

# ANARCHY, STATE, AND UTOPIA


---

ROBERT NOZICK

---

BASIC  
**B**  
BOOKS

A Member of The Perseus Books Group



PART

I

---

*State-of-Nature Theory,  
or How to Back into a State  
without Really Trying*

# CHAPTER

## 1

---

# Why State-of-Nature Theory?

---

**I**F the state did not exist would it be necessary to invent it? Would one be *needed*, and would it have to be *invented*? These questions arise for political philosophy and for a theory explaining political phenomena and are answered by investigating the “state of nature,” to use the terminology of traditional political theory. The justification for resuscitating this archaic notion would have to be the fruitfulness, interest, and far-reaching implications of the theory that results. For the (less trusting) readers who desire some assurance in advance, this chapter discusses reasons why it is important to pursue state-of-nature theory, reasons for thinking that theory would be a fruitful one. These reasons necessarily are somewhat abstract and metatheoretical. The best reason is the developed theory itself.

## POLITICAL PHILOSOPHY

The fundamental question of political philosophy, one that precedes questions about how the state should be organized, is whether there should be any state at all. Why not have anarchy? Since anarchist theory, if tenable, undercuts the whole subject of *political* philosophy, it is appropriate to begin political philosophy with an examination of its major theoretical alternative. Those who consider anarchism not an unattractive doctrine will think it possible that political philosophy *ends* here as well. Others impatiently will await what is to come afterwards. Yet, as we shall see, anarchists and anarchists alike, those who spring gingerly from the starting point as well as those reluctantly argued away from it, can agree that beginning the subject of political philosophy with state-of-nature theory has an *explanatory* purpose. (Such a purpose is absent when epistemology is begun with an attempt to refute the skeptic.)

Which anarchic situation should we investigate to answer the question of why not anarchy? Perhaps the one that would exist if the actual political situation didn't, while no other possible political one did. But apart from the gratuitous assumption that everyone everywhere would be in the same nonstate boat and the enormous unmanageability of pursuing that counterfactual to arrive at a particular situation, that situation would lack fundamental theoretical interest. To be sure, if that nonstate situation were sufficiently awful, there would be a reason to refrain from dismantling or destroying a particular state and replacing it with none, now.

It would be more promising to focus upon a fundamental abstract description that would encompass all situations of interest, including "where we would now be if." Were this description awful enough, the state would come out as a preferred alternative, viewed as affectionately as a trip to the dentist. Such awful descriptions rarely convince, and not merely because they fail to cheer. The subjects of psychology and sociology are far too feeble to support generalizing so pessimistically across all societies and persons, especially since the argument depends upon *not* making *such* pessimistic assumptions about how the *state* operates. Of

course, people know something of how actual states have operated, and they differ in their views. Given the enormous importance of the choice between the state and anarchy, caution might suggest one use the "minimax" criterion, and focus upon a pessimistic estimate of the nonstate situation: the state would be compared with the most pessimistically described Hobbesian state of nature. But in using the minimax criterion, this Hobbesian situation should be compared with the most pessimistically described possible state, including *future* ones. Such a comparison, surely, the worst state of nature would win. Those who view the state as an abomination will not find minimax very compelling, especially since it seems one could always bring back the state if that came to seem desirable. The "maximax" criterion, on the other hand, would proceed on the most optimistic assumptions about how things would work out—Godwin, if you like that sort of thing. But imprudent optimism also lacks conviction. Indeed, no proposed decision criterion for choice under uncertainty carries conviction here, nor does maximizing expected utility on the basis of such frail probabilities.

More to the point, especially for deciding what goals one should try to achieve, would be to focus upon a nonstate situation in which people generally satisfy moral constraints and generally act as they ought. Such an assumption is not wildly optimistic; it does not assume that all people act exactly as they should. Yet this state-of-nature situation is the best anarchic situation one reasonably could hope for. Hence investigating its nature and defects is of crucial importance to deciding whether there should be a state rather than anarchy. If one could show that the state would be superior even to this most favored situation of anarchy, the best that realistically can be hoped for, or would arise by a process involving no morally impermissible steps, or would be an improvement if it arose, this would provide a rationale for the state's existence; it would justify the state.\*

This investigation will raise the question of whether all the ac-

---

\* This contrasts with a theory that presents a state's arising from a state of nature by a natural and inevitable process of *deterioration*, rather as medical theory presents aging or dying. Such a theory would not "justify" the state, though it might resign us to its existence.

tions persons must do to set up and operate a state are themselves morally permissible. Some anarchists have claimed not merely that we would be better off without a state, but that any state necessarily violates people's moral rights and hence is intrinsically immoral. Our starting point then, though nonpolitical, is by intention far from nonmoral. Moral philosophy sets the background for, and boundaries of, political philosophy. What persons may and may not do to one another limits what they may do through the apparatus of a state, or do to establish such an apparatus. The moral prohibitions it is permissible to enforce are the source of whatever legitimacy the state's fundamental coercive power has. (Fundamental coercive power is power not resting upon any consent of the person to whom it is applied.) This provides a primary arena of state activity, perhaps the only legitimate arena. Furthermore, to the extent moral philosophy is unclear and gives rise to disagreements in people's moral judgments, it also sets problems which one might think could be appropriately handled in the political arena.

