

Chapter IV
The Principle
of Morality

Fried, Charles.
An Anatomy of Values: Problems of
Personal and Social Choice.
Harvard University Press, Cambridge,
MA: 1970.

Certain ends are constituted by the principles which score them or give them coherence. For instance, the principles which determine a good game of chance may lend coherence to a number of different games, and the rules of a particular game in turn govern many individual matches. But these relations are complex: the more remote the principle the less explicit is the agent's reference to it. For quite remote and general principles many agents may lack altogether the sophistication needed for their articulation. Nevertheless by analyzing our beliefs and preferences — aided perhaps by the probing of an inquirer — we can be brought to articulate these principles as principles which we have always held, and where such analysis reveals inconsistency or incoherence the articulated revision is reflected in the more specific and concrete behavior and ends which are scored by these principles.

In this chapter the analysis focuses on what I shall call the principle of morality, which is the term I shall use here for a particular principle of the more remote and general sort. In the terminology of this discussion a moral end is an end which is constituted by the principle of morality. A moral action is an action scored by this principle. It is a concept which refers to a very general principle —

V
ple
ty

like the concept of a good game of skill — in that it cannot be instantiated in specific ends or actions without the elaboration of intervening principles and arguments which derive from the entailments of the principle of morality in conjunction with other material premises. Thus fidelity, for instance, is a more particular principle derived from both the principle of morality and more particular principles having to do with promises, mutual expectations, and the like. And a particular act of fidelity, say the keeping of a burdensome undertaking, which would be an instance of a rational action, would involve the application of this general principle of fidelity to the circumstances of a particular case, and these circumstances, too, might impose their own rational structure.

Morality, moral principles, principles of justice and fidelity, and the other principles related to morality have usually been viewed as constraints upon actions and the pursuits of ends. In this account I shall present them rather as rational principles giving certain actions and ends their rational structure. In other words I look at a just act or an act of fidelity to one's trust not simply as instances in which impediments are placed and accepted in the way of pursuing other ends. Rather, such acts are viewed as performed for their own sake, and thus such acts are properly viewed as ends in themselves. These acts are performed in accordance with certain complex reflexive, rational principles — the principles of morality and of justice or of fidelity — and therefore they are instances of rational actions and rational ends.

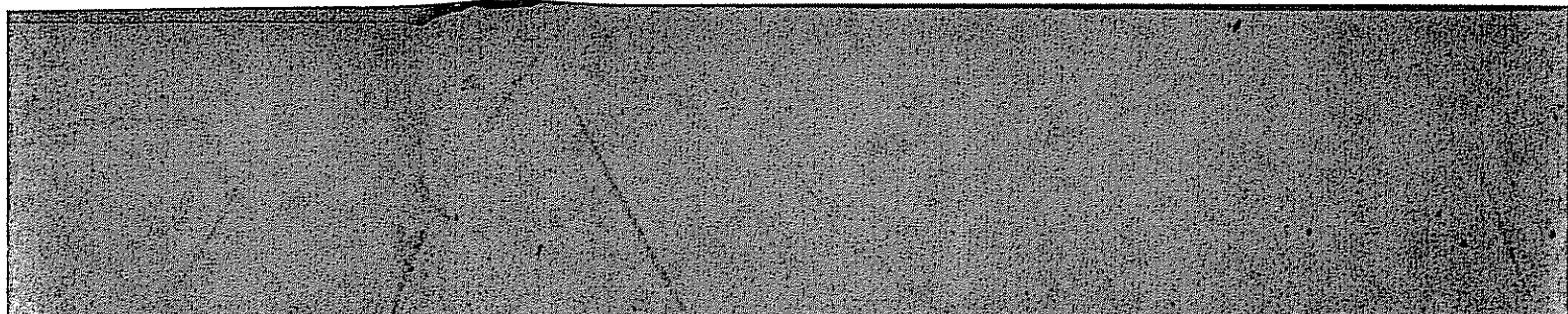
Much of what is said in these chapters about morality and its derivatives is not at all new. It is squarely in the tradition of Kantian moral philosophy. What is less familiar is the focus on the individual act as an end in itself, an end whose structure is determined by the complex of moral principles. In this way a quite definite sense is given to the notion that moral acts are done for their goodness alone.

The Principle of Morality Introduced

First, the domain in which the concept of morality (as I now define it) applies is the domain of all ends and actions which impinge in any significant way on other persons. (I shall on occasion use the term transaction to denote an action having impingements

principles which score them
; the principles which deter-
and coherence to a number of
rticular game in turn govern
lations are complex: the more
is the agent's reference to it.
les many agents may lack al-
r their articulation. Neverthe-
eferences — aided perhaps by
be brought to articulate these
ave always held, and where
r incoherence the articulated
fic and concrete behavior and
les.

on what I shall call the prin-
shall use here for a particular
neral sort. In the terminology
nd which is constituted by the
m is an action scored by this
; to a very general principle —



on other persons.) This aspect of my concept both includes and excludes aspects of what is often comprehended in the concept of morality. The concept perhaps is wider, in that it includes acts, ends, and relations of love and friendship and other situations in which obligations do not fully specify the content of the situation. It is narrower than some conceptions because it excludes acts and ends which have no significant impingements on others, but on the agent alone. Such totally "private" acts and ends are not, however, necessarily of lesser significance than those comprehended under the concept of morality; nor are they by reason of their privacy less susceptible to the kind of rational analysis I undertake in this essay as a whole.

Second, the principle which specifies the concept of morality is an expression of the concepts of equality, of impartiality, and of regard for all persons as ends in themselves. More precisely, morality requires that an action involving persons other than the agent be compatible with a principle which has the following formal properties. (1) *The interests, preferences, or desires of the agent have no special status or higher priority just because they belong to the agent; that is, an agent may not prefer his own interests as such.* (2) *The interests of no named party may be preferred simply because he is that named party.* Thus morality requires that justifications not rest in principle on first person references or on references to proper names. (3) *The interest of no party may be preferred simply because either that interest or the party having it is preferred by the agent or a named individual.* This is a perhaps obvious corollary to the first two propositions. (4) *The interest of no party to an action may be preferred except by reference to a principle which each party to the action does, or would, or should (if they know their own interest) recognize as according equal weight to his interests in the long run.* Thus, there will be situations where in conformity with morality persons may prefer their own or their families' interests, but these are situations in which all parties would or should agree that their own interests would best be served by everyone preferring his own or his family's interests.

There is, of course, an obvious problem lurking in the phrase "each party does, or would, or should recognize as according equal weight to his interests in the long run." If all parties do or, if asked, may be expected to agree that a principle is properly impartial,

concept both includes and excludes in the concept of morality, in that it includes acts, relationships and other situations in the content of the situation. Because it excludes acts and judgments on others, but on the means and ends are not, however, those comprehended under the principle by reason of their privacy less analysis I undertake in this essay

the concept of morality is an analysis of impartiality, and of reciprocity. More precisely, morality concerns persons other than the agent because the following formal properties of desires of the agent have no bearing on his own interests as such. (4) The interest of a person may be preferred simply because morality requires that justification be given in person references or on reference to the interest of no party may be preferred or the party having it is individual. This is a perhaps the most important position. (4) The interest of a person does, or would, or should be recognized as according equal weight to all interests. Thus, there will be situations in which all persons may prefer their interests. These are situations in which all persons in their own interests would best be served by their own or his family's interests. A problem lurking in the phrase "morality requires that justification be given in person references" is that we do not recognize as according equal weight to all interests. If all parties do or, if asked, a principle is properly impartial,

there is little problem, but what does it mean to say they *should* agree? If this were itself a moral standard, the criteria would be viciously circular. Although I am not prepared to give a full account of when parties should accord this recognition to a principle, I would propose that any account have the following additional formal property. (5) *It may be asserted that a person should acknowledge that his interests are recognized by a principle only if the interests, which he does or should acknowledge as his own, are interests which in some sense it is plausible to say he himself has.* This leaves open the question whether we are taking an extreme libertarian position where only those interests are taken into account which the person actually asserts for himself; or a Platonic position by which a person is asserted to have certain "real" interests, whether he acknowledges them or not; or intermediate positions which impose various rationality requirements on assertions of interest.

Finally, there is a rather technical issue raised by the notions of equality and impartiality. Propositions 1-4 have been put in terms of impermissible bases for preferring the interests of any person. This leaves open the following question as to the basis for an inequality. There are two choices open: one has the characteristic that A is preferred to B, and B is preferred to C; and the other would treat A, B, and C equally. Is it a sufficient justification for the choice of the first in spite of its inequalities, that the *average* level of the satisfaction of interest is higher in the first than it would be in the second, even though C is worse off in the first choice than in the second? In other words, is the availability of a higher average level of well-being a justification for making a particular person worse off than he would be if all were treated equally; may the interest of an individual be sacrificed for this conception of the general good?

To meet this issue, I posit a further formal requirement of the concept of morality, which goes beyond propositions 1-4. (6) *The principle of an end or action is consistent with morality only if that principle gives equal weight to the interests of each person affected by an action, except as departures from equality improve — overall and in the long run — the position of the least preferred person, in relation to what it would be in a position of absolute equality.*

This proposition derives from John Rawls's second principle of justice which states that "inequalities are arbitrary unless it is reasonable to expect that they will work out for everyone's advantage." * It is obviously related also to Kant's universal principle of right: "Every action is just [right] that in itself or in its maxim is such that the freedom of the will of each can co-exist together with the freedom of everyone in accordance with a universal law";† and to the third formulation of the categorical imperatives: "Act so that you treat humanity, whether in your own person or in that of another, always as an end and never as a means only." The present formulation differs from that of Rawls in that it, like Kant's, applies to particular actions. It is more general than Kant's principle of right because there is no commitment to the proposition that the sole significant interest in respect to which equality must be maintained is the interest in freedom.

