense as the situation of society enables them to be expressed effectively, so that eventually they are regulative and reveal their prior place.¹³ The basic structure is then to secure the free internal life of the various communities of interests in which persons and groups seek to achieve, in forms of social union consistent with equal liberty, the ends and excellences to which they are drawn (§79). People want to exercise control over the laws and rules that govern their association, either by directly taking part themselves in its affairs or indirectly through representatives with whom they are affiliated by ties of culture and social situation.

So much for the grounds of the precedence of liberty covered in the first part of the argument for the two principles of justice. We must now turn to the second part of the argument and ask whether this precedence will be undermined by the various feelings and attitudes that are likely to be generated within a well-ordered society (§80). Now it may seem that even when the essential needs are satisfied and the requisite material means are attained, people's concern for their relative position in the distribution of wealth will persist. Thus if we suppose that everyone

13. See J. S. Mill, *Principles of Political Economy*, ed. by W. S. Ashley (London, Longmans Green, 1909), p. 210. The reference is to the first part of the last paragraph of §3 of ch. 1 of bk. II. If we read this passage to imply the notion of a hierarchy of interests, which leads to a lexical ordering, the view I express in the text is essentially Mill's. His contention here fits the passage in *Utilitarianism*, ch. II, pars. 6–8, which was cited along with other references in footnote 23 of ch. I.

wants a greater proportionate share, the result could be a growing desire for material abundance all the same. Since each strives for an end that cannot be collectively attained, society might conceivably become more and more preoccupied with raising productivity and improving economic efficiency. And these objectives might become so dominant as to undermine the precedence of liberty. Some have objected to the tendency to equality on precisely this ground, that it is thought to arouse in individuals an obsession with their relative share of social wealth. But while it is true that in a well-ordered society there is presumably a trend to greater equality, its members take little interest in their relative position as such. As we have seen, they are not much affected by envy and jealousy, and for the most part they do what seems best to them as judged by their own plan of life, and those of their associates, without being dismayed by the greater amenities and enjoyments of others socially more distant. Thus there are no strong propensities prompting them to curtail their liberties for the sake of greater absolute or relative economic welfare.

Of course, it does not follow that in a just society everyone is unconcerned with matters of status. The account of self-respect as perhaps the main primary good has stressed the great significance of how we think others value us. But in a well-ordered society the need for status is met by the public recognition of just institutions, together with the full and diverse internal life of the many free communities of interests that the equal liberties allow. The basis for self-respect in a just society is not then one's income share but the publicly affirmed distribution of fundamental rights and liberties. And this distribution being equal, everyone has a similar and secure status when they meet to conduct the common affairs of the wider society. No one is inclined to look beyond the constitutional affirmation of equality for further political ways of securing his status. Nor, on the other hand, are men disposed to acknowledge a less than equal liberty. For one thing, doing this would put them at a disadvantage and weaken their political position. It would also have the effect of publicly establishing their inferiority as defined by the basic structure of society. This subordinate ranking in public life would indeed be humiliating and destructive of self-esteem. And so by acquiescing in a less than equal liberty one might lose on both counts. This is particularly likely to be true as a society becomes more just, since equal rights and the public attitudes of mutual respect have an essential place in maintaining a political balance and in assuring citizens of their own worth. Thus while the social and economic differences between the various sectors of society, the noncomparing groups as we may think of them, are not likely to generate

animosity, the hardships arising from political and civic inequality, and from cultural and ethnic discrimination, cannot be easily accepted. When it is the position of equal citizenship that answers to the need for status, the precedence of the equal liberties becomes all the more necessary. Having chosen a conception of justice that seeks to eliminate the significance of relative economic and social advantages as supports for men's self-confidence, it is essential that the priority of liberty be firmly maintained.

In a well-ordered society then self-respect is secured by the public affirmation of the status of equal citizenship for all; the distribution of material means is left to take care of itself in accordance with pure procedural justice regulated by just background institutions which narrow the range of inequalities so that excusable envy does not arise. Now this way of dealing with the problem of status has several advantages. Thus, suppose that how one is valued by others did depend upon one's relative place in the distribution of income and wealth. In this case having a higher status implies having more material means than a larger fraction of society. Everyone cannot have the highest status, and to improve one person's position is to lower that of someone else. Social cooperation to increase the conditions of self-respect is impossible. The means of status, so to speak, are fixed, and each man's gain is another's loss. Clearly this situation is a great misfortune. Persons are set at odds with one another in the pursuit of their self-esteem. Given the preeminence of this primary good, the parties in the original position surely do not want to find themselves so opposed. It would tend to make the good of social union difficult if not impossible to achieve. The best solution is to support the primary good of self-respect as far as possible by the assignment of the basic liberties that can indeed be made equal, defining the same status for all. At the same time, relative shares of material means are relegated to a subordinate place. Thus we arrive at another reason for factoring the social order into two parts as indicated by the principles of justice. While these principles permit inequalities in return for contributions that are for the benefit of all, the precedence of liberty entails equality in the social bases of respect.

Now it is quite possible that this idea cannot be carried through completely. To some extent men's sense of their own worth may hinge upon their institutional position and their income share. If, however, the account of social envy and jealousy is sound, then with the appropriate background arrangements, these inclinations should not be excessive. But theoretically we can if necessary include self-respect in the primary

goods, the index of which defines expectations. Then in applications of the difference principle, this index can allow for the effects of excusable envy (§80); the expectations of the less advantaged are lower the more severe these effects. Whether some adjustment for self-respect has to be made is best decided from the standpoint of the legislative stage where the parties have more information and the principle of political determination applies. Admittedly this problem is an unwelcome complication. Since simplicity is itself desirable in a public conception of justice (§49), the conditions that elicit excusable envy should if possible be avoided. I have mentioned this point not to settle it, but only to note that when necessary the expectations of the less advantaged can be understood so as to include the primary good of self-respect.

The second part of the argument seems to confirm, then, the priority of liberty. Still, some may object to this account of the priority of liberty that societies have other ways of affirming self-respect and of coping with envy and other disruptive inclinations. Thus in a feudal or in a caste system each person is believed to have his allotted station in the natural order of things. His comparisons are presumably confined to within his own estate or caste, these ranks becoming in effect so many noncomparing groups established independently of human control and sanctioned by religion and theology. Men resign themselves to their position should it ever occur to them to question it; and since all may view themselves as assigned their vocation, everyone is held to be equally fated and equally noble in the eyes of providence.¹⁴ This conception of society solves the problem of social justice by eliminating in thought the circumstances that give rise to it. The basic structure is said to be already determined, and not something for human beings to affect. On this view, it misconceives men's place in the world to suppose that the social order should match principles which they would as equals consent to.

Contrary to this idea, I have assumed all along that the parties are to be guided in their adoption of a conception of justice by a knowledge of the general facts about society. They take for granted then that institutions are not fixed but change over time, altered by natural circumstances and the activities and conflicts of social groups. The constraints of nature are recognized, but men are not powerless to shape their social arrangements.

14. On this point, see Max Weber, *Economy and Society*, ed. Guenther Roth and Claus Wittich (New York, Bedminster Press, 1968), vol. II, pp. 435f, 598f. See pp. 490–499 for general comments on the things looked for in religions by different social strata. Also consult Ernst Troeltsch, *The Social Teaching of the Christian Churches*, trans. Olive Wyon (London, George Allen and Unwin, 1931), vol. I, pp. 120–127, 132f, 134–138; and Scheler, *Ressentiment*, pp. 56f.

This assumption is likewise part of the background of the theory of justice. It follows that certain ways of dealing with envy and other aberrant propensities are closed to a well-ordered society. For example, it cannot keep them in check by promulgating false or unfounded beliefs. For our problem is how society should be arranged if it is to conform to principles that rational persons with true general beliefs would acknowledge in the original position. The publicity condition requires the parties to assume that as members of society they will also know the general facts. The reasoning leading up to the initial agreement is to be accessible to public understanding. Of course, in working out what the requisite principles are, we must rely upon current knowledge as recognized by common sense and the existing scientific consensus. But there is no reasonable alternative to doing this. We have to concede that as established beliefs change, it is possible that the principles of justice which it seems rational to acknowledge may likewise change. Thus when the belief in a fixed natural order sanctioning a hierarchical society is abandoned, assuming here that this belief is not true, a tendency is set up in the direction of the two principles of justice in serial order. The effective protection of the equal liberties becomes increasingly of first importance in support of self-respect and this affirms the precedence of the first principle.

83. HAPPINESS AND DOMINANT ENDS

In order to be in a position to take up the question of the good of justice, I shall discuss the manner in which just institutions frame our choice of a rational plan and incorporate the regulative element of our good. I shall approach this topic in a roundabout fashion by returning in this section to the concept of happiness and noting the temptation to think of it as determined by a dominant end. Doing this will lead naturally into the problems of hedonism and of the unity of the self. How these matters are connected should be apparent in due course.

Earlier I said that, with certain qualifications, a person is happy when he is in the way of a successful execution (more or less) of a rational plan of life drawn up under (more or less) favorable conditions, and he is reasonably confident that his intentions can be carried through (§63). Thus we are happy when our rational plans are going well, our more important aims being fulfilled, and we are with reason quite sure that our good fortune will continue. The achievement of happiness depends upon

circumstances and luck, and hence the gloss about favorable conditions. While I shall not discuss the concept of happiness in any detail, we should consider a few further points to bring out the connection with the problem of hedonism.

First of all, happiness has two aspects: one is the successful execution of a rational plan (the schedule of activities and aims) which a person strives to realize, the other is his state of mind, his sure confidence supported by good reasons that his success will endure. Being happy involves both a certain achievement in action and a rational assurance about the outcome.¹⁵ This definition of happiness is objective: plans are to be adjusted to the conditions of our life and our confidence must rest upon sound beliefs. Alternatively, happiness might be defined subjectively as follows: a person is happy when he believes that he is in the way of a successful execution (more or less) of a rational plan, and so on as before, adding the rider that if he is mistaken or deluded, then by contingency and coincidence nothing happens to disabuse him of his misconceptions. By good luck he is not cast out of his fool's paradise. Now the definition to be preferred is that which best fits the theory of justice and coheres with our considered judgments of value. At this point it suffices to observe, as I indicated a few pages back (§82), that we have assumed that the parties in the original position have correct beliefs. They acknowledge a conception of justice in the light of general truths about persons and their place in society. Thus it seems natural to suppose that in framing their plans of life they are similarly lucid. Of course none of this is strictly argument. Eventually one has to appraise the objective definition as a part of the moral theory to which it belongs.

Adopting this definition, and keeping in mind the account of rational plans presented earlier (§§63–65), we can interpret the special characteristics sometimes attributed to happiness.¹⁶ For example, happiness is self-contained: that is, it is chosen solely for its own sake. To be sure, a rational plan will include many (or at least several) final aims, and any of these may be pursued partly because it complements and furthers one or more other aims as well. Mutual support among ends pursued for their own sake is an important feature of rational plans, and therefore these ends are not usually sought solely for themselves. Nevertheless executing

15. For this point see Anthony Kenny, "Happiness," *Proceedings of the Aristotelian Society*, vol. 66 (1965–1966), pp. 101f.

16. Notably by Aristotle, *Nicomachean Ethics*, 1097a15–b21. For a discussion of Aristotle's account of happiness, see W. F. R. Hardie, *Aristotle's Ethical Theory* (Oxford, The Clarendon Press, 1968), ch. II.

the entire plan, and the enduring confidence with which this is done, is something that we want to do and to have only for itself. All considerations including those of right and justice (using here the full theory of good) have already been surveyed in drawing up the plan. And therefore the whole activity is self-contained.

Happiness is also self-sufficient: a rational plan when realized with assurance makes a life fully worthy of choice and demands nothing further in addition. When circumstances are especially favorable and the execution particularly successful, one's happiness is complete. Within the general conception one sought to follow, there is nothing essential that is lacking, no way in which it could have been distinctly better. So even if the material means that support our mode of life can always be imagined to be greater, and a different pattern of aims might often have been chosen, still the actual fulfillment of the plan itself may have, as compositions, paintings, and poems often do, a certain completeness which though marred by circumstance and human failing is evident from the whole. Thus some become exemplars of human flourishing and models for emulation, their lives being as instructive in how to live as any philosophical doctrine.

A person is happy then during those periods when he is successfully carrying through a rational plan and he is with reason confident that his efforts will come to fruition. He may be said to approach blessedness to the extent that conditions are supremely favorable and his life complete. Yet it does not follow that in advancing a rational plan one is pursuing happiness, not at least as this is normally meant. For one thing, happiness is not one aim among others that we aspire to, but the fulfillment of the whole design itself. But also I have supposed first that rational plans satisfy the constraints of right and justice (as the full theory of the good stipulates). To say of someone that he seeks happiness does not, it seems, imply that he is prepared either to violate or to affirm these restrictions. Therefore the acceptance of these limits should be made explicit. And secondly, the pursuit of happiness often suggests the pursuit of certain sorts of ends, for example, life, liberty, and one's own welfare.¹⁷ Thus persons who devote themselves selflessly to a righteous cause, or who dedicate their lives to furthering the well-being of others, are not normally thought to seek happiness. It would be misleading to say this of saints and heroes, or of those whose plan of life is in some marked degree supererogatory. They do not have the kinds of aims that fall under this

17. For these two qualifications, see Kenny, "Happiness," pp. 98f.

heading, admittedly not sharply defined. Yet saints and heroes, and persons whose intentions acknowledge the limits of right and justice, are in fact happy when their plans succeed. Although they do not strive for happiness, they may nevertheless be happy in advancing the claims of justice and the well-being of others, or in attaining the excellences to which they are attracted.

But how in general is it possible to choose among plans rationally? What procedure can an individual follow when faced with this sort of decision? I now want to return to this question. Previously I said that a rational plan is one that would be chosen with deliberative rationality from among the class of plans all of which satisfy the principles of rational choice and stand up to certain forms of critical reflection. We eventually reach a point though where we just have to decide which plan we most prefer without further guidance from principle (§64). There is however one device of deliberation that I have not yet mentioned, and this is to analyze our aims. That is, we can try to find a more detailed or more illuminating description of the object of our desires hoping that the counting principles will then settle the case. Thus it may happen that a fuller or deeper characterization of what we want discloses that an inclusive plan exists after all.

Let us consider again the example of planning a holiday (§63). Often when we ask ourselves why we wish to visit two distinct places, we find that certain more general ends stand in the background and that all of them can be fulfilled by going to one place rather than the other. Thus we may want to study certain styles of art, and further reflection may bring out that one plan is superior or equally good on all these counts. In this sense we may discover that our desire to go to Paris is more intense than our desire to go to Rome. Often however a finer description fails to be decisive. If we want to see both the most famous church in Christendom and the most famous museum, we may be stuck. Of course these desires too may be examined further. Nothing in the way that most desires are expressed shows whether there exists a more revealing characterization of what we really want. But we have to allow for the possibility, indeed for the probability, that sooner or later we will reach incomparable aims between which we must choose with deliberative rationality. We may trim, reshape, and transform our aims in a variety of ways as we try to fit them together. Using the principles of rational choice as guidelines, and formulating our desires in the most lucid form we can, we may narrow the scope of purely preferential choice, but we cannot eliminate it altogether.

The indeterminacy of decision seems to arise, then, from the fact that a person has many aims for which there is no ready standard of comparison to decide between them when they conflict. There are many stopping points in practical deliberation and many ways in which we characterize the things we want for their own sake. Thus it is easy to see why the idea of there being a single dominant end (as opposed to an inclusive end) at which it is rational to aim is highly appealing.¹⁸ For if there exists such an end to which all other ends are subordinate, then presumably all desires, insofar as they are rational, admit of an analysis which shows the counting principles to apply. The procedure for making a rational choice, and the conception of such a choice, would then be perfectly clear: deliberation would always concern means to ends, all lesser ends in turn being ordered as means to one single dominant end. The many finite chains of reasons eventually converge and meet at the same point. Hence a rational decision is always in principle possible, since only difficulties of computation and lack of information remain.

Now it is essential to understand what the dominant-end theorist wants: namely, a method of choice which the agent himself can always follow in order to make a rational decision. Thus there are three requirements: the conception of deliberation must specify (1) a first-person procedure which is (2) generally applicable and (3) guaranteed to lead to the best result (at least under favorable conditions of information and given the ability to calculate). We have no procedures meeting these conditions. A random device provides a general method but it would be rational only in special circumstances. In everyday life we employ schemes of deliberation acquired from our culture and modified during the course of our personal history. But there is no assurance that these forms of reflection are rational. Perhaps they only meet various minimum standards which enable us to get by, all the while falling far short of the best that we might do. Thus if we seek a general procedure by which to balance our conflicting aims so as to single out, or at least to identify in thought, the best course of action, the idea of a dominant end seems to give a simple and natural answer.

Let us consider then what this dominant end might be. It cannot be happiness itself, since this state is attained by executing a rational plan of life already set out independently. The most we can say is that happiness is an inclusive end, meaning that the plan itself, the realization of which

18. The terminology of "dominant" and "inclusive" ends is from W. F. R. Hardie, "The Final Good in Aristotle's Ethics," *Philosophy*, vol. 40 (1965). This usage is not adhered to in his *Aristotle's Ethical Theory*.

makes one happy, includes and orders a plurality of aims, whatever these are. On the other hand, it is most implausible to think of the dominant end as a personal or social objective such as the exercise of political power, or the achievement of social acclaim, or maximizing one's material possessions. Surely it is contrary to our considered judgments of value, and indeed inhuman, to be so taken with but one of these ends that we do not moderate the pursuit of it for the sake of anything else. For a dominant end is at least lexically prior to all other aims and seeking to advance it always takes absolute precedence. Thus Loyola holds that the dominant end is serving God, and by this means saving our soul. He is consistent in recognizing that furthering the divine intentions is the sole criterion for balancing subordinate aims. It is for this reason alone that we should prefer health to sickness, riches to poverty, honor to dishonor, a long life to a short one, and, one might add, friendship and affection to hatred and animosity. We must be indifferent, he says, to all attachments whatsoever, for these become inordinate once they prevent us from being like equalized scales in a balance, ready to take the course that we believe is most for the glory of God.¹⁹

It should be observed that this principle of indifference is compatible with our enjoying lesser pleasures and allowing ourselves to engage in play and amusements. For these activities relax the mind and rest the spirit so that we are better fitted to advance more important aims. Thus although Aquinas believes that the vision of God is the last end of all human knowledge and endeavor, he concedes play and amusements a place in our life. Nevertheless these pleasures are permitted only to the extent that the superordinate aim is thereby advanced, or at least not hindered. We should arrange things so that our indulgences in frivolity and jest, in affection and friendship, do not interfere with the fullest attainment of our final end.²⁰

The extreme nature of dominant-end views is often concealed by the vagueness and ambiguity of the end proposed. Thus if God is conceived (as surely he must be) as a moral being, then the end of serving him above all else is left unspecified to the extent that the divine intentions are not clear from revelation, or evident from natural reason. Within these limits a theological doctrine of morals is subject to the same problems of balancing principles and determining precedence which trouble other con-

19. See *The Spiritual Exercises*, The First Week, the remarks under the heading "Principle and Foundation"; and The Second Week, the remarks under the heading "Three Occasions When a Wise Choice Can Be Made."