#### EXPLANATORY POLITICAL THEORY

In addition to its importance for political philosophy, the investigation of this state of nature also will serve explanatory purposes. The possible ways of understanding the political realm are as follows: (1) to fully explain it in terms of the nonpolitical; (2) to view it as emerging from the nonpolitical but irreducible to it, a mode of organization of nonpolitical factors understandable only in terms of novel political principles; or (3) to view it as a completely autonomous realm. Since only the first promises full understanding of the whole political realm,<sup>1</sup> it stands as the most desirable theoretical alternative, to be abandoned only if known to be impossible. Let us call this most desirable and complete kind of explanation of a realm a *fundamental* explanation of the realm.

To explain fundamentally the political in terms of the nonpolitical, one might start either with a nonpolitical situation, showing how and why a political one later would arise out of it, or with a

political situation that is described nonpolitically, deriving its political features from its nonpolitical description. This latter derivation either will identify the political features with those features nonpolitically described, or will use scientific laws to connect distinct features. Except perhaps for this last mode, the illumination of the explanation will vary directly with the independent glow of the nonpolitical starting point (be it situation or description) and with the distance, real or apparent, of the starting point from its political result. The more fundamental the starting point (the more it picks out basic, important, and inescapable features of the human situation) and the less close it is or seems to its result (the less political or statelike it looks), the better. It would not increase understanding to reach the state from an arbitrary and otherwise unimportant starting point, obviously adjacent to it from the start. Whereas discovering that political features and relations were reducible to, or identical with, ostensibly very different nonpolitical ones would be an exciting result. Were these features fundamental, the political realm would be firmly and deeply based. So far are we from such a major theoretical advance that prudence alone would recommend that we pursue the alternative of showing how a political situation would arise out of a nonpolitical one; that is, that we begin a fundamental *explanatory* account with what is familiar within political philosophy as state-of-nature theory.

A theory of a state of nature that begins with fundamental general descriptions of morally permissible and impermissible actions, and of deeply based reasons why some persons in any society would violate these moral constraints, and goes on to describe how a state would arise from that state of nature will serve our explanatory purposes, *even if no actual state ever arose that way*. Hempel has discussed the notion of a potential explanation, which intuitively (and roughly) is what would be the correct explanation if everything mentioned in it were true and operated.<sup>2</sup> Let us say that a *law-defective* potential explanation is a potential explanation with a false lawlike statement and that a *fact-defective* potential explanation is a potential explanation with a false antecedent condition. A potential explanation that explains a phenomenon as the result of a process *P* will be defective (even though it is neither law-defective nor fact-defective) if some process *Q* other than *P* produced the

phenomenon, though *P* was capable of doing it. Had this other process *Q* not produced it, then *P* would have.\* Let us call a potential explanation that fails in this way actually to explain the phenomenon a *process-defective* potential explanation.

A *fundamental* potential explanation (an explanation that would explain the whole realm under consideration were it the actual explanation) carries important explanatory illumination even if it is *not* the correct explanation. To see how, in principle, a *whole realm* could fundamentally be explained greatly increases our understanding of the realm.† It is difficult to say more without examining types of cases; indeed, without examining particular cases, but this we cannot do here. Fact-defective fundamental potential explanations, if their false initial conditions “could have been true,” will carry great illumination; even wildly false initial conditions will illuminate, sometimes very greatly. Law-defective fundamental potential explanations may illuminate the nature of a realm almost as well as the correct explanations, especially if the “laws” together form an interesting and integrated theory. And process-defective fundamental potential explanations (which are neither law-defective nor fact-defective) fit our explanatory bill and purposes almost perfectly. These things could not be said as strongly, if at all, about nonfundamental explanation.

State-of-nature explanations of the political realm *are* fundamental potential explanations of this realm and pack explanatory

---

\* Or, perhaps yet *another* process *R* would have if *Q* hadn't, though had *R* not produced the phenomenon, then *P* would have, or. . . . So the footnoted sentence should read: *P* would have produced the phenomenon had no member of [*Q*, *R*, . . .] done so. We ignore here the complication that what would prevent *Q* from producing the phenomenon might also prevent *P* from doing so.