Although the full ramifications of this concept of morality are elaborate, and some of its features must be argued for in detail before they seem plausible, I shall at this stage of the argument only make the connection with the preceding discussion of rational action and rational ends. Morality is a principle or set of principles characterized, let us say briefly, by the notions of impartiality and equality. It is a principle which is of a higher order of generality than certain other moral principles, say fidelity or trust, which in turn are of a higher order of generality than some systems of conduct built on or exemplifying trust or fidelity. Finally, particular actions are structured by such rational systems from the lowest order of abstraction to the higher.

The concept of rational ends and rational actions that I have developed is entirely formal, and an infinite array of principles may be put forward to fulfill the defined constitutive role in a rational end. Morality — as I have specified it in this chapter — is just one such principle, and nothing has been said to show that it is a particularly significant one. I shall now try to show not only the significance of the concept of morality as defined, but also that it is a satisfactory formalization of a concept which corresponds to or lies behind a whole set of concepts of generally agreed significance and potency.

* John Rawls, "Justice as Fairness," *Philosophical Review*, 67 (1958): 164.
† *Metaphysical Elements of Justice* (Ladd trans., 1965), p. 35.

Rawls's second principle of justice is arbitrary unless it is worked out for everyone's advantage. Kant's universal principle of morality is that in itself or in its maximization of each can co-exist together in accordance with a universal law";† categorical imperatives: "Act so that you use your own person or in that of others never as a means only." The principle of Rawls in that it, like Kant's, is more general than Kant's principle of morality in its commitment to the proposition that in respect to which equality of freedom is required.

This concept of morality must be argued for in detail at this stage of the argument preceding discussion of rationality as a principle or set of principles by the notions of impartiality and justice is of a higher order of generality than some systems of principles, say fidelity or trust, or the principle of general trust or fidelity. Finally, parry off such rational systems from the other.

And rational actions that I have in mind are an infinite array of principles may play a constitutive role in a rational system. It in this chapter — is just one point said to show that it is a parry to show not only the significance of the principle defined, but also that it is a point which corresponds to or lies in the generally agreed significance and

The Principle of Morality as the Most General Principle Applicable to Relations with Other Persons

In this and the succeeding section I shall argue for the proposition that the principle of morality is the most general principle applicable to ends and actions having significant impingements on other persons. This argument will not be a logical proof. In this section I shall argue: (a) our dealings and relations with others often and significantly are ends we pursue for their own sakes and they are rational ends, so that what I call transactions make up an important subclass of the class of rational ends and actions; (b) therefore there must be a rational principle or set of rational principles for these rational ends, and among such possible principles are those which have the characteristic of incorporating into one's own rational ends the rational principles and ends of the other party to the transaction; (c) such ends thus are those in which the other party to the transaction is not used merely as a means to the accomplishment of one's own ends (there are adduced reasons why principles expressive of this reciprocity are of significance); and finally (d) the principle of morality is that principle which best expresses this kind of reciprocity. In the next section I shall argue why the principle of morality should be considered not merely a significant principle for transactions but, on certain assumptions, the controlling general principle of dealings with others.

Transactions as rational ends and actions. If no rational principles whatever applied to transactions — that is, if transactions could never be counted as rational ends in themselves — then it would follow of course that the principle of morality would never apply in our dealings with other persons. The reason is that the principle of morality is a general rational principle scoring a whole family of possible particular rational ends.

Let us begin by considering this preliminary question. Do relations with others ever involve ends in themselves, or are such dealings always instrumental to the attainment of goals which are separately identifiable from the relations with others and which are only means to their attainment? If one adopted a severely unitary form of hedonism — a single distinct entity like "pleasure" is the sole basis for all value and the end of all striving — then obviously transactions would be excluded (along with everything else except

Philosophical Review, 67 (1958): 164.
trans., 1965). p. 35.

"pleasure") from the category of ultimate ends. I have developed at length the arguments against this kind of unitary theory. The more plausible view recognizes a plurality of ends.

To the general statement of this pluralistic view I would now add that one's relations with others are the ground for a significant number of ends. Conversations, competitions, amusements, love, kindness, sex, domination, submission, all of these and many others are the forms of ends we attain in relations with other persons. And as to some or all of them it would be artificial to say that these ends are not ends in themselves, but means only to the attainment of some single end (or perhaps some small numbers of ends). In our relations with others, then, we find the occasions for a large number and for most of our significant ends and values. If it is true, as I shall later argue, that our ends define us, then the most significant thing about a man is the ends he chooses to and can attain, and then also we can see the sense in saying that man is a social animal: among his ultimate ends, those involving others occupy a prominent place.

But are these ends rational ends — ever, often, always? I think we may establish quite easily that ends involving others are primarily rational ends. When an end requires a certain kind of recognition of an outside object involved in that end, there is already present in implicit or inchoate form a rational principle. Does that mean that a dog chewing a bone or a sea anemone closing upon a speck of plankton performs a rational action and attains a rational end? Certainly not — but it may mean that for men even such "instinctual" activities as sex should be viewed as rational in the special sense I have defined. For the sea anemone and the dog the recognition of its object is not a rational recognition, because the process of rational analysis — of bringing to the surface the features of the object of the action — will not in any way illuminate or alter either the agent's conception of his end or his actions in respect to it. But with human ends involving other persons this rational recognition will frequently be present. Nor is this recognition purely instrumental. If a person is assembling a tricycle, then an understanding of the nature of his tools or of the parts he is operating on may of course be helpful, but it is not a better understanding of his *end*. He does not care how the tricycle gets

nate ends. I have developed a kind of unitary theory. The plurality of ends.

pluralistic view I would now are the ground for a significance, competitions, amusements, passion, all of these and many in relations with other people it would be artificial to say for ourselves, but means only to the perhaps some small numbers of men, we find the occasions for significant ends and values. If our ends define us, then the ends he chooses to and the sense in saying that man ends, those involving others

—ever, often, always? I think ends involving others are pre-requires a certain kind of recognition in that end, there is already a rational principle. Does that a sea anemone closing upon a rational action and attains a ray mean that for men even could be viewed as rational in the sea anemone and the dog a rational recognition, because of bringing to the surface the — will not in any way illumination of his end or his actions involving other persons this be present. Nor is this recognition person is assembling a tricycle, of his tools or of the parts helpful, but it is not a better not care how the tricycle gets

assembled, so long as it does — unless, of course, he is assembling the tricycle as a puzzle. But where the end essentially involves another person as part of its very structure, then to understand the other person, to recognize what sort of "thing" the other person is, will entail a better understanding of the end itself.

Now let us consider a bizarre example: a cannibal who wishes to prepare and consume a very special dish of human flesh. Does this example bear out the point I have been making? I think it does. Assuming that cannibalism has no ritual significance for our agent (that makes a more complex case) but only a culinary significance, if the cannibal is a true epicure he will be as potentially interested in the nature of his ingredients as Escoffier might have been in the coagulent properties of egg yolk in *sauce béarnaise*. Or to take a less bizarre example, a person who is doing anatomical sketches will have an interest in and manifest a recognition of the lines and form of the human body that I would call rational. The reason that cases of eating, drinking, and sex might be considered borderline cases is that it has been often asserted for all of them that the agent may be utterly unconcerned and unilluminated by a further understanding of his objects. He is simply driven instinctually to their pursuit, and reason is relevant only instrumentally, that is, only in telling him how to "get" his objects. I do not wish to argue that there are no such instinctual ends for men, or that ends involving other persons are necessarily never wholly instinctual. Perhaps acts in the realm of sex, aggression, or maternal care are sometimes wholly instinctual in the sense that recognition is merely the sensory apprehension of an object meeting some instinctually imprinted archetype. I only say that I doubt very much that such an instinctual analysis is *ever* wholly appropriate to ends involving others — "enemy" is a concept made up of significant conventional and rational elements, as is "appropriate sex object" or "object of maternal affection." But whether such ends are ever wholly instinctual or not, at least as regards persons of any maturity and civilization it is far more plausible to build on the assumption that most, if not all, of their ends involving others are rational ends.*

* Once again, I decline to offer a proof for this proposition that transactions involve rational principles, just as I declined in previous chapters to offer a

The rational principles of transactions: recognition and reciprocity. I have stated that transactions will involve significantly the kinds of recognition of their objects — that is, other persons — that is characteristic of rational principles and ends. Can anything more be said about the content of the rational principles for transactions? The examples I have just adduced are cannibalism, sex, figure drawing, and acts of aggression. In previous chapters, I have used examples such as acts of restoration and respect, and competitive or cooperative games. What are the principles of recognition, the rational principles, in these cases?

We can begin to specify by asking in each case what is the significance for the end being pursued that it requires a human object. In the case of (nonritual) cannibalism the significance resides in whatever special qualities of taste or texture there may be in human flesh, and this suggests that if some other kind of flesh had similar (or superior) qualities, then a human being would no longer be necessary to that end. Similarly in the case of the anatomical sketch, if what is desired is a rendering of strong musculature perhaps a human subject is also quite adventitious to the end. Contrast to such cases that of ritual cannibalism or — what may be surprisingly analogous — the case of an artist who is interested not in bodies and musculature as such, but in the human body. The ritual cannibal draws significance from his act just because it is human flesh. Transactions, then, may involve a recognition of various aspects of their objects — their taste (in the most bizarre example), their appearance — but the major significant recognition is of the humanity of the object. And so also in acts of kindness, hostility, love, sex, or play.* What does it mean to recognize the humanity of the object of one's end?

In general, to recognize the humanity of one's object is to recognize whatever is distinctive about a human object compared to other possible objects. One might wish to include features such as shape and size (I will now leave taste behind for good and all,

proof for the existence of rational ends and principles in general. I have said only so much as is necessary to give a sense of what I mean by these notions, leaving it to the reader to decide for himself whether they illuminate the experience to which they apply.