20. *Summa Contra Gentiles*, bk. III, ch. XXV.

ceptions. Since disputed questions commonly lie here, the solution propounded by the religious ethic is only apparent. And certainly when the dominant end is clearly specified as attaining some objective goal such as political power or material wealth, the underlying fanaticism and inhumanity are manifest. Human good is heterogeneous because the aims of the self are heterogeneous. Although to subordinate all our aims to one end does not strictly speaking violate the principles of rational choice (not the counting principles anyway), it still strikes us as irrational, or more likely as mad. The self is disfigured and put in the service of one of its ends for the sake of system.

84. HEDONISM AS A METHOD OF CHOICE

Traditionally hedonism is interpreted in one of two ways: either as the contention that the sole intrinsic good is pleasurable feeling, or as the psychological thesis that the only thing individuals strive for is pleasure. However I shall understand hedonism in a third way, namely, as trying to carry through the dominant-end conception of deliberation. It attempts to show how a rational choice is always possible, at least in principle. Although this effort fails, I shall examine it briefly for the light it throws upon the contrast between utilitarianism and the contract doctrine.

I imagine the hedonist to reason as follows. First he thinks that, if human life is to be guided by reason, there must exist a dominant end. There is no rational way to balance our competing aims against one another except as means to some higher end. Second, he interprets pleasure narrowly as agreeable feeling. Pleasantness as an attribute of feeling and sensation is thought to be the only plausible candidate for the role of the dominant end, and therefore it is the only thing good in itself. That, so conceived, pleasure alone is good is not postulated straightway as a first principle and then held to accord with our considered judgments of value. Rather pleasure is arrived at as the dominant end by a process of elimination. Granting that rational choices are possible, such an end must exist. At the same time this end cannot be happiness or any objective goal. To avoid the circularity of the one and the inhumanity and fanaticism of the other, the hedonist turns inwards. He finds the ultimate end in some definite quality of sensation or feeling identifiable by introspection. We can suppose, if we like, that pleasantness can be ostensibly defined as that attribute which is common to the feelings and experiences toward which we have a favorable attitude and wish to prolong, other things equal.

Thus, for purposes of illustration, one might say that pleasantness is that feature which is common to the experience of smelling roses, of tasting chocolate, of requited affection, and so on, and analogously for the opposite attribute of painfulness.²¹

The hedonist maintains, then, that a rational agent knows exactly how to proceed in determining his good: he is to ascertain which of the plans open to him promises the greatest net balance of pleasure over pain. This plan defines his rational choice, the best way to order his competing aims. The counting principles now apply trivially, since all good things are homogeneous and therefore comparable as means to the one end of pleasure. Of course these assessments are plagued by uncertainties and lack of information, and normally only the crudest estimates can be made. Yet for hedonism this is not a real difficulty: what counts is that the maximum of pleasure provides a clear idea of the good. We are now said to know the one thing the pursuit of which gives rational form to our life. Largely for these reasons Sidgwick thinks that pleasure must be the single rational end that is to guide deliberation.²²

It is important to note two points. First, when pleasure is regarded as a special attribute of feeling and sensation, it is conceived as a definite measure on which calculations can be based. By reckoning in terms of the intensity and duration of pleasant experiences, the necessary computations can theoretically be made. The method of hedonism provides a first-person procedure of choice as the standard of happiness does not. Second, taking pleasure as the dominant end does not imply that we have any particular objective goals. We find pleasure in the most varied activities and in the quest for any number of things. Therefore aiming to maximize pleasurable feeling seems at least to avoid the appearance of fanaticism and inhumanity while still defining a rational method for first-person choice. Furthermore, the two traditional interpretations of hedonism are now easily accounted for. If pleasure is indeed the only end the pursuit of which enables us to identify rational plans, then surely pleasure would appear to be the sole intrinsic good, and so we would have arrived at the principle of hedonism by an argument from the conditions of rational deliberation. A variant of psychological hedonism also follows: for although it is going too far to say that rational conduct would always consciously aim at pleasure, it would in any case be regulated by a schedule of activities designed to maximize the net balance of pleasurable

21. The illustration is from C. D. Broad, *Five Types of Ethical Theory* (London, Routledge and Kegan Paul, 1930), pp. 186f.

22. *The Methods of Ethics*, 7th ed. (London, Macmillan, 1907), pp. 405–407, 479.

feeling. Since it leads to the more familiar interpretations, the thesis that the pursuit of pleasure provides the only rational method of deliberation seems to be the fundamental idea of hedonism.

It seems obvious that hedonism fails to define a reasonable dominant end. We need only note that once pleasure is conceived, as it must be, in a sufficiently definite way so that its intensity and duration can enter into the agent's calculations, then it is no longer plausible that it should be taken as the sole rational aim.²³ Surely the preference for a certain attribute of feeling or sensation above all else is as unbalanced and inhuman as an overriding desire to maximize one's power over others or one's material wealth. No doubt it is for this reason that Sidgwick is reluctant to grant that pleasantness is a particular quality of feeling; yet he must concede this if pleasure is to serve, as he wants it to, as the ultimate criterion to weigh ideal values such as knowledge, beauty, and friendship against one another.²⁴

And then too there is the fact that there are different sorts of agreeable feelings themselves incomparable, as well as the quantitative dimensions of pleasure, intensity and duration. How are we to balance these when they conflict? Are we to choose a brief but intense pleasant experience of one kind of feeling over a less intense but longer pleasant experience of another? Aristotle says that the good man if necessary lays down his life for his friends, since he prefers a short period of intense pleasure to a long one of mild enjoyment, a twelvemonth of noble life to many years of humdrum existence.²⁵ But how does he decide this? Further, as Santayana observes, we must settle the relative worth of pleasure and pain. When Petrarch says that a thousand pleasures are not worth one pain, he adopts a standard for comparing them that is more basic than either. The person himself must make this decision, taking into account the full range of his inclinations and desires, present and future. Clearly we have made no advance beyond deliberative rationality. The problem of a plurality of ends arises all over again within the class of subjective feelings.²⁶

23. As Broad observes in *Five Types of Ethical Theory*, p. 187.

24. In *Methods of Ethics*, p. 127, Sidgwick denies that pleasure is a measurable quality of feeling independent of its relation from volition. This is the view of some writers, he says, but one he cannot accept. He defines pleasure "as a feeling which, when experienced by intelligent beings, is at least apprehended as desirable or—in cases of comparison—preferable." It would seem that the view he here rejects is the one he relies upon later as the final criterion to introduce coherence among ends. See pp. 405–407, 479. Otherwise the hedonist method of choice no longer provides instructions that can be followed.

25. *Nicomachean Ethics*, 1169a17–26.

26. *The Life of Reason in Common Sense* (New York, Charles Scribner, 1905), pp. 237f.

It may be objected that in economics and decision theory these problems are overcome. But this contention is based on a misunderstanding. In the theory of demand, for example, it is assumed that the consumer's preferences satisfy various postulates: they define a complete ordering over the set of alternatives and exhibit the properties of convexity and continuity, and the like. Given these assumptions, it can be shown that a utility function exists which matches these preferences in the sense that one alternative is chosen over another if and only if the value of the function for the selected alternative is greater. This function characterizes the individual's choices, what he in fact prefers, granted that his preferences meet certain stipulations. It asserts nothing at all about how a person arranges his decisions in such a coherent order to begin with, nor clearly can it claim to be a first-person procedure of choice that someone might reasonably follow, since it only records the outcome of his deliberations. At best the principles that economists have supposed the choices of rational individuals to satisfy can be presented as guidelines for us to consider when we make our decisions. But so understood, these criteria are just the principles of rational choice (or their analogues) and we are back once again with deliberative rationality.²⁷

It seems indisputable, then, that there is no dominant end the pursuit of which accords with our considered judgments of value. The inclusive end of realizing a rational plan of life is an entirely different thing. But the failure of hedonism to provide a rational procedure of choice should occasion no surprise. Wittgenstein showed that it is a mistake to postulate certain special experiences to explain how we distinguish memories from imaginings, beliefs from suppositions, and so on for other mental acts. Similarly, it is antecedently unlikely that certain kinds of agreeable feeling can define a unit of account the use of which explains the possibility of rational deliberation. Neither pleasure nor any other determinate end can play the role that the hedonist would assign it.²⁸

27. Thus to the objection that price theory must fail because it seeks to predict the unpredictable, the decisions of persons with free will, Walras says: "Actually, we have never attempted to predict decisions made under conditions of perfect freedom; we have only tried to express the effects of such decisions in terms of mathematics. In our theory each trader may be assumed to determine his utility or want curves as he pleases." *Elements of Pure Economics*, trans. William Jaffé (Homewood, Ill., Richard D. Irwin, 1954), p. 256. See also P. A. Samuelson, *Foundations of Economic Analysis* (Cambridge, Harvard University Press, 1947), the remarks pp. 90–92, 97f; and R. D. Luce and Howard Raiffa, *Games and Decisions* (New York, John Wiley and Sons, 1957), pp. 16, 21–24, 38.

28. See *The Philosophical Investigations* (Oxford, Basil Blackwell, 1953). The argument against postulating special experiences is made throughout for many different cases. For the application to pleasure, see the remarks of G. E. M. Anscombe, *Intention* (Oxford, Basil Blackwell, 1957). An-

Now philosophers have supposed that characteristic experiences exist and guide our mental life for many different reasons. So while it seems a simple matter to show that hedonism gets us nowhere, the important thing is to see why one might be driven to resort to such a desperate expedient. I have already noted one possible reason: the desire to narrow down the scope of purely preferential choice in determining our good. In a teleological theory any vagueness or ambiguity in the conception of the good is transferred to that of the right. Hence if the good of individuals is something that, so to speak, is just up to them to decide as individuals, so likewise within certain limits is that which is right. But it is natural to think that what is right is not a matter of mere preference, and therefore one tries to find a definite conception of the good.

There is, however, another reason: a teleological theory needs a way to compare the diverse goods of different individuals so that the total good can be maximized. How can these assessments be made? Even if certain ends serve to organize the plans of individuals taken singly, they do not suffice to define a conception of right. It would appear, then, that the turn inwards to the standard of agreeable feeling is an attempt to find a common denominator among the plurality of persons, an interpersonal currency as it were, by means of which the social ordering can be specified. And this suggestion is all the more compelling if it is already maintained that this standard is the aim of each person to the extent that he is rational.

By way of conclusion, I should not say that a teleological doctrine is necessarily driven to some form of hedonism in order to define a coherent theory. Yet it does seem that the tendency in this direction has a certain naturalness. Hedonism is, one might say, the symptomatic drift of teleological theories insofar as they try to formulate a clear and applicable method of moral reasoning. The weakness of hedonism reflects the impossibility of defining an appropriate definite end to be maximized. And this suggests that the structure of teleological doctrines is radically misconceived: from the start they relate the right and the good in the wrong

scombe says: "We might adapt a remark of Wittgenstein's about meaning and say 'Pleasure cannot be an impression; for no impression could have the consequences of pleasure.' They [the British Empiricists] were saying that something which they thought of as like a particular tickle or itch was quite obviously the point of doing anything whatsoever" (p. 77). See also Gilbert Ryle, "Pleasure," *Proceedings of the Aristotelian Society*, supp. vol. 28 (1954), and *Dilemmas* (Cambridge, The University Press, 1954), ch. IV; Anthony Kenny, *Action, Emotion and Will* (London, Routledge and Kegan Paul, 1963), ch. VI; and C. C. W. Taylor, "Pleasure," *Analysis*, supp. vol. (1963). These studies present what seems to be the more correct view. In the text I try to explain the motivation from the standpoint of moral philosophy of the so-called British Empiricist conception of pleasure. That it is fallacious I pretty much take for granted, as the writers mentioned have, I believe, shown.

way. We should not attempt to give form to our life by first looking to the good independently defined. It is not our aims that primarily reveal our nature but rather the principles that we would acknowledge to govern the background conditions under which these aims are to be formed and the manner in which they are to be pursued. For the self is prior to the ends which are affirmed by it; even a dominant end must be chosen from among numerous possibilities. There is no way to get beyond deliberative rationality. We should therefore reverse the relation between the right and the good proposed by teleological doctrines and view the right as prior. The moral theory is then developed by working in the opposite direction. I shall now try to explain these last remarks in the light of the contract doctrine.

85. THE UNITY OF THE SELF

The outcome of the preceding discussion is that there is no one aim by reference to which all of our choices can reasonably be made. Significant intuitionist elements enter into determining the good, and in a teleological theory these are bound to affect the right. The classical utilitarian tries to avoid this consequence by the doctrine of hedonism, but to no avail. We cannot, however, stop here; we must find a constructive solution to the problem of choice which hedonism seeks to answer. Thus we are faced once again with the question: if there is no single end that determines the appropriate pattern of aims, how is a rational plan of life actually to be identified? Now the answer to this question has already been given: a rational plan is one that would be chosen with deliberative rationality as defined by the full theory of the good. It remains to make sure that, within the context of a contract doctrine, this answer is perfectly satisfactory and that the problems which beset hedonism do not arise.

As I have said, moral personality is characterized by two capacities: one for a conception of the good, the other for a sense of justice. When realized, the first is expressed by a rational plan of life, the second by a regulative desire to act upon certain principles of right. Thus a moral person is a subject with ends he has chosen, and his fundamental preference is for conditions that enable him to frame a mode of life that expresses his nature as a free and equal rational being as fully as circumstances permit. Now the unity of the person is manifest in the coherence of his plan, this unity being founded on the higher-order desire to follow, in ways consistent with his sense of right and justice, the principles of

rational choice. Of course, a person shapes his aims not all at once but only gradually; but in ways that justice allows, he is able to formulate and to follow a plan of life and thereby to fashion his own unity.

The distinctive feature of a dominant-end conception is how it supposes the self's unity is achieved. Thus in hedonism the self becomes one by trying to maximize the sum of pleasurable experiences within its psychic boundaries. A rational self must establish its unity in this manner. Since pleasure is the dominant end, the individual is indifferent to all aspects of himself, viewing his natural assets of mind and body, and even his natural inclinations and attachments, as so many materials for obtaining pleasant experiences. Moreover, it is not by aiming at pleasure as his pleasure but simply as pleasure that gives unity to the self. Whether it is his pleasure or that of others as well which is to be advanced raises a further matter that can be put aside so long as we are dealing with one person's good. But once we consider the problem of social choice, the utilitarian principle in its hedonistic form is perfectly natural. For if any one individual must order his deliberations by seeking the dominant end of pleasure and can secure his rational personhood in no other way, then it seems that a number of persons in their joint efforts should strive to order their collective actions by maximizing the pleasurable experiences of the group. Thus just as one saint when alone is to work for the glory of God, so the members of an association of saints are to cooperate together to do whatever is necessary for the same end. The difference between the individual and the social case is that the resources of the self, its mental and physical capacities and its emotional sensibilities and desires, are placed in a different context. In both instances these materials are in the service of the dominant end. But depending on the other agencies available to cooperate with them, it is the pleasure of the self or of the social group that is to be maximized.

Further, if the same sorts of considerations that lead to hedonism as a theory of first-person choice are applied to the theory of right, the principle of utility seems quite plausible. For let us suppose first that happiness (defined in terms of agreeable feeling) is the sole good. Then, as even intuitionists concede, it is at least a *prima facie* principle of right to maximize happiness. If this principle is not alone regulative, there must be some other criterion such as distribution which is to be assigned some weight. But by reference to what dominant end of social conduct are these standards to be balanced? Since this end must exist if judgments of right are to be reasoned and not arbitrary, the principle of utility appears to specify the required goal. No other principle has the features necessary

to define the ultimate end of right conduct. I believe that it is essentially this reasoning that underlies Mill's so-called proof of utility.²⁹

Now in justice as fairness a complete reversal of perspective is brought about by the priority of right and the Kantian interpretation. To see this we have only to recall the features of the original position and the nature of the principles that are chosen. The parties regard moral personality and not the capacity for pleasure and pain as the fundamental aspect of the self. They do not know what final aims persons have, and all dominant-end conceptions are rejected. Thus it would not occur to them to acknowledge the principle of utility in its hedonistic form. There is no more reason for the parties to agree to this criterion than to maximize any other particular objective. They think of themselves as beings who can and do choose their final ends (always plural in number). Just as one person is to decide upon his plan of life in the light of full information (no restrictions being imposed in this case), so a plurality of persons are to settle the terms of their cooperation in a situation that gives all fair representation as moral beings. The parties' aim in the original position is to establish just and favorable conditions for each to fashion his own unity. Their fundamental interest in liberty and in the means to make fair use of it is the expression of their seeing themselves as primarily moral persons with an equal right to choose their mode of life. Thus they acknowledge the two principles of justice to be ranked in serial order as circumstances permit.

We must now connect these remarks with the problem of the indeterminacy of choice with which we began. The main idea is that given the priority of right, the choice of our conception of the good is framed within definite limits. The principles of justice and their realization in social forms define the bounds within which our deliberations take place. The essential unity of the self is already provided by the conception of right. Moreover, in a well-ordered society this unity is the same for all; everyone's conception of the good as given by his rational plan is a subplan of the larger comprehensive plan that regulates the community as a social union of social unions. The many associations of varying sizes

29. See *Utilitarianism*, ch. IV. This much-discussed chapter, and especially par. 3, is noteworthy for the fact that Mill seems to believe that if he can establish that happiness is the sole good, he has shown that the principle of utility is the criterion of right. The chapter title refers to the proof of the principle of utility; but what we are given is an argument to the effect that happiness alone is good. Now nothing so far follows about the conception of right. It is only by looking back at the first chapter of the essay, and attending to Mill's notion of the structure of a moral theory, as I discussed it in §8 and outlined it in the text above, that we can set out all the premises in the light of which Mill thought his argument a proof.

and aims, being adjusted to one another by the public conception of justice, simplify decision by offering definite ideals and forms of life that have been developed and tested by innumerable individuals, sometimes for generations. Thus in drawing up our plan of life we do not start *de novo*; we are not required to choose from countless possibilities without given structure or fixed contours. So while there is no algorithm for settling upon our good, no first-person procedure of choice, the priority of right and justice securely constrains these deliberations so that they become more manageable. Since the basic rights and liberties are already firmly established, our choices cannot distort our claims upon one another.