† This claim needs to be qualified. It will not increase our understanding of a realm to be told as a potential explanation what we know to be false: that by doing a certain dance, ghosts or witches or goblins made the realm that way. It is plausible to think that an explanation of a realm must present an underlying mechanism yielding the realm. (Or do something else equally productive of understanding.) But to say this is not to state precisely the deep conditions an underlying mechanism must satisfy to explain a realm. The precise qualification of the claim in the text awaits advances in the theory of explanation. Yet other difficulties call for such advances; see Jaegwon Kim, “Causation, Nomic Subsumption, and the Concept of Event,” *The Journal of Philosophy*, 70, no. 8 (April 26, 1973), 217–236.

punch and illumination, even if incorrect. We learn much by seeing how the state could have arisen, even if it didn't arise that way. If it didn't arise that way, we also would learn much by determining why it didn't; by trying to explain why the particular bit of the real world that diverges from the state-of-nature model is as it is.

Since considerations both of political philosophy and of explanatory political theory converge upon Locke's state of nature, we shall begin with that. More accurately, we shall begin with individuals in something sufficiently similar to Locke's state of nature so that many of the otherwise important differences may be ignored here. Only when some divergence between our conception and Locke's is relevant to *political* philosophy, to our argument about the state, will it be mentioned. The completely accurate statement of the moral background, including the precise statement of the moral theory and its underlying basis, would require a full-scale presentation and is a task for another time. (A lifetime?) That task is so crucial, the gap left without its accomplishment so yawning, that it is only a minor comfort to note that we here are following the respectable tradition of Locke, who does not provide anything remotely resembling a satisfactory explanation of the status and basis of the law of nature in his *Second Treatise*.

# CHAPTER

## 2

# The State of Nature

INDIVIDUALS in Locke's state of nature are in "a state of perfect freedom to order their actions and dispose of their possessions and persons as they think fit, within the bounds of the law of nature, without asking leave or dependency upon the will of any other man" (sect. 4).<sup>1</sup> The bounds of the law of nature require that "no one ought to harm another in his life, health, liberty, or possessions" (sect. 6). Some persons transgress these bounds, "invading others' rights and . . . doing hurt to one another," and in response people may defend themselves or others against such invaders of rights (chap. 3). The injured party and his agents may recover from the offender "so much as may make satisfaction for the harm he has suffered" (sect. 10); "everyone has a right to punish the transgressors of that law to such a degree as may hinder its violation" (sect. 7); each person may, and may only "retribute to [a criminal] so far as calm reason and conscience dictate, what is proportionate to his transgression, which is so much as may serve for reparation and restraint" (sect. 8).

There are "inconveniences of the state of nature" for which, says Locke, "I easily grant that civil government is the proper remedy" (sect. 13). To understand precisely what civil government remedies, we must do more than repeat Locke's list of the inconveniences of the state of nature. We also must consider what arrangements might be made within a state of nature to deal with

these inconveniences—to avoid them or to make them less likely to arise or to make them less serious on the occasions when they do arise. Only after the full resources of the state of nature are brought into play, namely all those voluntary arrangements and agreements persons might reach acting within their rights, and only after the effects of these are estimated, will we be in a position to see how serious are the inconveniences that yet remain to be remedied by the state, and to estimate whether the remedy is worse than the disease.\*

In a state of nature, the understood natural law may not provide for every contingency in a proper fashion (see sections 159 and 160 where Locke makes this point about legal systems, but contrast section 124), and men who judge in their own case will always give themselves the benefit of the doubt and assume that they are in the right. They will overestimate the amount of harm or damage they have suffered, and passions will lead them to attempt to punish others more than proportionately and to exact excessive compensation (sects. 13, 124, 125). Thus private and personal enforcement of one's rights (including those rights that are violated when one is excessively punished) leads to feuds, to an endless series of acts of retaliation and exactions of compensation. And there is no firm way to *settle* such a dispute, to *end* it and to have both parties know it is ended. Even if one party *says* he'll stop his

\* Proudhon has given us a description of the *state's* domestic "inconveniences." "To be GOVERNED is to be watched, inspected, spied upon, directed, law-driven, numbered, regulated, enrolled, indoctrinated, preached at, controlled, checked, estimated, valued, censured, commanded, by creatures who have neither the right nor the wisdom nor the virtue to do so. To be GOVERNED is to be at every operation, at every transaction noted, registered, counted, taxed, stamped, measured, numbered, assessed, licensed, authorized, admonished, prevented, forbidden, reformed, corrected, punished. It is, under pretext of public utility, and in the name of the general interest, to be placed under contribution, drilled, fleeced, exploited, monopolized, extorted from, squeezed, hoaxed, robbed; then, at the slightest resistance, the first word of complaint, to be repressed, fined, vilified, harassed, hunted down, abused, clubbed, disarmed, bound, choked, imprisoned, judged, condemned, shot, deported, sacrificed, sold, betrayed; and to crown all, mocked, ridiculed, derided, outraged, dishonored. That is government; that is its justice; that is its morality." P. J. Proudhon, *General Idea of the Revolution in the Nineteenth Century*, trans. John Beverly Robinson (London: Freedom Press, 1923), pp. 293–294, with some alterations from Benjamin Tucker's translation in *Instead of a Book* (New York, 1893), p. 26.