* Playing a game against a computer may be fun, but it is not the same as playing against a human opponent, unless one believes he is playing with whoever built and programmed the computer.

actions: recognition and reciprocity will involve significantly the other persons — principles and ends. Can anything rational principles for transduced are cannibalism, sex,

In previous chapters, I have action and respect, and compare the principles of recognition cases?

in each case what is the significance that it requires a human object — that the significance resides in the texture there may be in some other kind of flesh had human being would no longer in the case of the anatomical presence of strong musculature perceptible to the end. Cannibalism or — what may be an artist who is interested not but in the human body. The reason for his act just because it may involve a recognition of aesthetic taste (in the most bizarre case) the major significant recognition would so also in acts of kindness, does it mean to recognize the

recognition of one's object is to recognize a human object compared to other objects — that is, to wish to include features such as aesthetic taste behind for good and all,

and principles in general. I have said something of what I mean by these notions, myself whether they illuminate the ex-

perience may be fun, but it is not the same as one believes he is playing with who-

while adding the availability of the person's tissues and organs for study and transplantation), but the most important characteristic is the other's capacity for behavior and performances of his own. This capacity ranges from elementary reactions like jumping when burned or frightened to complex intellectual performances like performing a difficult theoretical task, joining in a musical improvisation, or cooperating in the governance of a polity. The most distinctive and unique human qualities and capacities for our purposes is the general capacity — exhibited in an infinity of particular acts and capacities — to pursue rational ends, to follow rational principles, and to engage in rational actions. Thus for instance, to join other players in a musical improvisation is not only to perform a difficult intellectual task, nor indeed is it only to perform that task as an end in itself — a rational end: it is to pursue a rational end which has as a part of its constituting rational principle the rational actions of others, and those rational actions and ends of the others have as part of their rational principles the same or analogous recognitions of the rational principle of the agent. These rational ends and actions I shall therefore call reciprocal: reciprocal ends, actions, and principles.

Before indicating why reciprocal ends and principles are so important, I shall contrast them to cases that are close to reciprocity, but yet are not truly reciprocal. Consider the case of intelligent compliance with a complex demand exacted under threat of extreme and immediate violence. The threatener certainly assumes the existence of significant human capacities: The capacity to understand and comply with the demand and the capacity to understand and respond to the threat. But this transaction does not exhibit what I have just defined as reciprocity because, although the threatener attains an end of his own, the only end of the victim's that is attained is the avoidance of the threatened consequence. Furthermore, the end of the threatener is not usually necessarily one that involves the participation of a threatened human agent, unless exercising dominance over another human agent is itself all or part of the threatener's ultimate end. Except for that special case, both the threat and its victim are mere instruments, standing outside the conception of whatever end the threatener seeks to attain: for example, having the victim assemble a tricycle for him. A bribe would exhibit similar characteristics, contrasting

it to the case of reciprocal pursuit of ends. In such cases, then, there are two reasons why there is no reciprocal rational action and end: (1) the agent does not use another person as a necessary constituting element of his (the agent's) end, but only as an instrument for attaining a separately identifiable end, and (2) the action of the other party to the transaction is not a necessary constituting element of that other party's end. Let us now consider cases in which reciprocity is lacking in one, but not both, of these respects.

A most interesting case of near reciprocity is what I shall call the case of perversion: the actor uses another person to attain his end, and it is a necessary constituting element of that end that another person be used (that is, the other person is not just a contingent instrumentality), but it is also a necessary element of the actor's rational principle that the other person thereby *not* attain an end of his own. Cases in which it is an essential part of the actor's rational end to exert power over another person or to inflict pain illustrate such a perverse rational principle. The principle (1) recognizes the other as an essential part of the end, (2) recognizes as an essential part of the plan that the other is capable of pursuing rational ends of his own, and (3) makes it an essential part of the plan to interfere with those ends.* I would add that in its most developed form this kind of perversion requires (4) that the victim know that it is the actor's end that his (the victim's) end be frustrated. (As truly reciprocal ends have the quality of two mirrors endlessly reflecting each other, these perverse cases may be called cases of the sinister mirror.)

The less sinister departure from perfect reciprocity is the one where the actor uses the other as a mere instrument of his end, but in so doing is indifferent whether the other person attains an end of his own. Such cases are far less dramatic. I need a task performed, and I can get it done cheapest and best if I pay you. The reason you accept my pay is that it allows you to do something which is an end of your own. I wish to have my watch repaired, and pay you to do it. You love repairing watches, and only by being paid can you afford to spend your time that way. But I do not care *why*

* Intermediate cases of cruelty and perversion are those in which it is of the essence of the actor's end to inflict pain on a nonhuman agent, one that does not have the capacity to pursue rational ends. Similarly, it can be part of an actor's rational end out of some sort of kindness to allow a nonrational creature to attain his nonrational end.

nds. In such cases, then, there is a local rational action and end: the person as a necessary constituent, but only as an instrument for the end, and (2) the action of the person as a necessary constituting element. I now consider cases in which both, of these respects.

Reciprocity is what I shall call the principle that another person to attain his end, the instrument of that end that another person is not just a contingent necessary element of the actor's end, but that the other person thereby *not* attain an end which is an essential part of the actor's end, rather than to inflict pain on the other person. The principle (1) recognizes that the other is capable of pursuing the end, and (2) recognizes that the other is capable of pursuing the end. I would add that in its most general form it requires (4) that the victim's end be frustrated. These are the quality of two mirrors in which these perverse cases may be called

perfect reciprocity is the one in which the other person is a mere instrument of his end, but in which the other person attains an end of his own. I need a task performed, and I need it best if I pay you. The reason I pay you is that you will have my watch repaired, and pay me for my watches, and only by being paid in that way. But I do not care *why* you do it. Perverse cases are those in which it is of the kind that is done by a nonhuman agent, one that does not have ends. Similarly, it can be part of an agent's kindness to allow a nonrational creature

you choose to repair my watch, and would let someone else who hated the work or let a machine do it if it could be done cheaper or better. If the departure from reciprocity in the perverse case is repellent it is because of the acknowledgment of the victim's human capacities; in this case any repellent quality derives from indifference to those same capacities.

I have now introduced and elucidated to some extent certain forms of principles that may obtain in respect to dealings with other persons. Principles involving what I call reciprocity are among those — they are candidates, they are possible principles of ends and actions. Moreover, I believe that we may assume even at this stage of the argument that some degree of reciprocity is an important aspect of some significant human ends and relations. What needs to be considered now is whether recognition of others and reciprocity have any special significance among all the possible principles in dealings with others, and what that significance is. My argument is that the capacity to entertain rational ends and principles is of peculiar significance, and that for that reason the principle of reciprocity has a peculiar significance.

I have stated frequently in this essay that our ends define us. This is not a point I shall try to prove. It seems to me implicit in the concept of an end that it is our ends that move us and express our values. The concepts of striving, preference, satisfaction, and value all are related to and depend on the concept of end. Nothing is more important to a person's understanding of himself than an understanding of his own ends. Indeed, it is an analytic truth in this system that a man's ends are the locus of significance for him. Now we come to a person's understanding and recognition of *other* persons. By the concept of reciprocity a person recognizes another person from the same perspective as he sees himself, and from the same perspective as that other person sees himself — that is, in terms of that other person's ends. If for any two persons, A and B, the most important aspect of A for A is A's ends, and of B for B is B's ends, then by the principle of reciprocity the most important aspect of B for A is B's ends, and vice versa. Thus each person recognizes as most important in others just what he recognizes as most important in himself, and just what they recognize as most important in themselves.

Now beyond ends as such, I have specified further the category

of rational ends. I must now make out the claim that there is a special significance to viewing oneself and others in terms of *rational* ends. Once again, I shall not attempt anything like a proof. The claim shall rest on three arguments. First, most human ends can be, and are more plausibly, conceived of as having a rational element. This argument is made in the preceding chapters. Second, human ends involving others are particularly likely to have a rational element, this being the element of recognizing the nature and capacities of those other persons. Finally, there is an argument that is considered in Chapter Three and in Chapter Six: the general capacity to entertain rational ends and principles corresponds to a tendency for order and simplicity to be created out of the multiplicity of elements of desire, observation, and need a person encounters. Together these three arguments sufficiently support the hypothesis that persons most appropriately regard themselves as entities having not only ends, but rational ends; and not only as having rational ends, but as having rational ends in respect to their dealings with other persons.

Reciprocity, therefore, is significant because when pursuing ends involving other persons, it is only by scoring those ends by principles consistent with—or expressive of—reciprocity that a human agent recognizes another in the terms that the agent recognizes himself and in the terms that that other person also recognizes.

Reciprocity and the principle of morality. It remains to demonstrate that the principle I have called the principle of morality is the principle expressing the notion of reciprocity.

Reciprocity is the recognition of the other participants in a transaction as entities having ends and rational ends. That recognition is not just a formality, a brief concession preceding the working out of the elements of the end itself. The recognition of this quality—I shall call it human personality, or personality—must be part of the structure of the end itself. The recognition of personality must, therefore, be part of the ordering principle of the end. We are seeking now a principle of ordering of rational actions and ends that expresses this recognition. Such a principle will be a very general one, since it must express this reciprocity in all conceivable forms of transactions. For this same reason such a principle

out the claim that there is a
 If and others in terms of ra-
 ttempt anything like a proof.
 ents. First, most human ends
 eived of as having a rational
 e preceding chapters. Second,
 rticularly likely to have a ra-
 ent of recognizing the nature
 is. Finally, there is an argu-

Three and in Chapter Six:
 ional ends and principles cor-
 d simplicity to be created out
 esire, observation, and need a
 ree arguments sufficiently sup-
 st appropriately regard them-
 ds, but rational ends; and not
 aving rational ends in respect

at because when pursuing ends
 y scoring those ends by prin-
 ve of—reciprocity that a hu-
 e terms that the agent recog-
 that other person also recog-

norality. It remains to demon-
 ed the principle of morality is
 of reciprocity.

re other participants in a tran-
 sational ends. That recognition
 cession preceding the working
 . The recognition of this qual-
 ity, or personality—must be
 elf. The recognition of person-
 ordering principle of the end.
 ordering of rational actions and
 Such a principle will be a very
 this reciprocity in all conceiv-
 s same reason such a principle

will not constitute a complete ordering of any particular end or
 action. A concrete end will be scored by much more detailed and
 particularized principles. All that is required is that such a detailed,
 particular principle conform to the general form of the general,
 partial ordering expressive of reciprocity and recognition of per-
 sonality.