Now given the precedence of right and justice, the indeterminacy of the conception of the good is much less troublesome. In fact, the considerations that lead a teleological theory to embrace the notion of a dominant end lose their force. First of all, the purely preferential elements in choice, while not eliminated, are nevertheless confined within the constraints of right already on hand. Since men's claims on one another are not affected, the indeterminacy is relatively innocuous. Moreover, within the limits allowed by the principles of right, there need be no standard of correctness beyond that of deliberative rationality. If a person's plan of life meets this criterion and if he succeeds in carrying it out, and in doing so finds it worthwhile, there are no grounds for saying that it would have been better if he had done something else. We should not simply assume that our rational good is uniquely determined. From the standpoint of the theory of justice, this assumption is unnecessary. Secondly, we are not required to go beyond deliberative rationality in order to define a clear and workable conception of right. The principles of justice have a definite content and the argument supporting them uses only the thin account of the good and its list of primary goods. Once the conception of justice is established, the priority of right guarantees the precedence of its principles. Thus the two considerations that make dominant-end conceptions attractive for teleological theories are both absent in the contract doctrine. Such is the effect of the reversal of structure.

Earlier when introducing the Kantian interpretation of justice as fairness, I mentioned that there is a sense in which the unanimity condition on the principles of justice is suited to express even the nature of a single self (§40). Offhand this suggestion seems paradoxical. How can the requirement of unanimity fail to be a constraint? One reason is that the veil of ignorance insures that everyone should reason in the same way and so the condition is satisfied as a matter of course. But a deeper explanation

lies in the fact that the contract doctrine has a structure opposite to that of a utilitarian theory. In the latter each person draws up his rational plan without hindrance under full information, and society then proceeds to maximize the aggregate fulfillment of the plans that result. In justice as fairness, on the other hand, all agree ahead of time upon the principles by which their claims on one another are to be settled. These principles are then given absolute precedence so that they regulate social institutions without question and each frames his plans in conformity with them. Plans that happen to be out of line must be revised. Thus the prior collective agreement sets up from the first certain fundamental structural features common to everyone's plan. The nature of the self as a free and equal moral person is the same for all, and the similarity in the basic form of rational plans expresses this fact. Moreover, as shown by the notion of society as a social union of social unions, the members of a community participate in one another's nature: we appreciate what others do as things we might have done but which they do for us, and what we do is similarly done for them. Since the self is realized in the activities of many selves, relations of justice that conform to principles which would be assented to by all are best fitted to express the nature of each. Eventually then the requirement of a unanimous agreement connects up with the idea of human beings who as members of a social union seek the values of community.

It may be thought that once the principles of justice are given precedence, then there is a dominant end that organizes our life after all. Yet this idea is based on a misunderstanding. To be sure the principles of justice are lexically prior to that of efficiency, and the first principle has precedence over the second. It follows that an ideal conception of the social order is set up which is to regulate the direction of change and the efforts of reform (§41). But it is the principles of individual duty and obligation that define the claim of this ideal upon persons and these do not make it all controlling. Furthermore, I have all along assumed that the proposed dominant end belongs to a teleological theory in which by definition the good is specified independently from the right. The role of this end is in part to make the conception of right reasonably precise. In justice as fairness there can be no dominant end in this sense, nor as we have seen is one needed for this purpose. Finally, the dominant end of a teleological theory is so defined that we can never finally achieve it and therefore the injunction to advance it always applies. Recall here the earlier remarks as to why the principle of utility is not really suitable for a lexical ordering: the later criteria will never come into play, except in special cases to break ties. The principles of justice, on the other hand,

represent more or less definite social aims and restrictions (§8). Once we realize a certain structure of institutions, we are at liberty to determine and to pursue our good within the limits which its arrangements allow.

In view of these reflections, the contrast between a teleological theory and the contract doctrine may be expressed in the following intuitive way: the former defines the good locally, for example, as a more or less homogeneous quality or attribute of experience, and regards it as an extensive magnitude which is to be maximized over some totality; whereas the latter moves in the opposite fashion by identifying a sequence of increasingly specific structural forms of right conduct each set within the preceding one, and in this manner working from a general framework for the whole to a sharper and sharper determination of its parts. Hedonistic utilitarianism is the classical instance of the first procedure and illustrates it with compelling simplicity. Justice as fairness exemplifies the second possibility. Thus the four-stage sequence (§31) formulates an order of agreements and enactments designed to build up in several steps a hierarchical structure of principles, standards, and rules, which when consistently applied and adhered to, lead to a definite constitution for social action.

Now this sequence does not aim at the complete specification of conduct. Rather the idea is to approximate the boundaries, however vague, within which individuals and associations are at liberty to advance their aims and deliberative rationality has free play. Ideally the approximation should converge in the sense that with further steps the cases left unaccounted for become of less and less importance. The notion guiding the entire construction is that of the original position and its Kantian interpretation: this notion contains within itself the elements that select which information is relevant at each stage, and generate a sequence of adjustments appropriate to the contingent conditions of the existing society.

86. THE GOOD OF THE SENSE OF JUSTICE

Now that all the parts of the theory of justice are before us, the argument for congruence can be completed. It suffices to tie together the various aspects of a well-ordered society and to see them in the appropriate context. The concepts of justice and goodness are linked with distinct principles and the question of congruence is whether these two families of criteria fit together. More precisely, each concept with its associated principles defines a point of view from which institutions, actions, and

plans of life can be assessed. A sense of justice is an effective desire to apply and to act from the principles of justice and so from the point of view of justice. Thus what is to be established is that it is rational (as defined by the thin theory of the good) for those in a well-ordered society to affirm their sense of justice as regulative of their plan of life. It remains to be shown that this disposition to take up and to be guided by the standpoint of justice accords with the individual's good.

Whether these two points of view are congruent is likely to be a crucial factor in determining stability. But congruence is not a foregone conclusion even in a well-ordered society. We must verify it. Of course, the rationality of choosing the principles of justice in the original position is not in question. The argument for this decision has already been made; and if it is sound, just institutions are collectively rational and to everyone's advantage from a suitably general perspective. It is also rational for each to urge others to support these arrangements and to fulfill their duties and obligations. The problem is whether the regulative desire to adopt the standpoint of justice belongs to a person's own good when viewed in the light of the thin theory with no restrictions on information. We should like to know that this desire is indeed rational; being rational for one, it is rational for all, and therefore no tendencies to instability exist. More precisely, consider any given person in a well-ordered society. He knows, I assume, that institutions are just and that others have (and will continue to have) a sense of justice similar to his, and therefore that they comply (and will continue to comply) with these arrangements. We want to show that on these suppositions it is rational for someone, as defined by the thin theory, to affirm his sense of justice. The plan of life which does this is his best reply to the similar plans of his associates; and being rational for anyone, it is rational for all.

It is important not to confuse this problem with that of justifying being a just man to an egoist. An egoist is someone committed to the point of view of his own interests. His final ends are related to himself: his wealth and position, his pleasures and social prestige, and so on. Such a man may act justly, that is, do things that a just man would do; but so long as he remains an egoist, he cannot do them for the just man's reasons. Having these reasons is inconsistent with being an egoist. It merely happens that on some occasions the point of view of justice and that of his own interests lead to the same course of action. Therefore I am not trying to show that in a well-ordered society an egoist would act from a sense of justice, nor even that he would act justly because so acting would best advance his ends. Nor, again, are we to argue that an egoist, finding

himself in a just society, would be well advised, given his aims, to transform himself into a just man. Rather, we are concerned with the goodness of the settled desire to take up the standpoint of justice. I assume that the members of a well-ordered society already have this desire. The question is whether this regulative sentiment is consistent with their good. We are not examining the justice or the moral worth of actions from certain points of view; we are assessing the goodness of the desire to adopt a particular point of view, that of justice itself. And we must evaluate this desire not from the egoist's standpoint, whatever this might be, but in the light of the thin theory of the good.

I shall assume that human actions spring from existing desires and that these can be changed only gradually. We cannot just decide at a given moment to alter our system of ends (§63). We act now as the sort of person we are and from the wants we have now, and not as the sort of person we might have been or from desires we would have had if earlier we had only chosen differently. Regulative aims are especially subject to this constraint. Thus we must decide well in advance whether to affirm our sense of justice by trying to assess our situation over a fairly extensive future. We cannot have things both ways. We cannot preserve a sense of justice and all that this implies while at the same time holding ourselves ready to act unjustly should doing so promise some personal advantage. A just person is not prepared to do certain things, and if he is tempted too easily, he was prepared after all.³⁰ Our question concerns then only those with a certain psychology and system of desires. It would obviously be demanding too much to require that stability should not depend upon definite restrictions in this respect.

Now on one interpretation the question has an obvious answer. Supposing that someone has an effective sense of justice, he will then have a regulative desire to comply with the corresponding principles. The criteria of rational choice must take this desire into account. If a person wants with deliberative rationality to act from the standpoint of justice above all else, it is rational for him so to act. Therefore in this form the question is trivial: being the sorts of persons they are, the members of a well-ordered society desire more than anything to act justly and fulfilling this desire is part of their good. Once we acquire a sense of justice that is truly final and effective, as the precedence of justice requires, we are confirmed in a plan of life that, insofar as we are rational, leads us to preserve and to

30. See Philippa Foot, "Moral Beliefs," *Proceedings of the Aristotelian Society*, vol. 59 (1958–1959), p. 104. I am much indebted to this essay, although I have not followed it on all counts.

encourage this sentiment. Since this fact is public knowledge, instability of the first kind does not exist, and hence neither does that of the second. The real problem of congruence is what happens if we imagine someone to give weight to his sense of justice only to the extent that it satisfies other descriptions which connect it with reasons specified by the thin theory of the good. We should not rely on the doctrine of the pure conscientious act (§72). Suppose, then, that the desire to act justly is not a final desire like that to avoid pain, misery, or apathy, or the desire to fulfill the inclusive interest. The theory of justice supplies other descriptions of what the sense of justice is a desire for; and we must use these to show that a person following the thin theory of the good would indeed confirm this sentiment as regulative of his plan of life.

So much then for defining the question. I now wish to note the grounds of congruence by reviewing various points already made. First of all, as the contract doctrine requires, the principles of justice are public: they characterize the commonly recognized moral convictions shared by the members of a well-ordered society (§23). We are not concerned with someone who is questioning these principles. By hypothesis, he concedes as everyone else does that they are the best choice from the standpoint of the original position. (Of course, this can always be doubted but it raises an entirely different matter.) Now since others are assumed to have (and to continue to have) an effective sense of justice, our hypothetical individual is considering in effect a policy of pretending to have certain moral sentiments, all the while being ready to act as a free-rider whenever the opportunity arises to further his personal interests. Since the conception of justice is public, he is debating whether to set out on a systematic course of deception and hypocrisy, professing without belief, as it suits his purpose, the accepted moral views. That deception and hypocrisy are wrongs does not, I assume, bother him; but he will have to reckon with the psychological cost of taking precautions and maintaining his pose, and with the loss of spontaneity and naturalness that results.³¹ In most societies as things are, such pretensions may not have a high price, since the injustice of institutions and the often squalid behavior of others renders one's own deceits easier to endure; but in a well-ordered society there is not this comfort.

These remarks are supported by the fact that there is a connection between acting justly and natural attitudes (§74). Given the content of the principles of justice and the laws of moral psychology, wanting to be fair

31. See Foot, *ibid.*, p. 104.

with our friends and wanting to give justice to those we care for is as much a part of these affections as the desire to be with them and to feel sad at their loss. Assuming therefore that one needs these attachments, the policy contemplated is presumably that of acting justly only toward those to whom we are bound by ties of affection and fellow feeling, and of respecting ways of life to which we are devoted. But in a well-ordered society these bonds extend rather widely, and include ties to institutional forms, assuming here that all three psychological laws are fully effective. In addition, we cannot in general select who is to be injured by our unfairness. For example, if we cheat on paying our taxes, or if we find some way to avoid doing our fair share for the community, everyone is hurt, our friends and associates along with the rest. To be sure, we might consider covertly passing on part of our gains to those we especially like, but this becomes a dubious and involved affair. Thus in a well-ordered society where effective bonds are extensive both to persons and to social forms, and we cannot select who is to lose by our defections, there are strong grounds for preserving one's sense of justice. Doing this protects in a natural and simple way the institutions and persons we care for and leads us to welcome new and broader social ties.

Another basic consideration is this: it follows from the Aristotelian Principle (and its companion effect) that participating in the life of a well-ordered society is a great good (§79). This conclusion depends upon the meaning of the principles of justice and their precedence in everyone's plans as well as upon the psychological features of our nature. It is the details of the contract view which establish this connection. Because such a society is a social union of social unions, it realizes to a preeminent degree the various forms of human activity; and given the social nature of humankind, the fact that our potentialities and inclinations far surpass what can be expressed in any one life, we depend upon the cooperative endeavors of others not only for the means of well-being but to bring to fruition our latent powers. And with a certain success all around, each enjoys the greater richness and diversity of the collective activity. Yet to share fully in this life we must acknowledge the principles of its regulative conception, and this means that we must affirm our sentiment of justice. To appreciate something as ours, we must have a certain allegiance to it. What binds a society's efforts into one social union is the mutual recognition and acceptance of the principles of justice; it is this general affirmation which extends the ties of identification over the whole community and permits the Aristotelian Principle to have its wider effect. Individual and group accomplishments are no longer seen

as just so many separate personal goods. Whereas not to confirm our sense of justice is to limit ourselves to a narrow view.

Finally, there is the reason connected with the Kantian interpretation: acting justly is something we want to do as free and equal rational beings (§40). The desire to act justly and the desire to express our nature as free moral persons turn out to specify what is practically speaking the same desire. When someone has true beliefs and a correct understanding of the theory of justice, these two desires move him in the same way. They are both dispositions to act from precisely the same principles: namely, those that would be chosen in the original position. Of course, this contention is based on a theory of justice. If this theory is unsound, the practical identity fails. But since we are concerned only with the special case of a well-ordered society as characterized by the theory, we are entitled to assume that its members have a lucid grasp of the public conception of justice upon which their relations are founded.

Let us suppose that these are the chief reasons (or typical thereof) which the thin account of the good allows for maintaining one's sense of justice. The question now arises whether they are decisive. Here we confront the familiar difficulty of a balance of motives which in many ways is similar to a balance of first principles. Sometimes the answer is found by comparing one balance of reasons with another, for surely if the first balance clearly favors one course of action then the second will also, should its reasons supporting this alternative be stronger and its reasons supporting the other alternatives be weaker. But arguing from such comparisons presupposes some configurations of reasons which evidently go one way rather than another to serve as a bench mark. Failing these, we cannot get beyond conditional comparisons: if the first balance favors a certain choice, then the second does also.

Now at this point it is obvious that the content of the principles of justice is a crucial element in the decision. Whether it is for a person's good that he have a regulative sense of justice depends upon what justice requires of him. The congruence of the right and the good is determined by the standards by which each concept is specified. As Sidgwick notes, utilitarianism is more strict than common sense in demanding the sacrifice of the agent's private interests when this is necessary for the greater happiness of all.³² It is also more exacting than the contract theory, for while beneficent acts going beyond our natural duties are good actions and evoke our esteem, they are not required as a matter of right. Utilitari-

32. *Methods of Ethics*, pp. 246–253, 499.

anism may seem to be a more exalted ideal, but the other side of it is that it may authorize the lesser welfare and liberty of some for the sake of a greater happiness of others who may already be more fortunate. A rational person, in framing his plan, would hesitate to give precedence to so stringent a principle. It is likely both to exceed his capacity for sympathy and to be hazardous to his freedom. Thus however improbable the congruence of the right and the good in justice as fairness, it is surely more probable than on the utilitarian view. The conditional balance of reasons favors the contract doctrine.

A somewhat different point is suggested by the following doubt: namely, that while the decision to preserve our sentiment of justice might be rational, we may in the end suffer a very great loss or even be ruined by it. As we have seen, a just person is not prepared to do certain things, and so in the face of evil circumstances he may decide to chance death rather than to act unjustly. Yet although it is true enough that for the sake of justice a man may lose his life where another would live to a later day, the just man does what all things considered he most wants; in this sense he is not defeated by ill fortune the possibility of which he foresaw. The question is on a par with the hazards of love; indeed, it is simply a special case. Those who love one another, or who acquire strong attachments to persons and to forms of life, at the same time become liable to ruin: their love makes them hostages to misfortune and the injustice of others. Friends and lovers take great chances to help each other; and members of families willingly do the same. Their being so disposed belongs to their attachments as much as any other inclination. Once we love we are vulnerable: there is no such thing as loving while being ready to consider whether to love, just like that. And the loves that may hurt the least are not the best loves. When we love we accept the dangers of injury and loss. In view of our general knowledge of the likely course of life, we do not think these risks so great as to cause us to cease loving. Should evils occur, they are the object of our aversion, and we resist those whose machinations bring them about. If we are loving we do not regret our love. Now if these things are true of love as the world is, or very often is, then a fortiori they would appear to be true of loves in a well-ordered society, and so of the sense of justice too. For in a society where others are just our loves expose us mainly to the accidents of nature and the contingency of circumstances. And similarly for the sentiment of justice which is connected to these affections. Taking as a bench mark the balance of reasons that leads us to affirm our loves as things are, it seems

that we should be ready once we come of age to maintain our sense of justice in the more favorable conditions of a just society.

One special feature of the desire to express our nature as moral persons strengthens this conclusion. With other inclinations of the self, there is a choice of degree and scope. Our policy of deception and hypocrisy need not be completely systematic; our affective ties to institutions and to other persons can be more or less strong, and our participation in the wider life of society more or less full. There is a continuum of possibilities and not an all or nothing decision, although for simplicity I have spoken pretty much in these terms. But the desire to express our nature as a free and equal rational being can be fulfilled only by acting on the principles of right and justice as having first priority. This is a consequence of the condition of finality: since these principles are regulative, the desire to act upon them is satisfied only to the extent that it is likewise regulative with respect to other desires. It is acting from this precedence that expresses our freedom from contingency and happenstance. Therefore in order to realize our nature we have no alternative but to plan to preserve our sense of justice as governing our other aims. This sentiment cannot be fulfilled if it is compromised and balanced against other ends as but one desire among the rest. It is a desire to conduct oneself in a certain way above all else, a striving that contains within itself its own priority. Other aims can be achieved by a plan that allows a place for each, since their satisfaction is possible independent of their place in the ordering. But this is not the case with the sense of right and justice; and therefore acting wrongly is always liable to arouse feelings of guilt and shame, the emotions aroused by the defeat of our regulative moral sentiments. Of course, this does not mean that the realization of our nature as a free and rational being is itself an all or nothing affair. To the contrary, how far we succeed in expressing our nature depends upon how consistently we act from our sense of justice as finally regulative. What we cannot do is express our nature by following a plan that views the sense of justice as but one desire to be weighed against others. For this sentiment reveals what the person is, and to compromise it is not to achieve for the self free reign but to give way to the contingencies and accidents of the world.

One last question must be mentioned. Suppose that even in a well-ordered society there are some persons for whom the affirmation of their sense of justice is not a good. Given their aims and wants and the peculiarities of their nature, the thin account of the good does not define reasons sufficient for them to maintain this regulative sentiment. It has been

argued that to these persons one cannot truthfully recommend justice as a virtue.³³ And this is surely correct, assuming such a recommendation to imply that rational grounds (identified by the thin theory) counsel this course for them as individuals. But then the further question remains whether those who do affirm their sense of justice are treating these persons unjustly in requiring them to comply with just institutions.