The principle of morality as detailed above is the *general* prin-
 ciple expressive of the *general* recognition of human personality
 (defined as the characteristic of having rational ends) in any deal-
 ing with other persons. To summarize, the principle of morality
 accomplishes this recognition by requiring that the most general
 principle of transactions place all persons in a position of parity.
 In this way the equality of all persons at the most general level is
 the starting point for any more particular principle. Why does this
 equality or impartiality of the principle of morality express recog-
 nition of personality? My thesis is: all persons are alike in respect
 to the characteristic that they conceive of themselves as entities
 having ends and rational ends; and any essential preference be-
 tween persons entails a violation of reciprocity in the direction of
 using the other person as an instrument.

Consider the case where the principle of morality is not met be-
 cause the actor assumes in his rational principle that his (the ac-
 tor's) interests or ends are to be given preference for the reason
 that they are *his* interests. Let us call this the egoist principle.
 Clearly such a principle entails the assumption that all other per-
 sons stand in the relation of instruments to the actor's ends, since
 in the particular end the other person is not realizing his (the other
 person's) ends, while the actor realizes his own end. Consider, then,
 a more plausible general principle for transactions: that the actor
 may prefer his own ends or the ends of some other person not
 because of the identity of the person whose ends are preferred, but
 because of the nature of the end—that is, whoever's end it is, that
 end is entitled to preference because it is an end of some particular
 sort. But a principle permitting a preference of this sort also vio-
 lates reciprocity since it allows the person whose ends are disfa-
 vored by the principle to be used as an instrument (in the above
 sense) for the favored ends.

Finally, as a third alternative to the principle of morality, con-
 sider the principle by which a person's ends could be preferred if

by so doing the ends of a larger number of persons or a larger number of ends would be attained. This (let us call it utilitarian) principle also violates reciprocity, since it allows the disfavored person to be used as the instrument for the attainment of the "greater good" — as measured in terms of the number of persons benefited or the number of ends attained.

There may be other principles besides these for transactions, but if these are the chief alternatives — as I believe they are — then the principle of morality is the only principle that completely expresses reciprocity, and it does so by requiring that the interest of no party to a transaction may be preferred except by reference to a principle which each party could recognize as according equal weight to his interests in the long run.

Now the principle of morality is sufficiently general and abstract that it allows for the derivation of more particular principles in terms of which in a particular case the ends of all persons will not appear to be given equal weight. Some of these derivative principles will be considered below. To anticipate that discussion, consider the case of a father who takes his son fishing one Sunday, and thus misses the opportunity to take some one else's boy on such an expedition. It may appear that such a transaction is not in accordance with the principle of morality, since it prefers a designated person. But if the more particular principle of that transaction refers to the relation of any person with his own child, and if — as I believe may be the case — such a more particular principle regarding the relations of parent to child can be shown to be consistent with the principle of morality, then what appears to be a breach of reciprocity is in reality not. Or take the even clearer case of a person who restores a stolen umbrella to its rightful owner during a thundershower. The actor in that case appears to be treating the owner differently from the way he is treating other persons in need of the umbrella — should he not at least hold a lottery? But if we see his action as scored by a more particular principle about returning unjustly acquired property, the claims of the others that they are being unequally treated may be defeated. For the more particular principle again may be derivable from or compatible with the general principle of morality. Or finally consider the case by which a person discharges his obligations out of fidelity to an office or an undertaking. If as a judge he finds for one

er of persons or a larger number of persons (let us call it utilitarian) principle allows the disfavored person the attainment of the "greater number of persons benefited

des these for transactions, but as I believe they are — then principle that completely excluding that the interest of preferred except by reference to recognize as according equal

efficiently general and abstract more particular principles in the ends of all persons will not some of these derivative principles anticipate that discussion, consider his son fishing one Sunday, take some one else's boy on

that such a transaction is not of morality, since it prefers a particular principle of that transaction person with his own child, case — such a more particular parent to child can be shown of morality, then what appears reality not. Or take the even a stolen umbrella to its right-

The actor in that case appears y from the way he is treating la — should he not at least hold as scored by a more particular

acquired property, the claims equally treated may be defeated. gain may be derivable from or ple of morality. Or finally discharges his obligations out of g. If as a judge he finds for one

party as against another, he acts in accordance with morality, for morality requires fidelity to one's office, and the facts of the dispute require a particular decision in fidelity to that office.

Personality and respect. Two terms I have introduced in the foregoing argument must be defined formally: personality and respect. Personality is the aspect of a human person which is made up of his capacity to entertain rational principles, actions, and ends, of his actual disposition to do so, and of the particular dispositions to entertain particular principles, actions, and ends. In short, a man's personality is a function of his rational ends and principles — both potential and actual. Respect is the disposition to entertain rational principles in accordance with the principle of morality — that is, rational principles which treat other persons implicated in them as ends rather than means. These two terms, and particularly the term respect, are drawn from Kantian moral philosophy, and the fact that they fit so well into this analysis shows how close to Kant these arguments are.

Why Should We Be Moral?

The question why should we be moral is used to raise an issue which was avoided in the preceding discussion of the principle of morality. In that discussion no more was done than: (1) to introduce morality as a general rational principle applicable to any end and action involving other persons; and (2) to show how, via the concepts of reciprocity, personality, and respect, the principle of morality is a concept of peculiar significance for actions and ends involving others. But the usual conception of morality accords to it the status of a categorical principle. It is not just one principle among many, nor just a peculiarly significant principle. How does the analysis I have been developing allow us to show that morality is not only a possible and even a plausible and appealing general principle for transactions, but the mandatory general principle with which any more particular principles and ends must be consistent?

It would seem that the tack I have been taking will make it particularly hard to accord such a status to morality. The focus of this essay has been on particular acts and ends (although I have shown

that particular ends are scored by principles, some of which can apply to whole families of other ends). Thus by showing that moral ends — that is, actions and ends scored by and consistent with the principle of morality — are ends persons have on particular occasions, the particularity of morality, its concreteness, is brought to light. That is a gain. The cost is that there seems to be nothing to require that a person always act morally, or at least that his ends always be consistent with the principle of morality. I have shown how morality can be an end in itself and on a particular occasion; but why exclude from the repertoire of human ends particular acts scored by inconsistent principles like the egoist principle, the utilitarian principle, or the perverse principle? These questions are hard to answer. They are hard questions not only because of this essay's focus on particular concrete ends, but also because of the evident fact that with rare exceptions even the most moral of men sometimes act on principles inconsistent with morality. So if I come up with an argument for the priority of the principle of morality, I will be putting forth a norm that departs from the almost universal, observed behavior of men. On the other hand, the belief that if there is a principle of morality it does have priority is almost as universal as the behavior that violates this conviction. In spite of these difficulties, I shall attempt to account for the priority of morality in terms of the focus and concepts of this essay.

The man who is not moral. What if a man chose never to score his ends by the principle of morality? He would never choose ends exhibiting reciprocity with others. But so what? There is no categorical answer to this question. One can only point to the system of particular ends that are controlled by morality and reciprocity, and ask, "Do you or do you not want to be a man who excludes from his repertoire of ends all of that?" The challenge is a serious one, for as I have and will argue, we *are* our ends, and so in asking that question, one asks, "What kind of man do you choose to be?"

What, then, are the ends from which such a man would be excluded? In general, they are all actions and ends in which a person treats another as an entity, the realization of whose ends is in principle of equal importance to the realization of his own ends. Such a man would be excluded from all ends recognizing the personality of others, from all ends exhibiting respect for another

principles, some of which can be shown to be consistent with the moral principles that we have on particular occasions. The concreteness, is brought to light by showing that his ends are not only rational, or at least that his ends are rational, but also that they are rational in the sense of morality. I have shown that the principle of morality is not only consistent with the moral principles of human ends particular acts of morality, but also that it is consistent with the egoist principle, the utilitarian principle, the principle of reciprocity, and the principle of justice. These questions are raised not only because of the moral principles of human ends, but also because of the moral principles of human ends even the most moral of men are not consistent with morality. So if I come to the principle of morality, I find that it is consistent with the moral principles of human ends, but it is not consistent with the moral principles of human ends. On the other hand, the belief that the principle of morality does not have priority is at least in part a violation of this conviction. In order to account for the priority of the moral principles of human ends, I must show that the moral principles of human ends are consistent with the moral principles of human ends.

If a man chose never to score, he would never choose ends that are not consistent with the moral principles of human ends. But so what? There is no categorical imperative that can only point to the system of moral principles that is consistent with morality and reciprocity, but that is not to be a man who excludes others. The challenge is a serious one: "The challenge is a serious one: are our ends, and so in asking, 'What kind of man do you choose to be?'"

Which such a man would be excluded from the moral principles of human ends and ends in which a person is not consistent with the moral principles of human ends is in the realization of his own ends. In all ends recognizing the principle of reciprocity, exhibiting respect for another

person. In particular, he would not perform acts of justice, of generosity, of trust, of faithfulness, of love, or of friendship. This is a strong assertion. I shall support it in respect to justice in the next section, and in respect to love, trust, and friendship in the next chapter. My thesis is that all of these, justice, generosity, trust, faithfulness, love, and friendship, are concepts that relate to particular rational ends, and that the principles of each of them are concrete derivatives — with the addition of various additional premises — from the most general principle of morality. To put this point differently, all of these have in common the feature that they depend on respect for the personality of the other person involved. Moreover, to pursue these ends is to stand in a certain relation to the other persons implicated in them — relations of love, trust, and so on. And finally, to stand in such relations is a necessary condition of experiencing certain emotions, for example, guilt, resentment, love, trust. These matters are all gone into in detail in the next chapter. For present purposes it is important only to assert that feeling love, friendship, or guilt depends on standing in relations of love, friendship, or justice to others, and it is a necessary condition for such relations that one be disposed to entertain ends exhibiting respect for other persons. He who cannot respect others (in the technical sense I use here) cannot love, trust, feel guilt or resentment.