Now unhappily we are not yet in a position to answer this query properly, since it presupposes a theory of punishment and I have said very little about this part of the theory of justice (§39). I have assumed strict compliance with any conception that would be chosen and then considered which one on the list presented would be adopted. However, we may reason much as we did in the case of civil disobedience, another part of partial compliance theory. Thus granting that adherence to whatever conception is acknowledged will be imperfect if left completely voluntary, under what conditions would the persons in the original position agree that stabilizing penal devices can be employed? Would they insist that a person can be required to do only what is to his advantage as defined by the thin theory?

It seems clear, in the light of the contract doctrine as a whole, that they would not. For this restriction amounts in effect to general egoism which, as we have seen, would be rejected. Moreover, the principles of right and justice are collectively rational; and it is in the interest of each that everyone else should comply with just arrangements. It is also the case that the general affirmation of the sense of justice is a great social asset, establishing the basis for mutual trust and confidence from which all normally benefit. Thus in agreeing to penalties that stabilize a scheme of cooperation the parties accept the same kind of constraint on self-interest that they acknowledge in choosing the principles of justice in the first place. Having agreed to these principles in view of the reasons already surveyed, it is rational to authorize the measures needed to maintain just institutions, assuming that the constraints of equal liberty and the rule of law are duly recognized (§§38–39). Those who find that being disposed to act justly is not a good for them cannot deny these contentions. It is, of course, true that in their case just arrangements do not fully answer to their nature, and therefore, other things equal, they will be less happy than they would be if they could affirm their sense of justice. But here one can only say: their nature is their misfortune.

The main point then is that to justify a conception of justice we do not

33. See Foot, pp. 99–104.

have to contend that everyone, whatever his capacities and desires, has a sufficient reason (as defined by the thin theory) to preserve his sense of justice. For our good depends upon the sorts of persons we are, the kinds of wants and aspirations we have and are capable of. It can even happen that there are many who do not find a sense of justice for their good; but if so, the forces making for stability are weaker. Under such conditions penal devices will play a much larger role in the social system. The greater the lack of congruence, the greater the likelihood, other things equal, of instability with its attendant evils. Yet none of this nullifies the collective rationality of the principles of justice; it is still to the advantage of each that everyone else should honor them. At least this holds true so long as the conception of justice is not so unstable that some other conception would be preferable. But what I have tried to show is that the contract doctrine is superior to its rivals on this score, and therefore that the choice of principles in the original position need not be reconsidered. In fact, granted a reasonable interpretation of human sociability (provided by the account of how a sense of justice is acquired and by the idea of social union), justice as fairness appears to be a sufficiently stable conception. The hazards of the generalized prisoner's dilemma are removed by the match between the right and the good. Of course, under normal conditions public knowledge and confidence are always imperfect. So even in a just society it is reasonable to admit certain constraining arrangements to insure compliance, but their main purpose is to underwrite citizens' trust in one another. These mechanisms will seldom be invoked and will comprise but a minor part of the social scheme.

We are now at the end of this rather lengthy discussion of the stability of justice as fairness. The only further point to note is that congruence allows us to complete the sequence of applications of the definition of goodness. We can say first that, in a well-ordered society, being a good person (and in particular having an effective sense of justice) is indeed a good for that person; and second that this form of society is a good society. The first assertion follows from congruence; the second holds since a well-ordered society has the properties that it is rational to want in a society from the two relevant points of view. Thus a well-ordered society satisfies the principles of justice which are collectively rational from the perspective of the original position; and from the standpoint of the individual, the desire to affirm the public conception of justice as regulative of one's plan of life accords with the principles of rational choice. These conclusions support the values of community, and in reaching them my account of justice as fairness is completed.

87. CONCLUDING REMARKS ON JUSTIFICATION

I shall not try to summarize the presentation of the theory of justice. Instead I should like to end with a few comments about the kind of argument I have offered for it. Now that the whole conception is before us, we are in a position to note in a general way the sorts of things that can be said on its behalf. Doing this will clarify several points which may still be in doubt.

Philosophers commonly try to justify ethical theories in one of two ways. Sometimes they attempt to find self-evident principles from which a sufficient body of standards and precepts can be derived to account for our considered judgments. A justification of this kind we may think of as Cartesian. It presumes that first principles can be seen to be true, even necessarily so; deductive reasoning then transfers this conviction from premises to conclusion. A second approach (called naturalism by an abuse of language) is to introduce definitions of moral concepts in terms of presumptively non-moral ones, and then to show by accepted procedures of common sense and the sciences that the statements thus paired with the asserted moral judgments are true. Although on this view the first principles of ethics are not self-evident, the justification of moral convictions poses no special difficulties. They can be established, granting the definitions, in the same fashion as other statements about the world.

I have not adopted either of these conceptions of justification. For while some moral principles may seem natural and even obvious, there are great obstacles to maintaining that they are necessarily true, or even to explaining what is meant by this. Indeed, I have held that these principles are contingent in the sense that they are chosen in the original position in the light of general facts (§26). More likely candidates for necessary moral truths are the conditions imposed on the adoption of principles; but actually it seems best to regard these conditions simply as reasonable stipulations to be assessed eventually by the whole theory to which they belong. There is no set of conditions or first principles that can be plausibly claimed to be necessary or definitive of morality and thereby especially suited to carry the burden of justification. On the other hand, the method of naturalism so-called must first distinguish moral from non-moral concepts and then gain acceptance for the definitions laid down. For the justification to succeed, a clear theory of meaning is presupposed and this seems to be lacking. And in any case, definitions become the main part of the ethical doctrine, and thus in turn they need to be justified.

Therefore we do better, I think, to regard a moral theory just as any

other theory, making due allowances for its Socratic aspects (§9). There is no reason to suppose that its first principles or assumptions need to be self-evident, or that its concepts and criteria can be replaced by other notions which can be certified as non-moral.³⁴ Thus while I have maintained, for example, that something's being right, or just, can be understood as its being in accordance with the relevant principles that would be acknowledged in the original position, and that we can in this way replace the former notions by the latter, these definitions are set up within the theory itself (§18). I do not hold that the conception of the original position is itself without moral force, or that the family of concepts it draws upon is ethically neutral (§23). This question I simply leave aside. I have not proceeded then as if first principles, or conditions thereon, or definitions either, have special features that permit them a peculiar place in justifying a moral doctrine. They are central elements and devices of theory, but justification rests upon the entire conception and how it fits in with and organizes our considered judgments in reflective equilibrium. As we have noted before, justification is a matter of the mutual support of many considerations, of everything fitting together into one coherent view (§4). Accepting this idea allows us to leave questions of meaning and definition aside and to get on with the task of developing a substantive theory of justice.

The three parts of the exposition of this theory are intended to make a unified whole by supporting one another in roughly the following way. The first part presents the essentials of the theoretical structure, and the principles of justice are argued for on the basis of reasonable stipulations concerning the choice of such conceptions. I urged the naturalness of these conditions and presented reasons why they are accepted, but it was not claimed that they are self-evident, or required by the analysis of moral concepts or the meaning of ethical terms. In the second part I examined the sorts of institutions that justice enjoins and the kinds of duties and obligations it imposes on individuals. The aim throughout was to show that the theory proposed matches the fixed points of our considered convictions better than other familiar doctrines, and that it leads us to revise and extrapolate our judgments in what seem on reflection to be more satisfactory ways. First principles and particular judgments appear on

34. The view proposed here accords with the account in §9 which follows "Outline for Ethics" (1951). But it has benefited from the conception of justification found in W. V. Quine, *Word and Object* (Cambridge, M.I.T. Press, 1960), ch. 1 and elsewhere. See also his *Ontological Relativity and Other Essays* (New York, Columbia University Press, 1969), Essay 4. For a development of this conception to include explicitly moral thought and judgment, see Morton White, *Toward Reunion in Philosophy* (Cambridge, Harvard University Press, 1956), pt. III, esp. pp. 254–258, 263, 266f.

balance to hang together reasonably well, at least in comparison with alternative theories. Finally we checked to see in the third part if justice as fairness is a feasible conception. This forced us to raise the question of stability and whether the right and the good as defined are congruent. These considerations do not determine the initial acknowledgment of principles in the first part of the argument, but confirm it (§81). They show that our nature is such as to allow the original choice to be carried through. In this sense we might say that humankind has a moral nature.

Now some may hold that this kind of justification faces two sorts of difficulties. First, it is open to the general complaint that it appeals to the mere fact of agreement. Second, there is the more specific objection to the argument I have presented that it depends upon a particular list of conceptions of justice between which the parties in the original position are to choose, and it assumes not only an agreement among persons in their considered judgments, but also in what they regard as reasonable conditions to impose on the choice of first principles. It may be said that the agreement in considered convictions is constantly changing and varies between one society, or part thereof, and another. Some of the so-called fixed points may not really be fixed, nor will everyone accept the same principles for filling in the gaps in their existing judgments. And any list of conceptions of justice, or consensus about what counts as reasonable conditions on principles, is surely more or less arbitrary. The case presented for justice as fairness, so the contention runs, does not escape these limitations.

In regard to the general objection the reply is that justification is argument addressed to those who disagree with us, or to ourselves when we are of two minds. It presumes a clash of views between persons or within one person, and seeks to convince others, or ourselves, of the reasonableness of the principles upon which our claims and judgments are founded. Being designed to reconcile by reason, justification proceeds from what all parties to the discussion hold in common. Ideally, to justify a conception of justice to someone is to give him a proof of its principles from premises that we both accept, these principles having in turn consequences that match our considered judgments. Thus mere proof is not justification. A proof simply displays logical relations between propositions. But proofs become justification once the starting points are mutually recognized, or the conclusions so comprehensive and compelling as to persuade us of the soundness of the conception expressed by their premises.

It is perfectly proper, then, that the argument for the principles of jus-

tice should proceed from some consensus. This is the nature of justification. Yet the more specific objections are correct in implying that the force of the argument depends on the features of the consensus appealed to. Here several points deserve notice. To begin with, while it should be granted that any list of alternatives may be to some extent arbitrary, the objection is mistaken if it is read as holding that all lists are equally so. A list that includes the leading traditional theories is less arbitrary than one which leaves out the more obvious candidates. Certainly the argument for the principles of justice would be strengthened by showing that they are still the best choice from a more comprehensive list more systematically evaluated. I do not know how far this can be done. I doubt, however, that the principles of justice (as I have defined them) will be the preferred conception on anything resembling a complete list. (Here I assume that, given an upper bound on complexity and other constraints, the class of reasonable and practicable alternatives is effectively finite.) Even if the argument I have offered is sound, it only shows that a finally adequate theory (if such exists) will look more like the contract view than any of the other doctrines we discussed. And even this conclusion is not proved in any strict sense.

Nevertheless, in comparing justice as fairness with these conceptions, the list used is not simply *ad hoc*: it includes representative theories from the tradition of moral philosophy which comprises the historical consensus about what so far seem to be the more reasonable and practicable moral conceptions. With time further possibilities will be worked out, thereby providing a more convincing basis for justification as the leading conception is subjected to a more severe test. But these things we can only anticipate. For the present it is appropriate to try to reformulate the contract doctrine and to compare it with a few familiar alternatives. This procedure is not arbitrary; we can advance in no other way.

Turning to the particular difficulty about the consensus on reasonable conditions, one should point out that one of the aims of moral philosophy is to look for possible bases of agreement where none seem to exist. It must attempt to extend the range of some existing consensus and to frame more discriminating moral conceptions for our consideration. Justifying grounds do not lie ready to hand: they need to be discovered and suitably expressed, sometimes by lucky guesses, sometimes by noting the requirements of theory. It is with this aim in mind that the various conditions on the choice of first principles are brought together in the notion of the original position. The idea is that by putting together enough reasonable constraints into a single conception, it will become obvious that one

among the alternatives presented is to be preferred. We should like it to happen that the superiority of a particular view (among those currently known) is the result, perhaps the unexpected result, of this newly observed consensus.

Again, the set of conditions incorporated into the notion of the original position is not without an explanation. It is possible to maintain that these requirements are reasonable and to connect them with the purpose of moral principles and their role in establishing the ties of community. The grounds for ordering and finality, say, seem clear enough. And we can now see that publicity can be explained as insuring that the process of justification can be perfectly carried through (in the limit so to speak) without untoward effects. For publicity allows that all can justify their conduct to everyone else (when their conduct is justifiable) without self-defeating or other disturbing consequences. If we take seriously the idea of a social union and of society as a social union of such unions, then surely publicity is a natural condition. It helps to establish that a well-ordered society is one activity in the sense that its members follow and know of one another, that they follow the same regulative conception; and everyone shares in the benefits of the endeavors of all in ways to which each is known to consent. Society is not partitioned with respect to the mutual recognition of its first principles. And, indeed, this must be so if the binding action of the conception of justice and of the Aristotelian principle (and its companion effect) are to take place.

To be sure, the function of moral principles is not uniquely defined; it admits of various interpretations. We might try to choose between them by seeing which one uses the weakest set of conditions to characterize the initial situation. The difficulty with this suggestion is that while weaker conditions are indeed to be preferred, other things equal, there is no weakest set; a minimum does not exist short of no conditions at all and this is of no interest. Therefore we must look for a constrained minimum, a set of weak conditions that still enables us to construct a workable theory of justice. Certain parts of justice as fairness should be viewed in this way. I have several times noted the minimal nature of the conditions on principles when taken singly. For example, the assumption of mutually disinterested motivation is not a demanding stipulation. Not only does it enable us to base the theory upon a reasonably precise notion of rational choice, but it asks little of the parties: in this way the principles chosen can adjust wider and deeper conflicts, an obvious desideratum (§40). It has the further advantage of separating off the more evident moral elements of the original position in the form of general conditions and the

veil of ignorance and the like, so that we can see more clearly how justice requires us to go beyond a concern for our own interests.

The discussion of freedom of conscience illustrates most clearly the assumption of mutual disinterest. Here the opposition of the parties is very great, yet one can still show that if any agreement is possible, it is that on the principle of equal liberty. And, as we noted, this idea can be extended to conflicts between moral doctrines as well (§33). If the parties assume that in society they affirm some moral conception (the content of which is unknown to them), they can still assent to the first principle. This principle therefore appears to hold a special place among moral views; it defines an agreement in the limit once we postulate sufficiently wide disparities consistent with certain minimal conditions for a practical conception of justice.

I should now like to take note of several objections that are independent from the method of justification and concern instead certain features of the theory of justice itself. One of these is the criticism that the contract view is a narrowly individualistic doctrine. To this difficulty, the preceding remarks supply the answer. For once the point of the assumption of mutual disinterest is understood, the objection seems misplaced. Within the framework of justice as fairness we can reformulate and establish Kantian themes by using a suitably general conception of rational choice. For example, we have found interpretations of autonomy and of the moral law as an expression of our nature as free and equal rational beings; the categorical imperative also has its analogue, as does the idea of never treating persons as means only, or indeed as means at all. Further, in the last part the theory of justice has been shown to account for the values of community as well; and this strengthens the earlier contention that embedded in the principles of justice there is an ideal of the person that provides an Archimedean point for judging the basic structure of society (§41). These aspects of the theory of justice are developed slowly beginning from what looks like an unduly rationalistic conception that makes no provision for social values. The original position is first used to determine the content of justice, the principles which define it. Not until later is justice seen as part of our good and connected with our natural sociability. The merits of the idea of the original position cannot be assessed by focusing on some single feature of it, but, as I have often observed, only by the whole theory which is built upon it.

If justice as fairness is more convincing than the older presentations of the contract doctrine, I believe that it is because the original position, as indicated above, unites in one conception a reasonably clear problem of

choice with conditions that are widely recognized as fitting to impose on the adoption of moral principles. This initial situation combines the requisite clarity with the relevant ethical constraints. It is partly to preserve this clarity that I have avoided attributing to the parties any ethical motivation. They decide solely on the basis of what seems best calculated to further their interests so far as they can ascertain them. In this way we can exploit the intuitive idea of rational prudential choice. We can, however, define ethical variations of the initial situation by supposing the parties to be influenced by moral considerations. It is a mistake to object that the notion of the original agreement would no longer be ethically neutral. For this notion already includes moral features and must do so, for example, the formal conditions on principles and the veil of ignorance. I have simply divided up the description of the original position so that these elements do not occur in the characterization of the parties, although even here there might be a question as to what counts as a moral element and what does not. There is no need to settle this problem. What is important is that the various features of the original position should be expressed in the simplest and most compelling way.

Occasionally I have touched upon some possible ethical variations of the initial situation (§17). For example, one might assume that the parties hold the principle that no one should be advantaged by unmerited assets and contingencies, and therefore they choose a conception of justice that mitigates the effects of natural accident and social fortune. Or else they may be said to accept a principle of reciprocity requiring that distributive arrangements always lie on the upward sloping portion of the contribution curve. Again, some notion of fair and willing cooperation may limit the conceptions of justice which the parties are prepared to entertain. There is no *a priori* reason for thinking that these variations must be less convincing, or the moral constraints they express less widely shared. Moreover, we have seen that the possibilities just mentioned appear to confirm the difference principle, lending further support to it. Although I have not proposed a view of this kind, they certainly deserve further examination. The crucial thing is not to use principles that are contested. Thus to reject the principle of average utility by imposing a rule against taking chances in the original position would render the method fruitless, since some philosophers have sought to justify this principle by deriving it as the consequence of the appropriate impersonal attitude in certain risk situations. We must find other arguments against the utility criterion: the propriety of taking chances is among the things in dispute (§28). The idea

of the initial agreement can only succeed if its conditions are in fact widely recognized, or can become so.

Another fault, some may contend, is that the principles of justice are not derived from the notion of respect for persons, from a recognition of their inherent worth and dignity. Since the original position (as I have defined it) does not include this idea, not explicitly anyway, the argument for justice as fairness may be thought unsound. I believe, however, that while the principles of justice will be effective only if men have a sense of justice and do therefore respect one another, the notion of respect or of the inherent worth of persons is not a suitable basis for arriving at these principles. It is precisely these ideas that call for interpretation. The situation is analogous to that of benevolence: without the principles of right and justice, the aims of benevolence and the requirements of respect are both undefined; they presuppose these principles already independently derived (§30). Once the conception of justice is on hand, however, the ideas of respect and of human dignity can be given a more definite meaning. Among other things, respect for persons is shown by treating them in ways that they can see to be justified. But more than this, it is manifest in the content of the principles to which we appeal. Thus to respect persons is to recognize that they possess an inviolability founded on justice that even the welfare of society as a whole cannot override. It is to affirm that the loss of freedom for some is not made right by a greater welfare enjoyed by others. The lexical priorities of justice represent the value of persons that Kant says is beyond all price.³⁵ The theory of justice provides a rendering of these ideas but we cannot start out from them. There is no way to avoid the complications of the original position, or of some similar construction, if our notions of respect and the natural basis of equality are to be systematically presented.