My answer, therefore, to the question, "Why should a man *ever* be moral?" is that a moral person, by virtue of having certain ends, stands in certain relationships to others and feels certain emotions. He is a certain kind of man. And the consequences of never being moral are that one is another, quite different sort of person. Then the question is, does one want to be excluded from these ends, relations, and emotions. This is all I shall say now about why men act on the principle of morality.

Must we always be moral? So it seems there may be good reason to accept moral ends not only as possible ends but as ends which a man would on some occasions choose to attain. But this is not enough. The concept of morality that we usually have is categorical — one must *always* act consistently with it. It is not like a game that is such fun that most men would choose to play it from time to time.

The chief reason that morality as a rational principle makes a

claim to absolute priority over other principles derives from the content of the principle itself. Unlike some other rational principles applicable to some other rational ends, it is universal in its very statement. It states that other persons are to be recognized in a certain way, so that every action involving dealings with others is either in accordance with the principle or in violation of it. Indeed, it is universal in the sense that even if an end does not involve others at all, the very failure to involve others may either be in accordance with or in violation of the principle of morality. A person who sits at home and does a mathematical problem may be entitled to ignore others in this way or not, depending on the circumstances; and that is a question to be resolved by reference to the principle of morality and some one or more of its derivative principles — that is, sometimes it is my moral right, as it is the moral right of everyone, to ignore others, and sometimes it is not. Thus the principle itself applies in some way universally, if only by way of permission to ignore others.

Granted the principle is universal, why must a man universally comply with it? Might not a man at one time act in a way that implies acceptance of morality and another time not act in that way? Of course, a man is physically able to do this, and all men do. What does this mean? It means that at the time he acts morally he acknowledges the principle of morality, and he acknowledges its universality. For it is of the essence of the principle that if its universality is not acknowledged, the principle itself is not acknowledged. Consider the case of a person who restores unjustly acquired property. For this to be an act of restitution — an act in accordance with the principle of morality and, more specifically, justice — it is not sufficient for the actor just to transfer physically the property. He must also acknowledge some principle about restitution. What can that principle be? Can it be that one restores unjustly acquired property: (a) if one no longer needs it, (b) if one feels like it, (c) once in a while, (d) if it was acquired during a full moon? Obviously, none of these will do, and the reason they will not is because they lack the appropriate universality; *not* because they admit of exceptions. So also a person who shows respect to another not only bows, or smiles, or tips his hat, but he acknowledges something about that other person and his own relations to him. What is it that he acknowledges? That he will tip his hat

r principles derives from the
 e some other rational princi-
 al ends, it is universal in its
 ersons are to be recognized in
 nvolving dealings with others
 inciple or in violation of it.
 that even if an end does not
 : to involve others may either
 1 of the principle of morality.
 a mathematical problem may
 way or not, depending on the
 n to be resolved by reference
 e one or more of its derivative
 my moral right, as it is the
 thers, and sometimes it is not.
 some way universally, if only
 s.

, why must a man universally
 one time act in a way that im-
 ther time not act in that way?
 : to do this, and all men do.
 et at the time he acts morally
 rality, and he acknowledges its
 of the principle that if its uni-
 rinciple itself is not acknowl-
 who restores unjustly acquired
 restitution — an act in accord-
 and, more specifically, justice
 just to transfer physically the
 e some principle about restitu-
 can it be that one restores un-
 no longer needs it, (b) if one
 l) if it was acquired during a
 e will do, and the reason they
 appropriate universality; *not* be-
 lso a person who shows respect
 or tips his hat, but he acknowl-
 erson and his own relations to
 lges? That he will tip his hat

when the spirit moves him, that if he has pleasant thoughts about the other person when he encounters him, he will express them? None of these will do, again, because the very concept of showing respect is a universal one. And in general, the actions and ends deriving from morality require the acknowledgment at least of some principles whose very statement is universal. One simply does not recognize the personality of another unless on the occasion where it is recognized it is recognized as being universally owed respect. For a man to love another person he must at least once recognize that other person as a person deserving reciprocity — he must recognize that person's humanity. But that recognition is a recognition that the other person is owed reciprocity, owed recognition as a person generally, universally.

If so much is accepted, there is still left unanswered the following question: what of the person who on one occasion acts morally — fully, truly, sincerely respects another, recognizes his personality — and does so with a full acknowledgment of the universality of those concepts, but on another occasion acts in a way inconsistent with these principles and acknowledgments? It must be clear that such inconsistency not only is possible but is exemplified to some extent in almost everybody's behavior. Does this then mean that common understanding is mistaken in according to morality a universal and categorical status? Not at all. What follows only is that persons are not perfectly — perhaps not even approximately — consistent in their principles. To recognize morality is to acknowledge its universality, but such an acknowledgment does not compel one to act accordingly thereafter. First, a person can always change his mind, and conclude that he prefers or accepts a different universal principle from the one he accepted before. Second, one can accept a principle and simply fail to act on it — this is the problem of *akrasia*, weakness of will. Both these reasons for violating the principle of morality once one has acknowledged it are related to familiar and difficult philosophical problems. For the purposes of this argument it is not necessary to solve them here.

Take the case of the person who accepts the principle of morality and then changes his mind, choosing instead the perverse principle. The point still holds that we are what our ends are — we are just, loving, trusting persons if our ends are scored by those

principles. So the man who changes his principle is a man who himself has radically changed, become a new kind of person. Whereas he once was capable of entering into relations of love and trust, capable of feeling love, guilt, and resentment, he is no longer. A somewhat more complicated version of this is the case where a person does not just make a great change, does not just suffer a conversion, but instead shifts back and forth without coming to rest. If such vacillations are extreme enough, one does not know what to say about him — what his ends are, what his values are, whether he is a person capable of friendship; indeed, one does not know who he is at all.

The second reason, weakness of will, can be treated analogously. Our conception of ourselves and others as trusting and trustworthy or loving (or what have you) persons allows for certain lapses and shortcomings, but after a point such lapses make it unclear whether a person is moral but extraordinarily weak-willed, or whether in fact he really does not subscribe to morality at all. In either case he can no longer enter into reciprocal relations of love, trust, or friendship.

Finally, I should like to anticipate very briefly what shall be discussed at length in Chapter Six: consistency as an end or general principle in itself. My thesis is that one significant human disposition is to seek consistency as an end in itself. And this disposition to order our particular ends into consistent wholes — a kind of highest order rational principle and end — would entail a consistent working out of the various principles one adopted for his significant ends. This would provide an additional ground for acting in a way that recognized over time the priority and categorical character of the principle of morality. To fail to do so would be to introduce significant inconsistency over time in the general ensemble of ends one pursued and principles one accepted.

And so to the question why should we be moral, I would answer that no obligation to be moral can be adduced. One either is or is not moral, with all the ensuing consequences for the individual and his relations. But if one is moral, then categorical obligations relevant to all one's ends arise. Obligation and its categorical quality arise out of the concept of morality; they do not precede

principle is a man who him-
 self kind of person. Whereas
 relations of love and trust,
 commitment, he is no longer. A
 this is the case where a per-
 son, does not just suffer a con-
 straint without coming to rest.
 In fact, one does not know what
 what his values are, whether
 indeed, one does not know

can be treated analogously.
 as trusting and trustworthy
 allows for certain lapses and
 lapses make it unclear whether
 weak-willed, or whether in
 ability at all. In either case he
 relations of love, trust, or

every briefly what shall be dis-
 tinguished as an end or general
 the significant human disposi-
 tion itself. And this disposition
 consistent wholes — a kind of
 end — would entail a con-
 ceptions one adopted for his
 an additional ground for
 time the priority and cate-
 gory-morality. To fail to do so
 consistency over time in the
 and principles one accepted.

we be moral, I would answer
 adduced. One either is or is
 sequences for the individual
 then categorical obligations
 obligation and its categorical
 morality; they do not precede

it. If one asks "why should I be moral," the answer is "you are
 under no obligation to be moral, but if you are moral, these and
 these are your obligations."

Justice and the Obligations of Fairness. The Rawlsian System

A whole family of principles derives from the most general prin-
 ciple of morality: the principles of justice, which refer to the struc-
 ture of institutions and practices, and the principles of fairness,
 which relate to the obligations of individuals involved in more or
 less formal institutions and in practices such as promising. This
 family of principles is important because it illustrates a part of the
 process of particularization of the general principle of morality.
 Without such particularization that most general principle can
 never apply to particular ends and actions. These do not represent
 the complete system of principles derivable from morality. There
 are more specific principles regarding natural obligations, such as
 those arising out of family relationships, and principles about in-
 juries to others apart from institutional contexts. Analysis of the
 principles of justice and fairness should indicate, however, how the
 working out of such principles might go.

The principles and arguments I present here are drawn directly
 and explicitly from the work of John Rawls. I should say that I will
 not at all times stay with Rawls's precise terminology, in part be-
 cause that terminology is in process of development and in part
 because some simplification (and perhaps distortion) is necessary in
 a brief account of his often very complex argument.