These remarks bring us back to the common sense conviction, which we noted at the outset, that justice is the first virtue of social institutions (§1). I have tried to set forth a theory that enables us to understand and to assess these feelings about the primacy of justice. Justice as fairness is the outcome: it articulates these opinions and supports their general tendency. And while, of course, it is not a fully satisfactory theory, it offers, I believe, an alternative to the utilitarian view which has for so long held the preeminent place in our moral philosophy. I have tried to present the

35. See *The Foundations of the Metaphysics of Morals*, pp. 434–436, vol. IV of the Academy Edition.

theory of justice as a viable systematic doctrine so that the idea of maximizing the good does not hold sway by default. The criticism of teleological theories cannot fruitfully proceed piecemeal. We must attempt to construct another kind of view which has the same virtues of clarity and system but which yields a more discriminating interpretation of our moral sensibilities.

Finally, we may remind ourselves that the hypothetical nature of the original position invites the question: why should we take any interest in it, moral or otherwise? Recall the answer: the conditions embodied in the description of this situation are ones that we do in fact accept. Or if we do not, then we can be persuaded to do so by philosophical considerations of the sort occasionally introduced. Each aspect of the original position can be given a supporting explanation. Thus what we are doing is to combine into one conception the totality of conditions that we are ready upon due reflection to recognize as reasonable in our conduct with regard to one another (§4). Once we grasp this conception, we can at any time look at the social world from the required point of view. It suffices to reason in certain ways and to follow the conclusions reached. This standpoint is also objective and expresses our autonomy (§78). Without conflating all persons into one but recognizing them as distinct and separate, it enables us to be impartial, even between persons who are not contemporaries but who belong to many generations. Thus to see our place in society from the perspective of this position is to see it *sub specie aeternitatis*: it is to regard the human situation not only from all social but also from all temporal points of view. The perspective of eternity is not a perspective from a certain place beyond the world, nor the point of view of a transcendent being; rather it is a certain form of thought and feeling that rational persons can adopt within the world. And having done so, they can, whatever their generation, bring together into one scheme all individual perspectives and arrive together at regulative principles that can be affirmed by everyone as he lives by them, each from his own standpoint. Purity of heart, if one could attain it, would be to see clearly and to act with grace and self-command from this point of view.

Conversion Table

Index

CONVERSION TABLE

Section	1971 Edition: Pages	1999 Edition: Pages
1	3–6	3–6
2	7–11	6–10
3	11–17	10–15
4	17–22	15–19
5	22–27	19–24
6	27–33	24–30
7	34–40	30–36
8	40–45	36–40
9	46–53	40–46
10	54–60	47–52
11	60–65	52–56
12	65–75	57–65
13	75–83	65–73
14	83–90	73–78
15	90–95	78–81
16	95–100	81–86
17	100–108	86–93
18	108–114	93–98
19	114–117	98–101
20	188–122	102–105
21	122–126	105–109
22	126–130	109–112
23	130–136	112–118
24	136–142	118–123
25	142–150	123–130
26	150–161	130–139
27	161–166	139–144
28	167–175	144–153
29	175–183	153–160
30	183–192	160–168

Conversion Table

Section	1971 Edition: Pages	1999 Edition: Pages
31	195–201	171–176
32	201–205	176–180
33	205–211	180–185
34	211–216	186–190
35	216–221	190–194
36	221–228	194–200
37	228–234	200–206
38	235–243	206–213
39	243–251	214–220
40	251–257	221–227
41	258–265	228–234
42	265–274	234–242
43	274–284	242–251
44	284–293	251–258
45	293–298	259–262
46	298–303	263–267
47	303–310	267–273
48	310–315	273–277
49	315–325	277–285
50	325–332	285–292
51	333–342	293–301
52	342–350	301–308
53	350–355	308–312
54	356–362	313–318
55	363–368	319–323
56	368–371	323–326
57	371–377	326–331
58	377–382	331–335
59	382–391	335–343
60	393–399	347–350
61	399–404	350–355
62	404–407	355–358
63	407–416	358–365
64	416–424	365–372
65	424–433	372–380
66	433–439	380–386
67	440–446	386–391

Conversion Table

Section	1971 Edition: Pages	1999 Edition: Pages
68	446–452	392–396
69	453–462	397–405
70	462–467	405–409
71	467–472	409–413
72	472–479	414–419
73	479–485	420–425
74	485–490	425–429
75	490–496	429–434
76	496–504	434–441
77	504–512	441–449
78	513–520	450–456
79	520–529	456–464
80	530–534	464–468
81	534–541	468–474
82	541–548	474–480
83	548–554	480–486
84	554–560	486–491
85	560–567	491–496
86	567–577	496–505
87	577–587	506–514

Index

- Acton, Lord, 189n
Advice, 356–357, 393–394
Accidents of natural endowment, *see* Distribution of natural assets
Aggregative-distributive dichotomy, 32–33, 39, 279–280
Aiken, H. D., 452–453n, 454
Aims, *see* Desires
Albritton, R. G., 333n
Allocative justice, 56, 77
Alston, W. P., 420n
Alternatives: presentation of, §21:105–109; ways and difficulties of specifying, 105–107; short list of, 107; conditional alternatives excluded, 108; argument for principles relative to list of, 108–109; problem of ad hoc list in justification, 508–509
Altruism, perfect, 164–165; limited, *see* Mutual disinterestedness
Amdur, Robert, 99n
Analysis of meaning: role of in moral theory, 44–45, 112–113; and contract definition of right, 95–96, 161; and definition of goodness, 352–353; and thin theory of good, 355–358, 396; and justification, 506–507
Animals, 15, 441, 448
Anscombe, G. E. M., 49n, 420n, 559n
Approval, 356–357, 373–374
Aquinas, St. Thomas, 189–190, 351n, 554
Archimedean point, 230–232, 511
Arguments for conceptions of justice, §20:102–105; second part of contract theories, 14; intuitive idea of, 102–103; and procedure of social theory, 103; ideally deductive, 104–105; various interpretations of initial situation, 105
Aristotelian Principle, §65:372–380; defined, 374–375; and principles of rational choice, 364, 376; as principle of motivation, 375–376; companion effect of, 375–376, 413, 463, 500–501; influence on rational plans of life, 376; as tendency, 376–377; and chains of activities, 378; evolutionary explanation, 378–379; as deep psychological fact, 379–380; helps account for judgments of value, 379–380; connection with self-respect and the excellences, 386–388, 390–391; in morality of association, 412–413; and social union, 458–459, 463; in argument for congruence, 500–501; and publicity, 510
Aristotle, 45n, 79, 214, 351n, 488; definition of justice, 10; and perfectionism, 22, 286; and Aristotelian Principle, 374n; and happiness, 481n; on envy and spite, 466n, 468n
Arrow, K. J., 116n, 124n, 172n, 228n, 233n, 284n, 311n, 314n
Artificial identification of interests, 49, 173, 399n
Ashby, W. R., 400n
Assurance problem, 237–238, 277n, 296, 305–306
Austin, J. L., 356n
Authoritative decision, principle of, 209
Autonomy, §78:450–456; defined, 451–452; Kant's interpretation of, 222–226; and responsibility of citizens in civil disobedience, 341–342; and moral education, 451; and original position, 452–454; relation to objectivity, 452–454; and erring conscience, 454–455; and virtues of integrity, 455–456
Avineri, Shlomo, 460n

Background institutions, §43:242–251; and interpretations of the second principle, 61–64; and types of procedural justice, 74–76; and branches of government, 243–251; social minimum and transfer branch, 244–245; taxation and distribution branch, 245–247; in socialist

- Background institutions (*continued*)
 regime, 247–249; exchange branch, 249–251
- Background justice, 73–78; and fair equality of opportunity, 73–74, 76; role of basic structure in, 73–74, 76–78; as condition of pure procedural justice, 74–77; as condition of obligations, 96–97, 302, 330; and fair wages (distributive shares), 268–271. *See also* Background institutions
- Bad man, 385–386
- Baier, Kurt, 113n, 115n, 329n
- Bandura, Albert, 402n
- Bargaining, 116n, 120–121
- Barry, Brian, 30n, 33n, 229n, 287n, 291n
- Basic liberties, enumerated, 53
- Basic structure of society, §2:6–10; defined, 6–7; when well-ordered, defined, 4–5; justice as first virtue of, 3–4; as primary subject of justice, 6–10, 47, 73–74; principles of justice for not generally applicable, 7; several virtues of and social ideals, 8–9; two parts of, 53, 174–175; as background for procedural justice, 76–77; as shaping human wants and ideals of the person, 229–232; and relative stability, 400–401
- Baumol, W. J., 33n, 132–133n, 236n, 238n, 240n, 283n
- Beardsmore, R. W., 129n
- Beck, L. W., 221n
- Bedau, H. A., 320n
- Benevolence, 298n; act of defined, 385
- Benevolence: act of defined, 385; and supererogatory acts, 100–101, 167; compared as formal condition to mutual disinterestedness combined with veil of ignorance, 128–129; conflict of loves in, 166; and idea of splitting, 166–167; as second-order notion, 167
- Benn, S. I., 444n
- Bennett, Jonathan, 124n
- Bentham, Jeremy, 20n, 26, 27n, 287, 399; his utilitarianism contrasted with justice as fairness, 28–30; on the artificial identification of interests, 49, 173, 399n; Maine on his standard assumptions, 285
- Bergson, Abram, 229n, 240n
- Berlin, Isaiah, 177n, 202n, 254n
- Bickel, A. M., 342n
- Black, Duncan, 315n
- Blaug, Mark, 240n
- Blessedness, 842
- Boorman, Scott, 84n
- Bowman, M. J., 84n
- Bradley, F. H., 95n, 221n, 351n
- Braithwaite, R. B., 116n
- Brandt, R. B., 20n, 30n, 97n, 99n, 140n, 368n; and publicity condition, 159n
- Brierly, J. L., 332n
- Broad, C. D., 161n, 329n, 487n, 488n
- Broadly based properties, 382, 383–384
- Brown, Roger, 402n, 411n, 433n
- Buchanan, J. M., 61n, 173n, 235n, 236n
- Burke, Edmund, 217, 265
- Butler, Bishop, 467n
- Campbell, B. G., 378n
- Campbell, C. A., 373n, 381n
- Cantor, Georg, 45
- Cardinal utility, 78, 282–285
- Careers open to talents, 57–58, 62–63, 91–92
- Carnap, Rudolf, 146n
- Cartesian justification, 506–507
- Cassirer, Ernst,
- Caste systems, 84–85, 87–88, 91
- Categorical imperatives, 222–223, 511
- Cavell, Stanley, 389n
- Chain connection, 69–72
- Chains of activities, 377–378
- Chains of reasons, 432, 434
- Chakravarty, Sukamoy, 253n, 262n
- Chomsky, Noam, 41n
- Circumstances of justice §22:109–112; defined, 109; objective elements of, 109–110; subjective elements of, 110; existence of assumed in original position, 111; duties and obligations to third parties in, 111; motivational assumptions in, 111–112; presumes conflict of ideals, 112; role of in Kantian interpretation, 223–224, 226; two interpretations of, 456–457
- Civil disobedience, §55:319–323, §57:326–331, §59:335–343; defined, 320–323; to oppose savings policies, 261–262; as problem for democratic theory, 319; three parts of theory of, 319; and sense of justice, 320, 328, 339–340; as a political act, 320–321; and fidelity to law, 322; distinguished from militant action, 322–323; appropriate objects of, 326–327; three conditions of justification of, 326–329; problem of fairness in, 328–329; wisdom of, 330; and principle of fairness, 330–331; role of in maintaining democratic constitution, 335–338; theory of founded on conception of justice, 337–338; place of in constitutional theory,

- 338; and types of consensus, 340; citizen's responsibility in, 341–343
- Civility, duty of, 312
- Clark, J. B., 271n, 272n
- Clark, J. M., 271n, 272n
- Close-knitness, 70–72
- Cohen, Marshall, 320n
- Commending, 373–374
- Common good, defined, 205, 217
- Common interest, principle of: defined, 83, 217; and equal citizenship, 83; and toleration, 186–190; as political convention, 280
- Completeness, approximative, defined, 300–301, 496
- Complexity, limits on, 40, 114, 122–123, 453–454
- Condorcet, Marquis de, 314–315
- Congruence: concept of defined, 350, 399; problem of defined, 450–451, 496–499; argument for, 499–505. *See also* Sense of justice, good of
- Conscience: *see* Equal liberty of
- Conscientious evasion, 324
- Conscientious refusal: §56:323–326, §58:331–335; defined, 323–324; distinguished from conscientious evasion, 324; contrasted with civil disobedience, 324–325; toleration of and erring conscience, 325; general pacifism as natural departure, 325; law of nations derived, 331–333; justification of in case of particular war, 333–335; and conscription, 333–334; and contingent pacifism, 334–335; discriminating preferred to general pacifism, 335
- Conscription, 333–334
- Consensus, 16–17, 453–454, 508–510; overlapping vs. strict, 340
- Considered judgments: defined, 42; fixed points of, 17–18, 280, 507–509; role of in justification, 17–19, 104, 507–509; and reflective equilibrium, 18–19, 42–44; as facts to be explained by moral theory, 44
- Constant, Benjamin, 177, 195
- Constitution: justice of, 172–173, 194–206, 314; case of imperfect procedural justice, 173–174, 194, 311, 316; majority rule in, 200–203, 312–314; of ideal procedure of public deliberation, 315–316
- Constitutional convention, stage of, 172–174
- Constitutional democracy: defined, 195–196, 199–200; example of just basic structure, 171; constitutional devices of, 197; historical failure of, 198–199; and sense of justice, 214; liberties of, best founded on justice, 214; and just savings problems, 260–262; place of majority rule in, 312–314; economic theory of, 316–318; place of civil disobedience in theory of, 319, 338; role of civil disobedience in, 335–338. *See also* Equal participation, principle of
- Constitutive rules, 49–50, 303
- Content of conceptions of right: pointless excluded, 129–130; and psychological understandability, 418–419, 426–427
- Contingencies of social circumstance: and intuitive idea of the principles of justice, 14; in system of natural liberty, 62–63; in liberal equality, 63–64; in democratic equality and difference principle, 65, 67–68, 88–90
- Continuity, principle of, 369
- Contribution, precept of, 269, 270, 271, 273–274
- Contribution curve, 66–67, 89–90
- Coordination, problem of, 5–6
- Cooper, J. M., 374n
- Counter-examples, limited usefulness of, 45
- Counting principles of rational choice, 364, 365; in first-person procedure of choice, 483; and dominant ends, 484, 485–486; in hedonism, 486–487
- Culture: claims of in justice as fairness, 87, 289, 291–292, 387–388; claims of in perfectionism, 218, 285–286, 288–289; role of historical tradition of in social union, 460–461
- Dahl, R. A., 198n, 317n
- Darwin, Charles, 440n
- Davidson, Donald, 124n, 300–301n
- Definition, role of in moral theory, 44, 95–96, 112–113, 130, 176–177, 506–507
- Deliberative rationality, §64:365–372; defined following Sidgwick, 366–368; in definition of rational plan of life, 358–359; and ways of judging intensity of desires, 368–369; time-related principles and the principle of continuity, 369–370; as hypothetical criterion, 370, 371–372; and regret and self-reproach, 370–371; and responsibility to self, 371; does not imply excessive planning, 371–372; and failure of hedonism, 489; no way beyond, 490–491; and unity of the self, 491–492

- Democratic equality, 57, 65–73. *See also*
 Difference principle
 Deontological theories, 26–27, 35–36
 De Roover, R. A., 240n
 Descriptive theory of goodness, 356–358
 Desert, and distributive justice, 13–14, 88–89, 273–277
 Desires: rational (as interests) defined, 359; and rational plans, 359–361; choice of systems of, 364–365, 498; estimating relative intensity of, 368, 483; to act on principles of rational choice, as regulative, 364–365; laws governing changes in systems of, 375, 432–433; sense of justice as regulative, 430, 491, 497–499, 503; final always plural in number, 432, 484, 493; derivative and final distinguished, 432–433; and unity of the self, 492; final, and finality condition, 498–499, 503
 Devlin, Patrick, 291n
 Dewey, John, 351n, 358n
 Difference principle, §13:65–73; defined, 65–68; in democratic equality, 65; and perfectly just scheme, 68–70, 71–72; and scheme just throughout, 68; how related to efficiency, 69, 71; and chain connection, 69–72; and close-knitness, 69–72; and average utility, 71; lexical form of, 72; and interpersonal comparisons, 79; relation to principle of redress, 86–87; regards distribution of natural talents as common asset, 87–88, 156, 448, 512; and reciprocity, 88–90; as interpretation of principle of fraternity, 90–91; and meritocratic society, 91–92; and eugenics, 92–93; reasoning leading to, 130–131; objection that allows excessive inequalities, 135–137; and Kantian interpretation, 156–157, 437; and social minimum, 252–253; meaning of in savings problem, 253–254, 258; and priority of fair opportunity, 265–267; and mixed conceptions, 278–282; regulates balance of precepts, 280; as political convention of democracy, 280; relative clarity of, 281–282; and problem of envy, 465–466, 470; self-respect and index of expectations, 478–479; and ethical variations of initial situation, 512–513
 Diggs, B. J., 49n
 Distribution branch, 245–248
 Distribution of natural assets: and intuitive idea of the principles of justice, 11, 13; in system of natural liberty, 62–63; in liberal equality, 63–64; in democratic equality and difference principle, 87, 156, 447–448, 512; as neither just nor unjust, 87; and moral worth, 88–89, 274; and eugenics 92–93; relation to basis of equality, 443–447; and two conceptions of equality, 447–448; how characterizes human sociability and social union, 459–460
 Distributive justice: main problem of, 4, 6–7, 53, 74; and pure procedural justice, 76–77; as happiness according to virtue, 273–275; not opposite of retributive justice, 276–277
 Division of labor, 463–464, 474
 Dobzhansky, Theodosius, 92n
 Dominant end: defined, 484; not used in justice as fairness, 463, 495–496; cannot be happiness, 484–485; Loyola and Aquinas as illustrating, 485; extreme nature of, 485–486; and counting principles, 486; use of in hedonism, 486–490; necessity for choice of, 490–491; and unity of the self, 492–493
 Dostoevsky, Fyodor, 398n
 Downs, Anthony, 317n, 431n
 Dreben, Burton, 324n
 Due process of law, 210
 Duncan-Jones, A. E., 354n
 Duties to self, 218–219
 Duty, all things considered, 299–301
 Duty, *prima facie*, 299–301
 Duty to comply, *see* Political duty
 Dworkin, Gerald, 219n
 Dworkin, Ronald, 291n, 307n, 339n
 Economic systems, §42:234–242; use of economic theory, 234; private vs. public ownership, 235–239; public goods, 235–239; problems of isolation and assurance, 237–238; prisoner's dilemma, 238n; use of markets, 239–242; allocative and distributive function of prices, 241–242; choice between private-property economy and socialism not decided by justice alone, 241–242
 Economic theory of democracy, 316–318, 431–432
 Edgeworth, F. Y.: 20n, 26, 28, 29, 58n; his argument for utility principle criticized, 147–148; conflation of persons in, 164n; and cardinal utility, 282
 Education, 87, 92, 220, 452
 Effective means, principle of, 352–353, 361–362, 486
 Efficiency, principle of: defined, 58–60; and Pareto optimality, 58; applied to basic structure, 61–62; not a principle