Justice and the original position. Rawls defines justice as that
 concept that applies to the design of institutions and practices inso-
 far as they assign to persons and classes of persons benefits and
 burdens, offices, privileges, liabilities, and the like. He puts forward,
 as specifying this concept, his two principles of justice: "first, each
 person participating in a practice, or affected by it, has an equal
 right to the most extensive liberty compatible with a like liberty for
 all; and second, inequalities are arbitrary unless it is reasonable to
 expect that they will work out for everyone's advantage, and pro-
 vided the positions and offices to which they attach, or from which

they may be gained, are open to all. These principles express justice as a complex of three ideas: liberty, equality, and reward for services contributing to the common good." *

The first principle requires maximum equal liberty. This means that equal liberty is not sufficient if everyone could have more liberty. The second principle, that departures from equality are permissible if they work out to everyone's advantage, has been specified by Rawls as follows: a departure from equality of distribution (and I shall say in a moment what goods are being distributed) is permissible only if the result of such an inequality is to make the worst-off person better off than he would be in a situation of pure equality. The first principle of equal maximum liberty applies primarily to the rights of citizenship, political rights, civil liberties, and religious liberties, and entails guarantees of the suffrage, due process of law, freedom of speech and thought, and freedom of religion. The second principle applies primarily to what Rawls calls distributive shares, or the distribution of economic and related benefits and burdens in society. It should be noted that Rawls believes that the two principles of justice, applying as they do to the design of institutions, refer in the first instance to the liberties and the social and economic position of classes and categories as they are defined by the institution. Rawls views his two principles of justice as superconstitutional, in the sense that these are the most general criteria against which constitutions are to be judged. If the constitution meets these criteria, then it is a just constitution, and the laws, social arrangements, and particular actions taken in accordance with such a just constitution will also be just, at least *prima facie*.

The argument for these two principles as the most general principles for judging institutions is based on notions of reciprocity, fairness, and ultimately on the notion of the equal dignity of all persons. In this it shows its derivation from Kantian moral philosophy in particular and the contractarian tradition in general. Rawls comes to the two principles from these general notions by means of what he calls an analytical construction. This is a kind of simulation in which persons with defined characteristics and in a defined situation choose the principles in terms of which the justness of their institutions will be judged; or, put differently, they

* In "Justice as Fairness"

torality

ese principles express justice
ality, and reward for services

n equal liberty. This means
everyone could have more
partures from equality are
yone's advantage, has been
rture from equality of dis-
t what goods are being dis-
ult of such an inequality is
than he would be in a situa-
le of equal maximum liberty
enship, political rights, civil
ntails guarantees of the suf-
peech and thought, and free-
e applies primarily to what
ie distribution of economic
ciety. It should be noted that
of justice, applying as they
in the first instance to the
position of classes and cate-
gory. Rawls views his two
principle, in the sense that these
which constitutions are to be
criteria, then it is a just com-
ments, and particular actions
constitution will also be just,

bles as the most general prin-
d on notions of reciprocity,
a of the equal dignity of all
1 from Kantian moral philos-
-tarian tradition in general.
om these general notions by
onstruction. This is a kind of
fined characteristics and in a
s in terms of which the just-
ged; or, put differently, they

choose the criteria by which claims and complaints against con-
crete institutions will be judged. The defining characteristics of the
persons and of the situation in this heuristic model are not in-
tended to correspond to any realistic situation, but rather each fea-
ture expresses some significant aspects of the general notions from
which justice derives.

The analytical construction is that of an "original position," a
hypothetical state of nature, in which persons choose the principles
of their institutions. The defining features of these persons are:
(1) they are rational, in the sense that they can recognize their
own interests, that their interests can be coherently ordered, and
that if need be they can postpone immediate gratification for the
sake of their long run interests as they conceive them; (2) they
are self-interested in the very broad sense that whatever their in-
terests, tastes, preferences, and values may be — whether they be
for sensual gratification, acts of charity, contemplation, or whatever
— they wish to maximize those interests; (3) they are capable of un-
derstanding and adhering to moral principles, which for these pur-
poses Rawls defines as principles which constrain the pursuit of
self-interest.

The situation in which these abstract persons are placed is one in
which (1) they know the general facts about human nature and
human society, and that on leaving the original position they will
take up some role or other in human society; but (2) they know
nothing — not their sex, nor generation, nor country, nor intel-
ligence, nor state of health, nor tastes, nor religion, nor values —
about themselves as concrete individuals nor about their concrete
situation; they do know, however, that they will have tastes, in-
terests, values, and a concrete situation. (3) Behind this veil of
ignorance they must choose finally and irrevocably the moral prin-
ciples in terms of which to judge the concrete institutions in which
they will find themselves in real life. These principles are moral
in the sense that once having chosen them the abstract persons
must abide by them, and they can expect others to abide by them,
even when it is not in the best interests of the person abiding by
them to do so. (4) Rawls specifies that the principles must be
chosen unanimously and without opportunity for coalition. This
last seems unnecessary because all participants in this simulation
are defined in the same way and have the same knowledge — for

instance, they do not even know whether in real life they will or will not be risk averse — and there is therefore no reason for any one of them to choose differently. Nor is there any reason for them to form coalitions, since they cannot tell when such a coalition will be in their own best interests.

Now each of these features of the original position is expressive of some general notion that stands behind this system of moral principles. First, the ignorance of the participants about their own interests and circumstances assures that their choice will have the quality of impartiality and equality. As Rawls puts it, they will choose principles as if their enemies would assign them places. (If you want a fair division of a pie between two persons, let one person divide the pie, and the other choose the first piece.) This also gives a sense to the notion that social arrangements must be such that no one is an instrument for serving the interests of another, unless at least his interests are also thereby served. This expression of the Kantian notion of not using another as a means only is a consequence of the original position, since a participant being rational, self-interested, and ignorant of his particular characteristics and circumstances cannot discover a principle which would allow him to use another as a means alone, and he would also want a principle that would protect him against being so used.

The definition of these persons as self-interested — that is, disposed to pursue their interests and values whatever they might be — is an expression of these notions: that all persons have an equal dignity, no matter what their values and interests; that our interests or ends define us, so that to compromise our interests in some essential way is to compromise our essence as persons; and third, that the principles of justice do not express a complete ordering of values and interests, but only a partial ordering dealing with the resolution of conflicts between different persons whose interests may be in conflict. In regard to this third point, the notion is that some ethical, aesthetic, or other principle may resolve conflicts between and establish a complete ordering among all ends and interests, perhaps even ruling some out altogether. It is not the job of compulsory social institutions to do this, but only to establish some order among persons when their ends conflict. Furthermore this third point expresses the notion that complete orderings for an individual person, if they are to have value, must be freely chosen.

orality

er in real life they will or
herefore no reason for any
; there any reason for them
when such a coalition will

ginal position is expressive
hind this system of moral
articipants about their own
their choice will have the
as Rawls puts it, they will
ould assign them places. (If
ween two persons, let one
hooose the first piece.) This
ocial arrangements must be
serving the interests of an-
so thereby served. This ex-
ing another as a means only
n, since a participant being
f his particular characteris-
a principle which would al-
ne, and he would also want
st being so used.

self-interested — that is, dis-
ues whatever they might be
at all persons have an equal
nd interests; that our inter-
omise our interests in some
ence as persons; and third,
express a complete ordering
rtial ordering dealing with
rent persons whose interests
ird point, the notion is that
ple may resolve conflicts be-
ing among all ends and in-
together. It is not the job of
s, but only to establish some
s conflict. Furthermore this
complete orderings for an
alue, must be freely chosen.

But freedom cannot be so complete *between* persons, at least where the liberty of one conflicts with that of another. Finally, we can see also how the ignorance and self-interestedness of the participants is a way of guaranteeing that the principles they choose will have the character of Kant's principle of right: that all actions must be compatible with a maxim by which a like freedom is accorded to all.

The features of the original position that the principles chosen are chosen by moral persons as moral principles, and once and for all, express a number of important aspects of justice. First, that justice is a moral principle — that it will sometimes operate as a constraint on the pursuit of self-interest, or that it is an end in itself, and not just a way of getting to other ends. Second, that the principle is chosen once and for all expresses the notion that justice, being a moral principle, is universal and categorical, and cannot be avoided just because it may become inconvenient.

To return to the two principles of justice, it is Rawls's thesis that persons in the original position would choose, as the criteria for institutions and practices, the two principles of justice, because these are the principles which would offer to each the best protection of all his interests, whatever they may turn out to be. And if the analytical construction does indeed yield the two principles, Rawls, in the contractarian tradition, has found a way to give a determinate sense to notions such as equal dignity, the essential liberty of man, and the like.

Rawls's principle of right; its relation to Kant. Rawls uses the analytical construction not only to generate principles for judging long-term social arrangements. He uses it as well to develop the principles for *individual* obligation within such arrangements, and to generate the principles of beneficence and of natural obligation. Thus we might say that the original position and analytical construction is the most general schema for obligations of one person to another. Rawls does not put forward the analytical construction as itself being a principle, but in one sense it clearly is the most general principle of the class of obligations, for it stands behind all such more particular principles and generates them, with the addition of further premises. The analytical construction, viewed as a most general principle, is expressive of certain general notions such as liberty, equality, the right of all men to be treated as ends

in themselves, and moral nature as defined by freedom and rationality.

The close affinity of Rawls's system to Kant's general ethical system and to his philosophy of right in particular should be evident. I shall not attempt to make the connection in any detail. I would point out that the specification of the features of the original position gives expression to the main features of Kant's account, which unfortunately in Kant's writing are never presented in such a way that one can be sure how they might be made to apply to particular situations. First, there is Kant's notion that morality is freely chosen, never imposed — the principles of justice and other principles of obligation are freely chosen in the original position. Second, there is Kant's notion that a moral being is essentially a free and rational being. This idea causes trouble for Kant, because he never quite shows how man's substantive ends gather moral significance; unless one finds a place for these substantive ends the system is in danger of circularity or vacuity, as Hegel was the first to point out. Rawls meets this point by according substantive interests to the persons in the original position and by positing that they are self-interested, while at the same time depriving them of knowledge of what their substantive interests are. This leaves them as purposive beings (that is, beings with substantive ends) who are free, rational and capable of understanding, and disposed to follow moral principles. Finally, Rawls's whole complex scheme, beginning with the analytical construction, passing through the two general principles of justice, and ending in the principles governing individual obligations, gives sense to the rather vague Kantian notion that morality or right consists in the compatibility of the maxim of a particular act with a universal law having a particular characteristic: namely, that it is a law which free and rational beings would choose.

Individual obligations. As I have stated, the principles of justice apply to institutions, not to individual obligation. An institution may comport with justice and thus be immune to complaint on that score, but still further principles are needed to account for the moral obligation of particular persons to fulfill the roles and obligations defined for them by the institution. For example, a society's constitution and laws may comport with justice, but what

defined by freedom and ra-

to Kant's general ethical sys-
particular should be evident.
tion in any detail. I would
features of the original posi-
res of Kant's account, which
ver presented in such a way
be made to apply to particu-
tion that morality is freely
s of justice and other prin-
n the original position. Sec-
al being is essentially a free
ouble for Kant, because he
tive ends gather moral sig-
r these substantive ends the
uity, as Hegel was the first
by according substantive in-
osition and by positing that
ame time depriving them of
terests are. This leaves them
th substantive ends) who are
iding, and disposed to follow
ole complex scheme, begin-
passing through the two gen-
in the principles governing
he rather vague Kantian no-
in the compatibility of the
ersal law having a particular
which free and rational beings

ated, the principles of justice
al obligation. An institution
immune to complaint on that
e needed to account for the
is to fulfill the roles and ob-
stitution. For example, a so-
mport with justice, but what

is the basis of the citizen's moral obligation to obey those laws? Now to pose this question is not to deny the moral character of just laws — they still have some moral force, even if no individual obligation followed. Thus a just society may enforce its laws against individuals, and the justice of the laws gives the society a moral title to do so and deprives individuals of any justified complaint against the laws. But society's moral right to coerce compliance with just laws does not entail a moral obligation on the part of individuals to obey these just laws.

The question thus arises what is the individual's moral obligation in institutional settings. Rawls considers three plausible alternatives: (1) that the individual has no moral obligation; (2) that the individual has an obligation to obey the rules and practices of any just institution; (3) that the individual has an obligation to obey the rules of a just institution if he has or intends to accept benefits under that institution — this is the principle of fairness, of doing your part if you have benefited from others doing their part. Rawls tests these three candidates by the analytical construction.

The first fails because the benefits to all under a just society would be significantly less if the citizens were morally free to disobey its laws, so that obedience could only be assured by threats and coercion. The cost of the enforcement mechanism as well as the cost of a lower level of compliance under such a system could be saved, and the saving distributed to all, if compliance could be assured in the main by each citizen's self-enforcement via his sense of moral obligation. Thus it appears that morality — by running counter to self-interest in particular cases — is in the long run interest of all. But being a moral obligation it is binding even where in a particular case it runs counter to self-interest. On this argument Rawls concludes that in respect to individual obligation some moral principle is preferable, and that the analytical construction leads to the acceptance of some moral constraints by individuals. It remains to discover what that moral obligation is.

The second candidate — full obligation to any just institution — is also rejected because it entails an excessive restriction on liberty. This can be seen by comparing it to the third principle, the principle of fair play or of fairness. By this principle there is a moral obligation to a just institution but *only* if the individual accepts or intends to accept benefits from the sacrifices of others who

accept the moral constraints of the institution. With this third alternative available, Rawls argues, there would be no reason for rational, self-interested individuals to accept the greater constraint on liberty implicit in the second principle. For the third principle promises the benefits of the second but leaves open an option of refusing the burdens if one is willing to forgo the benefits. There is, however, another way one could come to this same conclusion. Under both the second and the third alternatives the obligation only attaches to just institutions, and it might be argued that it is a corollary of the first principle of justice that the rules of a just institution not apply to unwilling participants who have not accepted benefits under the institution.*

Utility and self-interest. A striking feature of Rawls's account is the contrast he draws between his principles — of right, justice, and fairness — and the utilitarian principle or any principle of self-interest.

The contrast to self-interest has already been adverted to. Perhaps the notion of self-interest should be more clearly specified, especially in the context of my essay, which defines morality as a substantive end. Rawls appears to conceive of self-interest as those substantive ends of a person which do not derive from moral considerations. His argument is that, though we accept right, justice, and fair play in the context of other substantive pursuits, we accept these principles as constraints on the other pursuits. Thus, though the most advantageous principles to adopt in the original position are justice and fairness, it does not at all follow that in particular situations in the actual world, the most advantageous thing to do is to adhere to those principles. On the contrary, if these moral principles were coextensive with simple advantage in concrete cases, they would be superfluous. Nor is there any reason to believe that principles generated in the deliberately contrived model of the original position — contrived to express important values — are the same as would be generated by self-interest in concrete, actual cases. It may be in my interest to have, for instance, an institution of property, on the assumption that I can only have what all would agree to in the original position, but surely the best

* It should be added that the principle of fairness applies not only to social institutions like states and organizations, but to promises, and indeed to any just practice from which a person accepts or intends to accept benefits.

institution. With this third there would be no reason for to accept the greater constraint principle. For the third principle but leaves open an option of to forgo the benefits. There is, come to this same conclusion. rd alternatives the obligation it might be argued that it is a justice that the rules of a just participants who have not ac-

feature of Rawls's account is principles — of right, justice, and principle or any principle of self-

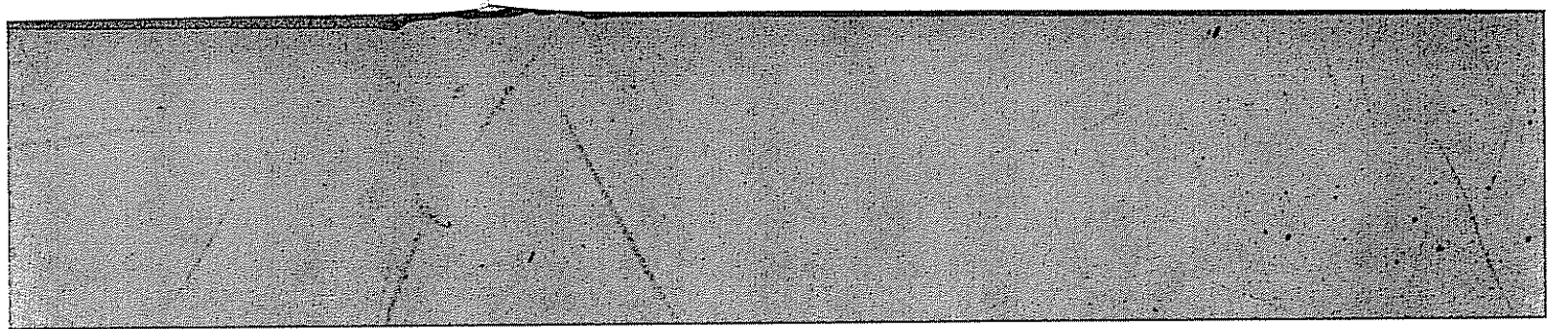
already been adverted to. Per- could be more clearly specified, y, which defines morality as a conceive of self-interest as those do not derive from moral con- ough we accept right, justice, er substantive pursuits, we ac- on the other pursuits. Thus, iples to adopt in the original does not at all follow that in world, the most advantageous principles. On the contrary, if sive with simple advantage in luous. Nor is there any reason in the deliberately contrived ontrived to express important enerated by self-interest in con- terest to have, for instance, an mption that I can only have al position, but surely the best of fairness applies not only to social ut to promises, and indeed to any or intends to accept benefits.

for my narrow advantage would be if everyone else were bound to honor just institutions, to respect property, to keep his promises, and so on, and only I was not. In this crucial sense justice, fairness, and morality in general — though they take self-interest into account — are obvious constraints upon self-interest.

A much more difficult and philosophically controversial issue concerns the contrast between the Rawlsian system of principles and various forms of utilitarianism. Rawls is very explicit in pointing out how both the principles of justice and of individual obligation can lead to different results from what utilitarianism prescribes as a rule for institutions or individuals. The principles of justice do not allow departures from equality to be justified simply on the ground that they lead to a higher sum of utility or a higher average utility. Only if the increased utility can be shown to benefit the worst-off person or class can the departure be justified. Thus Rawls concludes that a caste or slavery system can never be justified on the ground that it is more efficient — that is, leads to a higher level of utility. And this accords with the moral premise that no person is to be used as the mere instrument of any other person. If inequalities are justified by reason alone that the sum of utilities or average level of utility is thereby raised, then the less well-off in such a scheme are the uncompensated instruments of the greater well-being of the better-off. Rawls's principle of distributive justice precludes such a situation. In respect to the contrast between his scheme and utilitarianism, a passage from an unpublished paper of Rawls's summarizes his views well:

The striking feature of the principle of utility is that it doesn't matter, except indirectly, how this sum of satisfactions is distributed among individuals any more than it matters, except indirectly, how one man distributes his satisfactions over time . . .