- of justice, 62; role of in system of natural liberty, 62–63; role of in liberal equality, 63–64; relation to difference principle, 69, 71
- Efficiency, problem of, 5–6
- Egalitarianism, 471–472
- Egoism: types of listed, 107; inferior alternatives to principles of justice, 103, 117–118; excluded by constraints of right, 114, 117–118; general, as no agreement point, 117–118; justice as fairness not a case of, 127–128; and capacity for moral feelings, 427; problem of distinguished from problem of congruence, 497–498; and punishment, 504–505
- Eibl-Eibesfeldt, Irenäus, 378n, 379n, 440n
- Embedding: 28, 138–139, 231–232, 256–257, 285, 287
- Emotive theory of meaning, 357
- Ends, *see* Desires
- Entitlements, 276
- Envy, §§80–81:464–474; forms of defined, 466–468; problem of defined, 465–466, 468; and rationality, 124–125, 464–465; and two-part argument for principles of justice, 124, 464–465; special psychologies, 465–466, 474; not a moral feeling, 467, converses of, 467–468; excusable envy, 468; disposing conditions of, 469; in well-ordered society of two principles of justice, 469–471, 479; and equality, 471–474; and conditions of original position, 472; Freud on, and genesis of sense of justice, 472–473; and index of expectations, 478–479; how dealt with by caste and feudal systems, 479–480
- Equal liberty of conscience, §§33–35:180–194; and arguments for first principle of justice, 180–181; how leads to equal rights, 181–183; case for strengthened when descendants considered, 183; Mill's arguments for, 184–185; equal liberty insecure on teleological principles, 185; grounds for state's regulation of, 186–187; appeal to common sense and public knowledge in regulating, 188–190; and toleration, 188–190; toleration of the intolerant, 190–194; and stability of just institutions, 192–193; equal liberty and moral and cultural differences, 193–194, 583; and perfectionism, 287–290
- Equal participation, principle of §§36–37: 194–206; defined, 194, 196; two aspects of political justice, 194; features of a constitutional regime, 195–196, 199–200; extent of defined, 197; fair value of rights established by, 197–198; historical failure of constitutional regimes, 198–199; does not define an ideal of citizenship, 200; three ways of limiting, 200; justification for constitutional devices limiting extent, 200–202; and intensity of desire, 202–203; justification for inequalities of, 203–204; Mill on plural voting, 204–205; grounds for self-government, 205–206
- Equal respect, right to in determining principles for basic structure, 475
- Equality, basis of, §77:441–449; moral personality as 17, 289, 442–443, 505f; and natural rights, 442n; and natural attributes, 443–446; objection to procedural interpretation of, 444; in teleological theories, 445; as a potentiality, 445–446; simplicity of, relative to other views, 446–447; and reciprocity, 447; and two conceptions of equality, 447–448; and limits of justice as fairness, 448
- Equality, tendency to §17:86–93; principle of redress, 86–88; distribution of natural talents as a common asset, 87, 156, 447–448, 585; and reciprocity, 88–90; and harmony of interests, 89–90; principle of fraternity, 90–91; difference principle prevents meritocratic society, 91–92; eugenics, 92–93; and envy, 471–474
- Equality of consideration, 444
- Equality of fair opportunity, §14:73–78, §46:263–267; defined, 63; and the family, 64, 265, 447–448; and pure procedural justice, 73–77; role of in background justice, 74–75, 76–77; contrasted with allocative justice, 77; lexically prior to difference principle, 77–78; cases illustrating priority of, 264–265; priority rule for stated, 266–267; and two conceptions of equality, 447–448
- Equality of opportunity, formal, *see* Careers open to talents
- Equilibrium, 103, 399–401
- Equity, 209
- Erikson, Erik, 389n
- Erring conscience, 235–236, 454–455
- Eternity, perspective of, 514
- Ethics of creation, 137–138
- Eugenics, 92–93
- Evil man, 385–386
- Evolution, 378–379, 440–441

- Exchange branch, 249–251, 291–292
- Excellences, 388–391; defined, 389; and natural shame, 389–390; relation to moral virtues and moral shame, 390–391; and virtues of self-command, 391. *See also* Self-respect; Shame
- Expectations, defined, 56; and representative persons, 56–57; utilitarianism and accurate measure of, 78–79; how based on index of primary goods, 79–81; index problem for, 80; reasons for using primary goods as basis for, 80–81; lack of unity of in average utility, 150–152. *See also* Primary goods
- Exploitation, 272
- Fair political conduct, duty of, 210n
- Fairness, principle of, §18:96–98, §52:301–308; two parts of defined, 96, 301–302; covers all obligations, 96; Locke on and background justice, 96–97; characteristic features of, 97; political obligation for citizens generally problematical, 97–98, 295, 302; rejected as sole basis of political ties, 295–296; permits more discriminating account of requirements, 302–303; explains obligation to keep promises, 303–306; argument for, 304–306; requirements not founded on institutions alone, 306–307; and Prichard's question, 307–308; and political obligation for members of groups, 330–331
- Falk, W. D., 113n
- Family, institution of: and fair equality of opportunity, 64, 265, 447–448; and fraternity, 90; persons in original position as heads of, 111; in morality of authority, 405; in morality of association, 409–410
- Feinberg, Joel, 276n, 277n
- Fellner, William, 134n, 146n, 149n
- Fidelity, principle of, *see* Promises
- Fidelity to law, 322, 336–337
- Field, G. H., 374n, 418n
- Finality, as formal condition, 116–117; in argument for two principles, 153–155; in argument for congruence, 498–499, 503
- Findlay, J. N., 351n, 418n, 446n
- First principle of justice: first statement of, 53–54; final statement of, 220, 266; applies to first part of basic structure, 53, 174–175; as criterion for use in constitutional convention, 174; and equal liberty of conscience, 180–194; and political justice, 194–206; and rule of law, 206–213; meaning of priority of, 214–220; affirmed by mixed conceptions, 278; and perfectionism, 287–291; violations of, as appropriate object of civil disobedience, 326–327. *See also* Equal liberty of conscience; Political justice
- Firth, Roderick, 161n, 162n
- Fixed natural characteristics, 84–85
- Fixed points, of considered judgments, 17–18, 280, 507–508
- Flavell, John, 410n
- Fletcher, Ronald, 402n
- Foot, Philippa, 129n, 351n, 420n, 498n, 499n, 504n
- Formal constraints of concept of right, §23:112–118; not given by analysis of meaning, 112–113; propriety derives from function of moral principles, 113; generality, 113–114; universality, 114–115; publicity, 115; ordering, 115–116; finality, 116–117; exclude variants of egoism, 117–118
- Foster, G. M., 468n
- Four-stage sequence, §31:171–176; three kinds of political questions, 171–172; needed as schema to apply principles of justice, 172; as elaboration of original position, 172–173; constitutional convention, 172–174; as part of the theory of justice, 172n, 176; legislative stage, 174–175; division of labor between principles in, 174–175; stage of particular cases, 175; availability of knowledge in, 175–176
- Frankena, W. K., 22n, 113n, 444n, 446n
- Franklin, J. H., 338n
- Fraternity, 90–91
- Free and equal persons, 131–132, 339, 475; highest-order interests of, 131–132, 475–476
- Free association, principle of, 272–273, 289, 330–331
- Freedom of speech, 195–196, 197–198
- Free-rider: problem of, 236–239; egoism, 107, 117–118, 311–312, 340, 499
- Frege, Gottlob, 45
- Freud, Sigmund, 402n, 428, 472–474
- Fried, Charles, 223n, 322n, 371n
- Fuchs criterion, 84n
- Full system of principles, 299–300, 306, 434
- Fuller, Lon, 52n, 206n, 209n
- Galanter, Eugene, 358n
- Games as examples of social unions, 460–461

- Gauthier, D. P., 21n, 211n, 238n
 Geach, P. T., 356n
 Generality: as formal condition, 113–114, 160, 221; and variants of egoism, 114, 117–118
 Georgescu-Roegen, Nicholas, 38n
 Gewirth, Alan, 99n
 Gibbard, Allan, 14n, 20n, 294n
 Gierke, Otto, 10n
 Goethe, J. W., von, 286
 Goldman, Alvin, 358n
 Good, definition of, on thin theory, §§61–62:350–358; three-stage definition of, 350–351; illustrated by simpler cases, 352–353; moral neutrality of, 354–355; discussion of meaning in, 355–358
 Good, full theory of, 380–386; defined, 348, 349–350; and moral worth, 355, 381–384; and human goods, 373; and congruence, 496–503. *See also next entry*
 Good, thin theory of, §60:347–350; defined, 348–349; need for, 348–349, 350; accounts for primary goods, 348, 380–381, 392–393; contrasted with full theory of, 347–348, 349–350; three-stage definition of good in, 350–355; discussion of meaning in, 355–358; definition of good applied to plans of life, 358–361; principles of rational choice in, 361–364; and deliberative rationality, 364–372; general facts in, 372–373; and Aristotelian principle, 374–380; contrasted with concept of right, 392–394; and problem of congruence, 496–503
 Good (beneficent) act, 385
 Good (benevolent) action, 385
 Good faith, of original agreement, 153–154, 159
 Good society, 505
 Goodman, Nelson, 18n
 Goodness as rationality, *see* Good, thin theory of
 Gough, J. W., 10n
 Government, four branches of, 243–247, 249–251
 Greater likelihood, principle of, 362
 Gregor, M. J., 221n
 Grice, G. R., 10n
 Grudgingness, 467–468, 470–471, 473
 Guilt: feelings of, defined, 415–416, 421–422; Kant's ethics not an ethic of, 225; distinguished from shame, 391, 423–424; authority guilt, 407; and natural attitudes, 407, 412, 416, 427–429; association guilt, 412; principle guilt, 415–417; rational, defined, 416; psychological understandability of, 416–419; explanation of, 421–422; neurotic, 421; residue, 421–422; features of as moral sentiment, 422–424; and aspects of morality, 424; and finality condition, 503
 Halévy, Elie, 49n
 Happiness, §83:480–486; defined, 79, 480–481; as self-contained, 481–482; as self-sufficient, 482; and blessedness, defined, 482; not necessarily pursued by a rational plan of life, 482; of saints and heroes, 482–483; not a dominant end, 484–485
 Hardie, W. F. R., 9n, 45n, 374n, 481n, 484n
 Hare, R. M., 113n, 164n, 381n, 384n
 Harman, G. H., 145n
 Harmony of social interests, 89–90
 Harrison, Jonathan, 20n
 Harrod, R. F., 20n
 Harsanyi, J. C., 20n, 118n, 140n
 Hart, H. L. A., 5n, 48n, 96n, 97n, 109n, 180n, 210n, 212n, 277n, 291n, 301n
 Hedonism, §84:486–490; defined, 22, 486–487; as dominant-end method of first-person choice, 486–487; failure of, 488; not rescued by utility theory, 557; tendency to hedonism in teleological theories, 490–491; and unity of the self, and Mill's proof of utility, 492–493
 Hegel, G. W. F., 221n, 265, 457n
 Hempel, C. G., 124n
 Herzen, Alexander, 254
 Hicks, J. R., 147n
 Hobbes, Thomas, 10n, 211, 238, 304–305
 Hoffman, M. L., 402n, 403n
 Hollingsdale, J. R., 286n
 Homans, G. C., 430n
 Houthakker, H. S., 38n
 Human goods, 373–374, 379–380
 Humboldt, Wilhelm von, 459–460
 Hume, David, 8, 20n, 109n; criticism of Locke, 29n, 296n; and circumstances of justice, 110; and impartial sympathetic spectator, 161–163, 165–166, 223
 Hutcheson, Francis, 20n, 38n
 Ideal market process, 316–318
 Ideal observer, 161–162
 Ideal of the person, 231–232
 Ideal procedure, 314–318
 Ideal-regarding principles, 287
 Ideal theory: defined and contrasted with nonideal theory, 7–8, 215–216, 308–309; as fundamental part of theory of

- Ideal theory (*continued*)
 justice, 8, 212, 343; and penal sanctions, 212, 277; and priority rules, 212, 267. *See also* Strict compliance
- Idealism, 232–233
- Ideals, of roles, 404n, 409–410, 413, 414
- Illocutionary forces, 356–358
- Impartial sympathetic spectator: in definition of right, 161–162; in utilitarianism, 23–24, 26, 29–30, 161–166; Hume's account of, 161–163, 165–166, 233
- Impartiality, 163–166
- Inclusive end, 484, 489
- Inclusiveness, principle of, 362
- Indeterminacy of justice, 176, 318; of choice, 484, 490, 493–494
- Indifference, principle of, 485
- Indifference curves, 33–34
- Inheritance, 245–246
- Initial situation: defined, 15–16, 105; relation to original position, 16–17, 105; many interpretations of, 105, 109; as analytic method of comparing conceptions of justice, 105, 165; list of variations and elements of, 126–127; ethical variations of, 512
- Instability, two kinds of, 295–296, 435
- Institutions, §10:47–52; defined, 47–48; arrangement of major, primary subject of principles of justice, 6–7, 47; existence and publicity of rules of, 48–49; constitutive rules of distinguished from strategies, 49–50; and artificial identification of interests, 49; and formal justice, 50–52; as defining content of obligations, 97
- Integrity, virtues of, 455–456
- Interpersonal comparisons of well-being: in utilitarianism, 78, 284–285; in justice as fairness, 79–80, 81; role of primary goods in, 79, 81, 188; and unity of expectations, 151–153; and some procedures of cardinal utility, 282–285; moral presuppositions in, 284–285
- Intuitionism, §7:30–36; broad vs. traditional sense of, defined, 30–31; types of, by levels of generality, 31–33; represented by indifference curves, 33–34; and priority problem, 34–35, 36–37; incomplete but not irrational, 34–35, 36–37; may be either teleological or deontological, 35; in mixed conceptions, 279–281; in more common forms of perfectionism, 286–287, 290–291
- Isolation problem, 237–238, 295
- James, William, 390n
- Jealousy, 467–468, 472–473
- Jevons, W. S., 38n
- Jouvenal, Bertrand, 287n
- Judicial virtues, 453
- Just savings principle, §44:251–258; motivation assumption for, 111, 121, 254–256; needed to determine social minimum, 251–252; and time preference, 253, 259–262; in classical utilitarianism, 253, 262; construction of in contract theory, 253–258; relation to difference principle, 253–254; public savings policies and democratic principles, 260–262; and priority questions, 263–264; in final statement of two principles, 266–267; and principle of political settlement, 318. *See also* Time preference
- Just war, 332–335
- Justice, concept of: as distinct from conception of, defined, 5, 8–9; as first virtue of institutions, 3–4, 513; principles of assign basic rights and duties and regulate competing claims, 4–6, 9, 112; primary subject of, basic structure, 6–10, 47; but one part of a social ideal, 8–9
- Justice, conceptions of: as distinct from concept of, defined, 5, 8–9; role of principles of, 4–6, 9; distinguished from social ideals, 9; content of, 129–130; degrees of reasonableness of, 309–310. *See also* Stability, of conceptions of justice
- Justice, formal, 50–52, 156, 207–210, 441–442
- Justice, general conception of: principle of stated, 54; relation to two principles of justice (as special conception), 54–55, 217–218; lacks definite structure, 55
- Justice, natural duty of: defined, 99, 293–294; and political obligation, 100, 296, 310n; and toleration of the intolerant, 192; argument for, 293–296; and duty to a just constitution, 308–312; weight of and civil disobedience, 336
- Justice, primary subject of, 6–10. *See also* Basic structure
- Justice, role of, §1: 3–6; as first virtue of institutions, 3–4, 513; to assign basic rights and duties and to regulate competing claims, 4–5; concept and conceptions of distinguished, 5; and other social problems, 5–6
- Justice, substantive, 50–52
- Justice, two principles of special concep-

- tion of, §11:52–56, §26:130–139, §29:153–160; first and final statements of, 53, 266; defined as special conception when in lexical order, 53–54; as special case of general conception, 54–55; as long-run tendency of justice, 55, 132, 541f; primary goods in, 54–55, 79–81; consequences of applying to institutions, 55–56; representative persons in, 56, 81–86; initial argument for, 130–132; as maximin solution, 132–135 place of general facts in arguments for, 136–139; argument for from finality and strains of commitment, 153–154; argument for from publicity and constraints on agreements, 154–155; argument for from self-respect and treating persons as ends, 155–159. *See also* First principle of justice; Second principle of justice
- Justice as fairness, §3:10–15; intuitive idea of defined, 10–12; not an account of ordinary meaning, 9; hypothetical nature of, 11, 18–19, 104, 145, 514; name of explained, 11; analogue of state of nature in, 11; intuitive idea of principles of, 13–14; two parts of, 14; and propriety of term “contract,” 14–15; limited scope of, 15, 448; as a deontological theory, 26; and priority of right, 27–30; embedding of ideals in, 28, 138–139, 231; appeal to intuition in, 36–40; and priority problem, 36–40; and complexity of moral facts, 40; as a moral theory, 41–45, 104–105; simplifying devices in, 46, 77–78, 81, 123, 433, 453–454; and pure procedural justice, 104, 118; not egoistic, 127–129; reliance on general facts in, 137–139; concept of impartiality in, 165–166; claims of culture in, 288–289, 291–292, 387–388; as natural rights theory, 442–443n; structure of, 496; some objections to considered, 511–513
- Justice as regularity, *see* Justice, formal
- Justice between generations: and eugenics, 92–93; and motivational assumption in original position, 111, 121; and veil of ignorance, 118–119, 121; in savings problem, 121, 251–258; and liberty of conscience, 183; in historical tradition of social union, 460–462. *See also* Just savings
- Justification, §4:15–19, §87:506–514; as a problem of rational choice, 15–16; presupposes some consensus, 16–17, 508–509; role of considered judgments and principles in, 17–19, 105, 508–511; as mutual support of many considerations, 18–19, 507; Cartesian and naturalistic rejected, 506–507; three parts of exposition as, 507–508; objections to method of considered, 508–511; some objections to justice as fairness discussed, 511–513; and ethical variations of initial situation, 512–513
- Kaldor, Nicholas, 246n
- Kant, Immanuel: in tradition of social contract, 10; on priority of right, 28n, 38n, 513; theory of the good, 79–80, 351n; publicity condition in, 115, 221, 298n; veil of ignorance implicit in, 121–122, 222; doctrine not egoistic, 127–128; treating persons as ends interpreted, 156–159, 437; and Kantian interpretation of justice as fairness, 221–227; his ethics one of mutual respect, 225; and Rousseau, 225, 233; duty of mutual aid in, 297–298; on savings, 254; priority rule for requirements, 297–298; on moral learning, 402–403; on social union, 459n; definition of envy, 466
- Kantian interpretation of justice as fairness, §40:221–227; moral principles as object of rational choice, 221–222; notion of autonomy in, 222; principles of justice as categorical imperatives, 222–223; and mutual disinterestedness, 223–224; Sidgwick’s objection, 224–225; original position as procedural interpretation of Kant’s ethics, 226–227; in explanation of moral shame, 390–391; in psychological understandability of morality, 417, 418; in idea of social union, 462–463; effect on understanding unity of the self, 493–495; unanimity condition of, 494–495; in argument for congruence, 501
- Kaufmann, Walter, 469n
- Kenny, Anthony, 481n, 482n, 490n.
- Keynes, J. M., 146n, 263–264
- King, Martin Luther, 320n
- Kirchenheimer, Otto, 206n
- Kneale, W. K., 161n
- Knight, F. H., 199n, 274n, 314n
- Kohlberg, Lawrence, 403n, 404n
- Koopmans, T. C., 58n, 240n, 253n
- Kyburg, H. E., 149n
- Lamont, W. D., 351n
- Laplace, Marquis de, 146