There is, however, no reason to suppose that the principles which should regulate a group of men are simply the extension of the principle of rational choice for one man. To the contrary: if we assume that the correct principle for any thing must rest in part on the nature of that thing, and if we regard the plurality of distinct persons as an essential feature of associations (as the spuriousness of all organic theories of society con-



firms), we should not expect the principle of social choice to be utilitarian. Yet classical utilitarianism by means of the conception of the impartial sympathetic spectator conflates all persons into one, the many systems of individual desire into one system of ends; for by identifying in turn with the interests in conflict the impartial spectator treats all desires as if they were his own. Since this construction gives no place to the separateness of individuals but dissolves them into one, it must surely lead to an erroneous theory of right. Furthermore, if we believe (as we seem to) that, as a matter of principle, each member of society has an inviolability founded on justice which even the welfare of everyone else cannot override, and that a loss of freedom for some is not made right by a greater sum of satisfactions enjoyed by many, we shall have to look for another account of justice. The principle of utility cannot explain the fact that in a just society the liberties of equal citizenship are taken for granted and the rights secured by justice are not subject to political bargaining or to the calculus of social interests. Hence, on reflection, the notion of maximizing the good is not a plausible conception of right, nor does our reasoning in regard to fundamental liberties appear to conform to it.*

In respect to individual obligation and utility an analogous point holds. Rawls's principle of fairness requires not only that an individual forgo personal advantage in fidelity to his obligations to a just

* There has been considerable debate on this point: Why would the persons in the original position not choose, instead of the principles of justice—and particularly the corollary of making the worst-off best off—simply the principle of the highest average utility? The reason that is given for their doing so is that since they do not know their positions, the highest average utility assures them of the best chance of occupying the most satisfying position, while Rawls's principle may mean that considerable advantages may have to be forgone. Rawls answers this objection by arguing that to gamble in this way would be irrational in the absence of any knowledge about one's actual situation, that such a gamble would be inconsistent with obligations to one's children or a religion one might possibly have, and that the resulting institutions would be subject to peculiar strains and instabilities. I would add the point that it misconceives the analytical construction to argue as if the persons in the original position had the kind of continuity of identity and experience with the actual persons who are governed by the rules there chosen that would be necessary to console them in the real world for the unfortunate outcome of a lottery that never actually took place.

principle of social choice to be
 nism by means of the concep-
 spectator conflates all persons
 lividual desire into one system
 n with the interests in conflict
 esires as if they were his own.
 lace to the separateness of in-
 one, it must surely lead to an
 hermore, if we believe (as we
 iple, each member of society
 justice which even the welfare
 and that a loss of freedom for
 er sum of satisfactions enjoyed
 for another account of justice.
 explain the fact that in a just
 zenship are taken for granted
 e are not subject to political
 ocial interests. Hence, on reflec-
 ne good is not a plausible con-
 reasoning in regard to funda-
 1 to it.*

and utility an analogous point
 quires not only that an individ-
 lity to his obligations to a just

this point: Why would the persons
 d of the principles of justice—and
 worst-off best off—simply the prin-
 reason that is given for their doing
 positions, the highest average utility
 pying the most satisfying position,
 onsiderable advantages may have to
 by arguing that to gamble in this
 f any knowledge about one's actual
 nconsistent with obligations to one's
 have, and that the resulting institu-
 : and instabilities. I would add the
 nstruction to argue as if the persons
 ontinuity of identity and experience
 by the rules there chosen that would
 world for the unfortunate outcome

institution or practice, but also that he fulfill his obligations where
 to break them would lead to a greater sum of utility. This is the
 classical problem of keeping a promise when utility would be in-
 creased if I broke it, or obeying a just law — say a tax law — when
 cheating would lead to a higher level of utility; and it is not as-
 sumed that the beneficiary would always necessarily be the obliga-
 tion-breaker himself. The usual objection to a utilitarian concep-
 tion of obligation, "But what if everybody did the same?" is recog-
 nized to be inadequate. For in those factual circumstances where
 one person's acting on the utilitarian rule will not cause others to
 breach their obligations — as where the violation can be kept secret
 — the argument that this violation in other circumstances might
 lead others to do the same, thus undermining the institution, seems
 beside the point. Rawls, by recurring to the conception of fairness
 chosen as a public rule in the analytic construction, expresses the
 independent value that no person should be required to make sacri-
 fices to an institution on a principle that will not also be binding
 on others, for otherwise such a person is again used as an instru-
 ment to produce the benefit for others. Indeed, he's been had. Nor
 can it be argued that the utilitarian principle of obligation would
 itself ever be chosen or promulgated as a public principle of obliga-
 tion. For, as Kant pointed out, as a public principle of obligation
 utilitarianism would make all institutions and practices unstable.*
 Thus Rawls's principle of individual obligation meets the utilitarian
 objection not by showing that utility is indeed served by disregard-
 ing utility — this is an argument that rests on a fallacy — but by
 bringing in a value apart from utility that is served by fidelity to
 the obligations of institutions and practices.

The relation of Rawls's system to the concepts of this essay. The
 point of this elaborate presentation of Rawl's system of con-
 cepts is that I wish to incorporate them by reference, as it were,
 into the system of this essay. Specifically, I wish to incorporate the
 principles of justice and fairness as rational principles available to
 score rational ends and actions. They are less general than the
 principle of morality, but are derivable from that principle. They
 are, however, still too general to score rational ends directly. Fur-

* For Kant and Rawls it is an important criterion of the morality of a
 principle that it can be publicly promulgated. This is related to the notion
 of respect for the rationality of moral agents.

ther principles and material premises are needed to control concrete ends and actions. Thus the principles of justice apply to institutions, and institutions, as I shall argue in Part Two, are themselves systems of principles and rules. Justice applies to institutions, a particular just institution has a further system of rules — from very general ones like a constitution, to very particular ones like a particular contract drawn according to the civil law — and finally a concrete end will be scored by all of these together with the principles and material premises arising out of the circumstances of the concrete situation in which an agent pursues a particular end.

The principle of morality, it will be recalled, is the most general principle applicable to those rational actions and ends that impinge on other persons. It is put forward as that principle in terms of which a person can recognize in his dealings with other persons the personality of those other persons. It is that principle in terms of which an actor may assure that the principle of his rational end incorporates or is compatible with the rational ends of those with whom he interacts. Or, to state the point differently, it is that principle in terms of which an actor treats other persons as ends and never as means alone. It should be obvious that my principle of morality is equivalent to Rawls's principle of right, at least as I have understood it and presented it here. Rawls's principle of right is expressed in terms of the analytical construction. The determinants of the analytical construction are each expressive of a number of related values: liberty, equality, impartiality, morality as a constraint on self-interest, and respect for others as ends. And these are precisely the same conceptions which my principle of morality is intended to express. Both are related to Kant's categorical imperative and his "universal principle of justice." Thus I intend my principle of morality to be substantially equivalent to Rawls's principle of right.

There is another equivalence between these two principles which I would call an equivalence of function. Neither Rawls's principle of right nor my principle of morality is meant to apply directly; rather both are intended to generate further more particular principles. In the scheme of this work, the principle of morality is the most general principle for transactions. From it are derived more specific principles, until finally we arrive at the rational principle for a particular end or action.

are needed to control concrete
 es of justice apply to institu-
 e in Part Two, are themselves
 e applies to institutions, a par-
 e system of rules — from very
 very particular ones like a
 o the civil law — and finally a
 these together with the princi-
 at of the circumstances of the
 pursues a particular end.

be recalled, is the most gen-
 onal actions and ends that im-
 ard as that principle in terms
 is dealings with other persons
 s. It is that principle in terms
 e principle of his rational end
 e rational ends of those with
 oint differently, it is that prin-
 ats other persons as ends and
 obvious that my principle of
 iple of right, at least as I
 ere. Rawls's principle of right
 ical construction. The deter-
 ion are each expressive of a
 ality, impartiality, morality as
 pect for others as ends. And
 tions which my principle of
 are related to Kant's categori-
 niple of justice." Thus I in-
 e substantially equivalent to

en these two principles which
 on. Neither Rawls's principle
 is meant to apply directly;
 further more particular prin-
 e principle of morality is the
 s. From it are derived more
 ive at the rational principle

I have contrasted Rawls's principles to both self-interest and to utility. That contrast has a peculiar significance for this essay. A person acting out of morality, or out of justice, or out of fairness acts on a rational principle which is not completely equivalent to the principle he would adopt if he were seeking to further any other interests, either of his own (self-interest) or of others (maximizing the sum of utility). In the case of such rational actions and ends the actor does of course also pursue certain interests of his own or of other persons. There is no such thing as an act or end which is simply just, simply moral, simply fair; rather, these other interests are pursued subject to the constraints of morality and justice. But to pursue one's own ends or the ends of others morally and justly is not the same as simply pursuing them. The principle of morality and its derivatives constrain or shape the form and ordering of those other substantive pursuits. A person who returns stolen property serves the ends of another, but he does not do so simply to serve the ends of the other nor even to maximize the sum of utilities. Rather he does it in accordance with the principles of morality and fairness. Of course, those principles would not require that act if it was not an act that also served somebody's interests — one need not return the sword of a man who has gone mad. Thus the fact that these principles lead to particular concrete ends and actions, that have an ordering and content different from that entailed by other interests, gives a precise and palpable sense to the proposition that there are ends and actions which may be described as moral ends and actions. And also this conclusion provides an interpretation for Kant's claim that an act is good only if it is done for the sake of morality or duty.

Finally, by incorporating Rawls's principles into this account to serve as rational principles at our intermediate level of generality we can see better how rational principles not only are constitutive of particular ends but also imply an ordering of ends. For each of Rawls's principles applies to an infinity of situations, and does so categorically — that is, whenever a situation of a particular sort is present, then the principles apply and demand that an end of a particular sort be chosen. Justice is a constraint applicable to any institution or practice; fairness is a constraint in any situation of mutual aid, forbearance, and expectation. Thus the orderings implied by these principles can be quite dense and exigent, though

they are far from being complete — there is still a lot of room for diverse ends and actions within the framework they establish. I have shown that it is implicit in the principle of morality that if a person accepts the principle of morality for one action this commits him in a special sense to accepting the ordering of all his other ends. The same arguments hold true for the orderings implicit in justice and fairness — if a man would be a just man or a fair man, who does a just or a fair act, he must accept those principles not just as principles for this end or act but as principles for the ordering of all his ends and actions.

In the next chapter I shall take these principles to a further stage of particularity and show how they are implicated in the concepts of love, friendship, and trust; in emotions and relations of love and trust; in particular rational actions and ends expressive of those relations and emotions; and in the ordering of ends and relations implicit in love, friendship, and trust.