- Laplacean rule, 148
- Law of nations, 7, 93, 99, 401; and natural duties, 99; derivation of, 332–333; and just war and conscientious refusal, 333–335
- Least advantaged class, defined, 83–84
- Legal system, defined, 207–208
- Legality, principle of, *see* Rule of law
- Legislative stage, 174–175
- Legitimate expectations, §48:273–277; and just law, 207; not based on moral desert, 273–274; moral worth not rewarded by following precepts of justice, 273–275; how arise in well ordered society, 275–276; entitlements, deservingness, and moral worth distinguished, 276; distributive justice not opposite of retributive, 276–277
- Leibenstein, Harvey, 400n
- Leibniz, G. W. von, 273n
- Lessnoff, Michael, 264n
- Lewis, C. I., 164n, 165n, 351n
- Lewis, D. K., 115n
- Lexical order: defined, 37–38; in justice as fairness, 38–39; as simplifying device, 38–39, 40, 77–78; of two principles of justice, 53–54, 130–131; and lexical form of difference principle, 72. *See also* Priority *entries*
- Liberal equality, 57, 63–65
- Liberty, concept of, §32:176–180; as pattern of social forms, 55–56, 177, 180n, 210; triadic form of, 177; question of positive and negative, 176–177; total system of, 178, 201–202, 213, 220; assessed from standpoint of equal citizens, 178–179, 217; worth of liberty and the end of social justice, 179; political, fair value of, 197–199; and rule of law, 210–213; and paternalism, 218–220
- Life prospects, *see* Expectations
- Linguistic theory, 41, 430
- Little, I. M. D., 62n, 124n, 147n, 311n
- Locke, John, 10, 28–29, 97, 114, 189–190
- Loev, Gerald, 322n
- Lorenz, Konrad, 440n
- Lottery schemes, 329
- Louch, A. R., 291n
- Love: defined, 166, 406–407; of many persons, problem of, 166–167; law of in psychological laws, 405–408, 411–412, 416, 433; how related to moral feelings, 407, 425–429; in explaining psychological understandability of morality, 416–417 as family of dispositions, 426–427; hazards of in argument for congruence, 502–503. *See also* Benevolence; Love of mankind; Supererogation; Supererogatory acts
- Love of injustice, 385–386
- Love of mankind, 166–167, 417, 419, 424
- Lovejoy, A. O., 460n
- Loyola, St. Ignatius of, 485n
- Lucas, J. R., 109n, 116n, 206n
- Luce, R. D., 74n, 133n, 150n, 238n, 283n, 392n, 489n
- Lyons, David, 20n, 329n
- Mabbott, J. D., 20n, 124n, 410n
- MacCallum, G. G., 177n
- Maccoby, E. E., 408n
- Maine, H. S., 285
- Majority rule, §54:313–318; bare, defined, 197; and principle of participation, 197, 200–201; circumscribed by constitutional devices, 200–202; and intensity of desire, 202–203, 317–318; and bill of rights, 203; argument for in just constitution, 311–312; limits of majority principle, 311–312; status of, 313–314; role of in ideal procedure, 314–316; contrast with ideal market process, 316–318; and principle of political settlement, 318
- Marglin, S. A., 136n, 260n
- Markets, use of in economic systems, 239–242
- Marshall, Alfred, 229
- Marx, Karl, 229, 249n, 268n, 271–272n, 460n, 538n
- Maximal class of plans, 359, 365
- Maximin rule: defined, 132–133; as heuristic device for arranging arguments for principles of justice, 132–136, 153; situations when reasonable, 134; and original position, 134–135
- McCloskey, Herbert, 313n
- McCloskey, H. J., 30n
- McDougall, William, 389n, 404n
- Mead, G. H., 410n
- Meade, J. E., 241–242, 245n
- Meiklejohn, Alexander, 178n
- Mens rea*, 212
- Meritocratic society, 91–92
- Method of choice, first-person, 283, 484, 486–489, 492, 494
- Militant action, 322–323
- Mill, James, 402n
- Mill, J. S., 20n, 108, 180n, 197, 374n, 402n, 429n, 460n, 476; on force of precepts of justice, 23n; on priority problem, 36; lexical order in, 38n; as holding average utilitarianism, 140;

- arguments for liberty considered, 184–185; argument for plural voting, 204–205; on value of self-government, 205–206; on balancing precepts of justice, 268; on moral learning, 403, 439; proof of utility interpreted, 492–493
- Miller, G. A., 358n
- Mixed conceptions, §49:277–285; list of, 107; appeal of, 278; with social minimum and distributional constraints, 278–279; institutionistic features of, 279–281; and difference principle, 280–282; vagueness of, 281–283; cardinal utility and interpersonal comparisons, 282–285
- Moderate scarcity, 110, 226
- Moore, G. E., 30, 35, 287n
- Moral attitudes, *see* Moral sentiments
- Moral education, 429
- Moral geometry, 105, 109
- Moral learning theory: two traditions of, 401–404; presupposes moral theory, 404, 430, 434; in justice as fairness, 429–439
- Moral person (personality): defined, 11, 17, 442; as basis of equality, 17, 289, 442–447; and perfectionism, 289; and duty of mutual respect, 297; freedom and equality of 475; and unity of the self, 491–492
- Moral principles, function of, 113–114, 115–116, 510–511
- Moral psychology, principles of, §75:429–434; first law, 406, 429; second law, 412–413, 429; third law, 414–415, 429–430; refer to the principles of justice, 430; moral conceptions in psychological and social theory, 430–432, 434; as laws of transformation of systems of final ends, 432–433; as reciprocity principles, 433; and relative stability, 434–439; and evolution, 440–441
- Moral reasons, 97, 300–301, 306
- Moral sentiments, §§73–74:420–429; in what sense independent of contingencies, 416; some terms for explained, 420; relation to characteristic sensations and behavior, 420–421; moral principles in explanations of, 421–422; how resolved and relation to attitudes of others, 422–423; guilt and shame, 424; connection with natural attitudes, 425–427; as normal feature of human life, 427–429; egoists incapable of, 428; may take irrational forms, 428–429
- Moral theory: nature of, §9:40–46; as attempt to describe our moral capacities, 41–42; compared with linguistics, 41, 430; considered judgments in, 42–44, 507–509; and reflective equilibrium, 42–44; as a theory of moral sentiments, 44; place of definitions in, 44–45, 95–96, 112–113, 130; what to expect of, 46, 176, 319–320; simplifying devices in, 46, 77–78, 433; general facts in, 137–139, 234, 372–373, 399, 404–405; role of in psychological and social theory, 430–432, 434. *See also* Justification
- Moral worth of persons, §66:380–386; defined, 382–384; not a basis of distributive justice, 274–275; problem belongs to full theory of good, 349–350, 381–382; primary goods established, 380–381; distinguished from natural assets, 383; and idea of function, 384; definition of good extended to other cases, 384–385; as itself a good, and question of congruence, 496–504
- Morality of association, §71:409–413; as second stage of morality, 409; institutional context and role ideals of, 409–410; intellectual development in, 410–411; second psychological law and (association) guilt, 411–412; and Aristotelian principle, 413; features and virtues of, 414
- Morality of authority, §70:405–409; as first stage of morality, 405; family as institutional context of, 405; first psychological law and (authority) guilt, 406–407; conditions favorable to acquiring, 407–408; features and virtues of, 408–409
- Morality of principles, §72:414–419; as final stage of morality, 414; institutional context of, 414; third psychological law and (principle) guilt, 414–416; sense of justice and particular attachments, 416; psychological understandability of, 416–419; two forms of, their features and virtues, 419
- Morgan, G. A., 286n
- Murphy, J. G., 11n, 221n
- Musgrave, R. A., 243n, 247n
- Mutual advantage, *see* Reciprocity
- Mutual aid, duty of, 98, 297–298, 357
- Mutual disinterestedness: defined, 12; and circumstances of justice, 110–112; distinguished from egoism, 111; and meaning of rationality, 125; combined with veil of ignorance compared to benevolence, 128–129, 131; and principles

- Mutual disinterestedness (*continued*)
for guiding benevolence, 167; in
characterization of autonomy, 223–
224, 511
- Mutual respect, duty of, 94, 156, 297, 447
- Myrdal, Gunnar, 140n
- Nagel, Thomas, 166n, 371n
- Nash, J. F., 116n
- National interest, legitimate, 332–333
- Natural aristocracy, 57, 64–65
- Natural assets, *see* Distribution of natural
assets
- Natural attitudes: and moral sentiments,
405–408, 412–413, 414–418, 425–
429; and sense of justice in argument
for congruence, 499–500
- Natural abilities, and basis of equality,
444–447
- Natural duties, §19:98–101, §51:293–
301; defined, 98–99; order of choice
of, 93–95; duty of justice and other ex-
amples of, 98–99; relation to super-
erogatory acts, 100–101; to other gen-
erations, 183, 258; argument for duty
of justice, 293–296; duty of mutual re-
spect, 297; of mutual aid, 297–298; pri-
ority problem for, 298–299; *prima facie*
duty and duty all things considered,
299–301
- Natural justice, precepts of, 209–210
- Natural lottery, *see* Distribution of natural
assets
- Natural rights, 25, 28, 505n
- Naturalism, 506–507
- Near justice, state of, 311, 312; defined,
309, 319
- Necessary truth, 19, 44, 506
- Need, precept of, 244–245, 271, 274
- Neumann-Morgenstern definition of car-
dinal utility, 283–284
- Nietzsche, Friedrich, 22, 286, 469n
- No offense without a law, precept of, 209
- Noblesse oblige*, 64, 100
- Noncomparing groups, 387–388, 470–
471, 477–478, 479
- Nonideal theory, 215–218, 267, 308–309,
343; defined, 216. *See also* Partial com-
pliance
- Nozick, Robert, 30n
- Objective rationality of plans, 366, 370–
371
- Objectivity, 452–455
- Obligations, *see* Fairness, principle of
- Olson, Mancur, 236n
- Oppressive conscience, 428–429
- Oppenheim, Felix, 177n
- Ordering, as formal condition, 115–116,
117
- Original position, §4:15–19; defined as fa-
vored interpretation of initial situation,
16, 105; relation to social contract the-
ory, 10–11; as hypothetical situation,
11, 19, 104, 145, 587; as fair initial
status quo, 15–16, 104, 116n; and
justification, 18–19, 510–513; as guide
to intuition, 19; nature of arguments
from, 102–105; presentation of alterna-
tives in, 105–109; and circumstances
of justice, 109–112; formal constraints
of concept of right in, 112–118; and
veil of ignorance, 118–123; present
time of entry interpretation of, 119–
120, 254–255; unanimity in, 120, 233;
rationality of parties in, 123–130; list
of elements of, 126–127; highest-order
interests of parties, 131–132, 475–476;
choice under uncertainty in and maxi-
min rule, 134–135; conditions of con-
trasted with impartial sympathetic spec-
tator, 163; Kantian interpretation of,
221–227; as procedural interpretation
of Kant, 226–227, 233; and savings
problem, 121, 254–256; and time pref-
erence, 259–260; form of in deriving
law of nations, 331–332; and responsi-
bility to self, 371; and autonomy and
objectivity, 452–454; conditions of and
problem of envy, 472; embodies reci-
procity and equality between free and
equal moral persons, 475
- Ought implies can, precept of, 208, 213
- Overlapping consensus, 340
- Pacifism: general, 325, 335; contingent,
334–335
- Page, A. N., 58n
- Pareto, Vilfredo, 58, 103
- Partial compliance theory, 8, 212–213,
215–218, 309, 504–505
- Paternalism, 183, 218–220
- Paton, H. J., 221n, 351n
- Pattanaik, P. K., 140n
- Paul, G. A., 184n
- Pearce, I. F., 37–38n
- Pennock, J. R., 90n, 313n
- Perelman, Ch., 51n
- Perfectionism, §50:285–292; defined as a
teleological theory, 22; scope of intui-
tion in, 35–36, 286–287, 290–291;
claims of culture in, 218, 285–286;
two forms of distinguished, 285–287;
relation to want- and ideal-regarding

- principles, 287; argument against strict view from equal liberty, 287–290; interpretation of original position in, 288–289; analogue of standard assumptions in, 290; argument against moderate view, 290–292; claims of culture in justice as fairness, 291–292, 470; rejection of as political principle and democracy of association, 387–388, 462; as psychologically understandable, 417
- Pericles, 114
- Permissions, 100
- Perry, R. B., 21n, 90n, 351n, 358n; his view compared with justice as fairness, 122, 128; and principle of inclusiveness, 362
- Person, and rational plans, 358, 370–371
- Petrarch, 488
- Piaget, Jean, 403n, 404n
- Piers, Gerhart, 389n
- Pigou, A. C., 20n, 28, 272n
- Pitcher, George, 420n
- Pitkin, H. F., 97n, 99n, 200n, 217n
- Plamenatz, J. P., 99n
- Planning activity, rationally of, 371–372
- Plans of life; §63:358–365; defined 79–80, 358; rationality of defined, 358–359; as determining a person's good, 75–80, 358–359, 370; maximal class of, 359; features of, 359–361; subplans of, 361; principles of rational choice for, 361–364; and Aristotelian principle, 364, 376–377, 379; possibility of choice between, 364–365; objectively and subjectively rational defined, 366–367; satisfactory plans, 367; as determining shame, 390
- Plato, 398n, 457n
- Pleasure, 486–490
- Poincaré, Henri, 19n
- Pole, J. R., 204n
- Political duty, §53:308–312; for citizens generally, 100, 310n, 330–331; to just laws, 308; cases of ideal and partial compliance theory distinguished, 308–309; two contexts of unjust laws, 309–310; duty to unjust laws as duty to just constitution, 310–312; and majority rule, 311–314
- Political economy, §41:228–234; conception of defined, 228–229; and welfare economics, 228–229; need for ideal of the person in, 229–232; problem of Archimedean point and embedding of ideals, 230–232; and assumption of unanimity, 232–233; and values of community, 233–234
- Political justice, defined, 194. *See also* Equal participation, principle of
- Political obligation, *see* Fairness, principle of
- Political settlement, principle of, 318
- Population size, 140–141
- Postponement, principle of, 360, 369
- Potter, R. B., 332n
- Precepts of justice; §47:267–273; defined, 31–32; in utilitarianism, 23, 25, 268; in rule of law, 207–210; balance of, 244–245, 268, 270–271, 279–280; in pure procedural justice and fair wages, 268–270; subordinate place of, 270–271; and imperfections of competition, 272–273; and moral worth, 274
- Prescriptive theory of meaning, 357–358
- Pribram, K. H., 358n
- Price, Richard, 30n
- Prices, allocative and distributive functions of, 241–242
- Prichard, H. A., 30n, 305, 307–308
- Primary goods; §15:78–81; defined, 54–55, 79; social and natural distinguished, 54; self-respect as most important, 54, 155–156, 348, 386; index of basis of expectations, 79–81; role of in interpersonal comparisons, 79, 81, 285; index problem for, 80–81; reasons for using to define expectations, 80–81, and rationality of so doing in original position, 123; derived from general assumptions, 223, 230; accounted for by thin theory of the good, 348–349, 380–381, 392. *See also* Expectations
- Principle of insufficient reason, 144–145, 145–147, 148
- Priority of fair opportunity: defined, 77; cases illustrating, 264–265; rule of stated, 266–267
- Priority of justice, 3–4, 69, 263–264; rule of stated regard in savings, 266–267
- Priority of liberty, §39:214–220, §82:541–548; meaning of, 132, 214–220; rule of stated, 220, 266; cases illustrating, 200–205, 212–213, 215–216, 216–218; best secured by justice, 214; ideal and nonideal theory defined, 216–217; and paternalism, 218–220; and conscription, 333–334; reasons for from first part of argument for the principle, 131–132, 475–476; basis of in parties' fundamental aims and interests, 131, 475; basis of in parties' highest-order interests, 131–132, 475–476; reasons for priority from second part of argument, 475–480; and desire for economic

- Priority of liberty (*continued*)
 advantages, 476–477; argument for
 from desire for status and self-respect,
 477–479; and publicity condition and
 true general beliefs, 479–480
- Priority of right: defined, 27–28; in jus-
 tice as fairness, 27–28, 38n, 394–395,
 396, 494–496; and indeterminacy of
 the good, 394–396; and the unity of
 the self, 493; how affects indeterminacy
 of choice, 493–494
- Priority problem, §8:36–40; three ways
 of meeting, 36–39; in utilitarianism
 and intuitionism, 36; in justice as fair-
 ness, 37–40, 55; and lexical order, 37–
 39, 40; limiting the appeal to intuition
 in, 39; and natural duties, 98; rules for
 stated for justice, 220, 266–267; and
 principles for individuals, 298–299
- Prisoner's dilemma, 238, 505
- Private property economy, 235–236, 239–
 242
- Private society, 457–458
- Probability, concept of, 149–150
- Procedural justice, perfect, 74, 316
- Procedural justice, imperfect: defined, 74–
 75; in classical utilitarianism, 77; and
 just constitution, 173–174, 194, 310–
 312; and ideal procedure, 316
- Procedural justice, pure: defined, 74–75;
 and background justice, 58, 75–77;
 and fair equality of opportunity, 73–
 77; advantages of, 76–77; and original
 position, 104, 118; and fair wages,
 267–291
- Procedural justice, quasi-pure, 176, 318
- Promises, 97, 98, 303–306, 307–308
- Protestant reformers on toleration, 190
- Public forum, 197–198, 321, 328–329,
 330, 544f
- Public goods, 235–239, 295–296
- Publicity: as implicit in contract theory, 15,
 153; of rules, 48–49; as formal condi-
 tion, 113n, 115, 397–398; in argument
 for stability, 154–158; of general beliefs,
 397–398, 479–480; and envy, 479–
 480; in argument for congruence, 499;
 and justifications in social union, 510
- Punishment, 211–212, 276–277, 504–505
- Purely conscientious act, doctrine of,
 418, 499
- Purely preferential choice, *see* Indetermi-
 nacy of choice
- Purity of heart, 514
- Quasi-stability, 400n
- Quine, W. V., 95n, 113n, 507n
- Raiffa, Howard, 74n, 133n, 150n, 238n,
 283n, 392n, 489n
- Ramsey, F. P., 252–253n, 259n
- Ramsey, Paul, 332n
- Rancor, *see* Envy
- Raphael, D. D., 86n
- Rashdall, Hastings, 287n
- Rational choice, principles of: apply to
 plans of life, 358–359; at best deter-
 mine maximal class, 359, 365–366;
 time-related principles defined, 360,
 369–370; counting principles defined,
 361–364; specify higher-order desires,
 364; not unanimously chosen, 392; for
 choice under uncertainty, 392–393;
 and veil of ignorance, 394. *See also*
 Uncertainty, choice under
- Rationality of the parties, §25:123–130;
 defined, 123–124; and envy, 124–125,
 464–474; and mutual disinterestedness,
 125; how related to strict compliance
 condition, 125–126; elements of initial
 situation and variations listed, 126–
 127; an aspect of theoretically defined
 individuals, 127; relation to egoism
 and benevolence, 127–129; and con-
 tent of morality, 129–130
- Reasonableness, of conception of justice,
 277–278, 309–310
- Reciprocity: utilitarianism incompatible
 with, 13, 29–30, 437–438; in differ-
 ence principle, 88–90; in harmony of
 interests, 89–90; in argument from sta-
 bility and mutual respect, 155–156; in
 conditions of consensus, 340; as char-
 acterizing psychological laws, 433,
 437–439; and the basis of equality,
 447; embodied in original position,
 475; as ethical variation of initial situ-
 ation, 512
- Redress, principle of, 86–87
- Reflective equilibrium, 18–19, 42–45,
 104, 379, 381, 507–508
- Region of positive contributions, 68–69,
 71, 88
- Regret, 370–371, 388–389, 389–390, 421
- Relevant social positions, §16:81–86;
 defined, 81–82; and starting places, 82,
 85–86; two main cases of, 82–83, 84–
 85; equal citizenship as, 82–83; least
 advantaged as defined, 83–84; and
 fixed natural characteristics, 84–85;
 need for account of, 85–86
- Remorse, 421
- Representative persons, defined, 56
- Requirements, of obligation and natural
 duty, 96–101

- Rescher, Nicholas, 30n, 279n
- Resentment, 416, 423, 467, 468, 472–474
- Respect for persons, 455, 513, in Kant's ethics, 225. *See also* Equal Respect; Mutual Respect, duty of; Self-respect
- Responsibility, principle of, 212, 341–342, 455
- Responsibility to self, principle of, 371
- Retributive justice, 276–277
- Right, complete conception of, 93–96, 299–301, 306
- Right, concept of: contract definition of, 95–96, 161–162; formal constraints of, 112–118; generality, 113–114; universality, 114–115; publicity, 115; ordering, 114–115; finality, 116–117; ideal observer definition of, 95–96; contrasted with the good, §68:392–396; with respect to need for agreement, 392–393, to diversity of conceptions of, 393–394; and veil of ignorance, 394; priority of right in contrast to utilitarianism, 27–30, 437–439; and analysis of meaning, 396
- Rightness as fairness, 15, 95–96
- Rodes, Robert, 64n
- Ross, W. D., 30n, 35, 38n, 273n, 299–301, 351n, 354n, 418
- Rousseau, J. J., 10, 121n, 189–190, 233, 402–403, 406n, 473
- Royce, Josiah, 351n, 358
- Ruggiero, Guido, 177n
- Rule of law, §38:206–213; legal system defined, 207; precept of ought implies can, 208; precept of similar cases, 208–209; precept of no offense without a law, 209; precepts of natural justice, 209–210; relation to liberty, 210–211; and penal sanctions and principle of responsibility, 212; and cases illustrating priority of liberty, 212–213
- Runciman, W. G., 447n
- Ryle, Gilbert, 490n
- Sachs, David, 425n
- Samuelson, P. A., 489n
- Santayana, George, 64n, 488
- Savage, L. J., 149n
- Savings, *see* Just savings principle
- Scanlon, T. M., 382n
- Schaar, John, 91n
- Scheler, Max, 469n, 479n
- Schiller, Friedrich, 460n
- Schneewind, J. B., 45n
- Schopenhauer, Arthur, 127–128
- Schultz, R. A., 572n
- Schumpeter, J. A., 317n
- Scitovsky, Tibor, 228n
- Searle, J. R., 49n, 303n, 335n
- Second principle of justice, §12:57–65; first and second statements of, 53–54, 72; final statement of, 266 interpretations of, 57–65; and system of natural liberty, 57–58, 62–63, 65; and principle of efficiency, 58–62; and liberal equality, 63–65; and natural aristocracy, 64–65; and democratic equality and difference principle, 65–73; and legislative stage, 174–175
- Self, unity of, §85:174–175; in dominant-end views, 492; in hedonism and Mill's proof of utility, 492–493; in justice as fairness, 493–494; and unanimity condition, 494–495; structure of contract and teleological theories contrasted, 496
- Self-command, morality of, 419, 424
- Self-evidence, 19, 506–507
- Self-esteem, *see* Self-respect
- Self-government, value of, 205–206
- Self-interest, and finality condition, 117
- Self-protection, right of, 191–192
- Self-respect, §67:386–391; defined as most important primary good, 386; in argument for principles of justice, 155–157; effect on of utilitarianism, 157–158; in value of self-government, 205–206; as characterizing Kant's ethics, 225; associative circumstances supporting, 386–388; how related to shame and the excellences, 388–391; and morality of self-command, 391; how related to envy, 468, 469–471; in argument for priority of liberty, 477–480; basic right, and liberties of equal citizenship as basis for, 477–478; allowance for in index of expectations, 478–479; in feudal and caste systems, 479. *See also* Excellences; Shame
- Sen, A. K., 38n, 58n, 62n, 72n, 116n, 124n, 140n, 229n, 237n, 252n, 260n, 282n, 283n, 284n, 313n
- Sense of justice: defined, 41, 274–275, 442; as shown in considered judgments, 41–44; relation to strict compliance condition, 125–126; relation to love of mankind, 167, 417; public, stabilizes cooperation, 236, 295–296, 305–306, 435–436; used to define moral worth, 274–275; of majority addressed in civil disobedience, 320, 328, 339–340; how acquired at third stage, 414–416; as psychologically understandable, 416–419; capacity for

- Sense of justice (*continued*)
condition of human sociability, 433;
why stronger in justice as fairness,
436–440; Mill on, 439–440; and evolu-
tion, 440–441; capacity for as basis of
equality, 442–446; genesis of and
soundness of its dictates, 451–452;
defines the shared final end of society
as a social union of social unions, 462–
463; Freud on genesis of, 472–474;
See also next entry
- Sense of justice, good of, §86:496–505;
problem of belongs to thin theory, 350;
problem of defined 450–451, 496–499;
and obvious interpretation of, 498–
499; argument from connection with
natural attitudes, 499–500; argument
from Aristotelian principle and human
sociability, 500–501; argument from
Kantian interpretation, 501; balance of
reasons favoring, 501–502; and haz-
ards of love, 502–503; argument from
finality, 503; just conduct toward those
for whom not a good, 503–505; con-
nection with stability of justice as fair-
ness, 504–505
- Serial order, *see* Lexical order
- Shaftesbury, Lord, 20n
- Shame, 388–391; defined as injury to self-
respect, 388; natural, 389–390; moral,
390–391; as a moral feeling, con-
trasted with guilt, 391, 422, 423–424;
relation to morality of self-command,
391, 424; relation to aspects of moral-
ity and to supererogation, 424; connec-
tion with finality condition, 503. *See*
also Self-respect; Excellences
- Shand, A. F., 426n
- Sharp, F. C., 161n
- Shklar, J. N., 206n, 540n
- Sidgwick, Henry, 23n, 26, 28, 29, 79,
351n, 401; taken as representative of
classical utilitarianism, 20; on priority
problem, 36; conception of moral the-
ory, 45n; on formal justice, 51, 442n;
definition of equality of opportunity
adopted, 63n; rejected average utility,
161; conflation of persons in, 164n; his
objection to Kant, 221n, 224–225; on
time preference, 259–260; on delibera-
tive rationality, 366, 370; on psycho-
logical under standability of utilitarian-
ism, 417; hedonism in, 487, 488; on
strictness of utilitarianism in requiring
sacrifices, 501–502
- Similar cases, precept of, 208–209
- Simon, H. A., 124n, 367n
- Singer, Milton, 389n
- Slaveholder's argument, 145
- Slavery, 137, 218, 286
- Smart, J. J. C., 20n, 140n, 164n
- Smith, Adam, 20n, 49, 161n, 233, 419n,
457n, 460n
- Social contract, traditional theory of, 10–
11, 14–15, 28–29, 96–97
- Social ideal, defined, 9
- Social interdependency, facts of, 373–374
- Social minimum, 244–245, 251–252,
267, 278–280
- Social nature of humankind, 433, 458–
460, 463, 494–495
- Social union, §79:456–464; defined, 459–
460; two interpretations of circum-
stances of justice, 456–457; concept of
private society defined, 457; social na-
ture of humankind explained, 458–
460; illustrations of social union, 460–
462; well-ordered society as social
union of social unions explained, 462–
463; collective activity of justice in, a
value of community, 463; division of
labor in, 463–464
- Socialism, 235, 239–242, 247–249
- Socrates, 286
- Solow, R. M., 252n, 262n
- Sovereign, role of in stability, 211, 237–
238, 296, 435, 504–505
- Special psychologies, problem of, 124–
126, 464–465, 474
- Spiegelberg, Herbert, 86n
- Spite, 467–468, 470
- Splitting, idea of, 166–167
- Stability, inherent, 436
- Stability, of conceptions of justice:
defined, 398; knowledge of, counts
among general facts, 119; relation to
publicity condition in argument for
two principles, 154–159; relied on in
toleration of the intolerant, 192–193;
and problem of congruence, 350, 496–
503; distinguished from unchanging ba-
sic structure, 400–401; inherent stabl-
ity of and psychological laws, 436;
relative stability of and psychological
laws, 436–439; and tendency of evolu-
tion, 440–441
- Stability, of social cooperation: as social
problem, defined, 6, 434–436; role of
sovereign in maintaining, 211, 237–
238, 296, 435, 505; role of public
sense of justice in maintaining 236,
295–296, 305–306, 436; two kinds of
instability defined, 295–296, 435; in ar-
gument for duty of justice, 295–296;

- role of promises in, 304–305; role of civil disobedience in, 336–337
- Stability, relative, §76:434–441; for conceptions of justice, defined, 436; problem of explained, 398–399, 434–435; inherent stability defined and related to three psychological laws, 436; relative stability of principles of justice and utility, 436–439; Mill's view concerning not really utilitarian, 439–440; and tendency of evolution, 440–441
- Stability, of systems, 399–401
- Stace, W. T., 111n
- Standard assumptions of utilitarianism, 137–138, 189, 247, 284–285, 290, 445
- Starting places, 82–83, 85–86
- Status, desire for, 477–479
- Stein, Walter, 335n
- Steinhaus, Hugo, 74n
- Stevenson, C. L., 420n
- Stopping points, plurality of, 432, 434, 484
- Strains of commitment, 126, 153–154, 255, 371, 475
- Strawson, P. F., 113n
- Strict compliance, as formal condition, 127
- Strict compliance theory, see Ideal theory
- Strict liability, 213
- Structure of ethical theories, 21–22, 490–491, 493n, 493–496
- Subjective rationality of plans, 366, 370–371
- Supererogation, moralities of, 419, 424
- Supererogatory acts, 100–101, 167, 298–299, 385, 482–483
- Sympathy, in utilitarianism, 24, 155, 161–164, 437–438
- System of natural liberty, 57, 62–63, 65
- Tawney, R. H., 63n
- Taxation, 245–247, 249–251
- Taylor, C. C. W., 490n
- Teleological theories: defined, 21–22; intuitive appeal of, 22; may be intuitionistic, as in perfectionism, 35–36; equal liberty insecure in, 185, 289–290; basis of equality in, 445; hedonism as symptomatic drift of, 490–491; role of dominant ends in, 495–496; structure of contrasted with contract theory, 496
- Terrell, Huntington, 453n
- Thoreau, H. D., 320n, 323
- Thorpe, W. H., 378n
- Time preference, §45:259–262; defined, 259–262; in classical utilitarianism, 252–253, 262; Sidgwick on, 259–260; opposition to policies related to in constitutional regime, 261–262; as parameter for ad hoc adjustment, 262
- Time-related principles, 360–361, 369–370
- Tinbergen, Jan, 369n
- Tobin, James, 252n
- Toleration of the intolerant, 190–194
- Transfer branch, 244–245
- Transitivity, as formal condition, 116
- Treating persons as ends, 156–159, 437
- Trial by combat, 116
- Trivers, R. B., 440n
- Troeltsch, Ernst, 479n
- Tucker, A. W., 238n
- Tucker, R. C., 249n
- Tullock, Gordon, 61n, 173n
- Tussman, Joseph, 99n
- Unanimity: as formal condition, 106; as not unreasonable condition given veil of ignorance, 120–122; relation to Kant's ethics, 226; in philosophical tradition, 232–233; does not apply to thin theory of good, 392; relation to unity of the self, 494–495
- Uncertainty, choice under: knowledge of special attitudes toward excluded by veil of ignorance, 118, 149; two principles of justice and maximin rule for, 132–133; original position as case of, 134–135; use of Laplacean rule for in average utilitarianism, 146–148; absence of in classical utilitarianism and splitting, 165–166; various rules for and thin theory, 392–393
- Universality, as formal condition, 114, 160, 221
- Unjust man, 385–386
- Urmson, J. O., 20n, 351n, 352n, 357n, 419n
- Utilitarianism, average, §§27–28:139–153; defined, 139–140; preferred to classical in contract theory, 140–141; reasoning leading to and taking chances in, 141–144; slaveholder's argument for, 145; objected to as using principle of insufficient reason, 145–147; Edgeworth's objective interpretation of probability in unrealistic, 147–148; concept of probability discussed, 150–153; lack of unity of expectations in, 150–153; in argument from finality and stability, 154–159; role of sympathy in, 155, 437–438; as grounds for equal liberty, 181–182, 184–185; and embedding of ideals; use of

- Utilitarianism (*continued*)
 discounting in savings problem, 262;
 and mixed conceptions, 279; relative
 vagueness of, 281–282; as principle
 for individuals, 294–295, 298; and in-
 determinacy of right, 394–396, 494; as
 psychologically understandable, 417–
 418; relative stability of, 437–439; and
 evolution, 440–441; and basis of equal-
 ity, 445; as teleological theory, hedon-
 ism the tendency of, 490–491; Mill's
 proof of interpreted, 492–493; struc-
 ture of contrasted with contract theory,
 493–496; ostensibly less congruent
 than contract theory, 501–502
- Utilitarianism, classical, §§5–6:19–30,
 §30:160–168; defined, 20, 22–23, 139–
 140; extends principle of choice for
 one person to social choice, 21, 23–24;
 as teleological theory, 21–23; concep-
 tion of good in, 21–22; distribution in,
 23; status of precepts of justice in, 23,
 25, 269–270; impartial spectator in,
 24, 26, 161–166; conflation of persons
 in, 24, 26, 162–165; and priority of
 right, 27–29; solution to priority prob-
 lem, 36; and allocative justice, 77; rela-
 tion to imperfect procedural justice,
 77; interpersonal comparisons in, 78,
 284–285; standard assumptions of,
 137–138, 185, 247, 284–285, 445; reli-
 ance on general facts in, 137–139; con-
 sidered in argument from finality and
 stability, 154–159; role of sympathy in,
 155, 161–164; and treating persons as
 ends, 156–159; and impartial sympat-
 hetic spectator definition of right, 161–
 163; confuses impersonality with im-
 partiality, 163–166; relation to perfect
 altruism, 164–165; absence of taking
 chances in, 165; and idea of splitting,
 166–167; on savings problem, 253,
 262. *See also* Utilitarianism, average
- Values of community: problem of in jus-
 tice as fairness, 233–234, 456, 511;
 why justice in social union a case of,
 462–463; justice as case of, how re-
 lated to unanimity condition, 494–495;
 congruence confirms justice as case of,
 505
- Vanek, Jaroslav, 240n
- Veil of ignorance, §24:118–123; defined,
 11, 17, 118; knowledge of what ex-
 cluded by, 11, 16–17, 118–119, 149,
 175–176; does not obscure meaning of
 original position, 119; not irrational,
 120; effect on unanimity condition,
 120–121; excludes bargaining, 120–
 121; in savings problem, 121, 254–
 255; implicit in Kant's ethics, 121–
 122, 222; bounds on complexity of
 general facts, 122–123; with mutual
 disinterestedness compared to benevo-
 lence, 128–129; closes historical re-
 cord to parties, 160; variations of in
 four-stage sequence, 175–176; and just
 savings problem, 254–255; does not
 hold in applying principles of rational
 choice, 394
- Venturi, Franco, 254n
- Vickrey, W. S., 141n, 196n, 283n
- Viner, Jacob, 399n
- Virtues: defined, 167, 382; how distin-
 guished from natural assets, 382–383;
 as excellences, 390–391; of self-
 command, 391, 419, 424; of morality
 of authority, 408–409; of morality of
 association, 413; judicial, 453; of integ-
 rity, 455–456
- Vlachos, Georges, 11n
- Vlastos, Gregory, 9n
- Wages, fair, 268–271
- Walras, Leon, 489n
- Walzer, Michael, 99n, 330n, 335n
- Want-regarding principles, 287
- Warnock, G. F., 129n
- Warrender, Howard, 211n
- Weber, Max, 479n
- Wechsler, Herbert, 206n
- Well-ordered society, concept of,
 §69:397–405; defined, 4–5, 397–399,
 475; stability of conception of justice
 of, 192–193, 398–399; concepts of
 equilibrium and stability defined, 399–
 401; quasistability of 400n; two tradi-
 tions of moral learning, 401–404; ac-
 count of moral development in tied to
 a theory of justice, 404, 430, 434
- White, Morton, 507n
- White, R. W., 374n, 389n
- Whiteley, C. H., 97n
- Wicksell, Knut, 140, 249n
- Williams, B. A. O., 63n, 130n, 297n,
 420n, 511n
- Williams, G. C., 440n
- Wittgenstein, Ludwig, 420n, 489
- Wollheim, Richard, 180n, 210n
- Young, Michael, 91n
- Ziff, Paul, 351n, 353n, 355n
- Zinn, Howard, 320n, 322n