acts of retaliation, the other can rest secure only if he knows the first still does not feel entitled to gain recompense or to exact retribution, and therefore entitled to try when a promising occasion presents itself. Any method a single individual might use in an attempt irrevocably to bind himself into ending his part in a feud would offer insufficient assurance to the other party; tacit agreements to stop also would be unstable.<sup>2</sup> Such feelings of being mutually wronged can occur even with the clearest right and with joint agreement on the facts of each person's conduct; all the more is there opportunity for such retaliatory battle when the facts or the rights are to some extent unclear. Also, in a state of nature a person may lack the power to enforce his rights; he may be unable to punish or exact compensation from a stronger adversary who has violated them (sects. 123, 126).

#### PROTECTIVE ASSOCIATIONS

How might one deal with these troubles within a state of nature? Let us begin with the last. In a state of nature an individual may himself enforce his rights, defend himself, exact compensation, and punish (or at least try his best to do so). Others may join with him in his defense, at his call.<sup>3</sup> They may join with him to repulse an attacker or to go after an aggressor because they are public spirited, or because they are his friends, or because he has helped them in the past, or because they wish him to help them in the future, or in exchange for something. Groups of individuals may form mutual-protection associations: all will answer the call of any member for defense or for the enforcement of his rights. In union there is strength. Two inconveniences attend such simple mutual-protection associations: (1) everyone is always on call to serve a protective function (and how shall it be decided who shall answer the call for those protective functions that do not require the services of all members?); and (2) any member may call out his associates by saying his rights are being, or have been, violated. Protective associations will not want to be at the beck and call of their cantankerous or paranoid members, not to mention those of their members who might attempt, under the guise of self-defense, to

use the association to violate the rights of others. Difficulties will also arise if two different members of the same association are in dispute, each calling upon his fellow members to come to his aid.

A mutual-protection association might attempt to deal with conflict among its own members by a policy of nonintervention. But this policy would bring discord within the association and might lead to the formation of subgroups who might fight among themselves and thus cause the breakup of the association. This policy would also encourage potential aggressors to join as many mutual-protection associations as possible in order to gain immunity from retaliatory or defensive action, thus placing a great burden on the adequacy of the initial screening procedure of the association. Thus protective associations (almost all of those that will survive which people will join) will not follow a policy of nonintervention; they will use some procedure to determine how to act when some members claim that other members have violated their rights. Many arbitrary procedures can be imagined (for example, act on the side of that member who complains first), but most persons will want to join associations that follow some procedure to find out which claimant is correct. When a member of the association is in conflict with nonmembers, the association also will want to determine in some fashion who is in the right, if only to avoid constant and costly involvement in each member's quarrels, whether just or unjust. The inconvenience of everyone's being on call, whatever their activity at the moment or inclinations or comparative advantage, can be handled in the usual manner by division of labor and exchange. Some people will be *hired* to perform protective functions, and some entrepreneurs will go into the business of selling protective services. Different sorts of protective policies would be offered, at different prices, for those who may desire more extensive or elaborate protection.<sup>4</sup>

An individual might make more particular arrangements or commitments short of turning over to a private protective agency all functions of detection, apprehension, judicial determination of guilt, punishment, and exaction of compensation. Mindful of the dangers of being the judge in his own case, he might turn the decision as to whether he has indeed been wronged, and to what extent, to some other neutral or less involved party. In order for the occurrence of the social effect of justice's being seen to be



done, such a party would have to be generally respected and thought to be neutral and upright. Both parties to a dispute may so attempt to safeguard themselves against the appearance of partiality, and both might even agree upon the *same* person as the judge between them, and agree to abide by his decision. (Or there might be a specified process through which one of the parties dissatisfied with the decision could appeal it.) But, for obvious reasons, there will be strong tendencies for the above-mentioned functions to converge in the same agent or agency.

People sometimes now do take their disputes outside of the state's legal system to other judges or courts they have chosen, for example, to religious courts.<sup>5</sup> If all parties to a dispute find some activities of the state or its legal system so repellent that they want nothing to do with it, they might agree to forms of arbitration or judgment outside the apparatus of the state. People tend to forget the possibilities of acting independently of the state. (Similarly, persons who want to be paternalistically regulated forget the possibilities of contracting into particular limitations on their own behavior or appointing a given paternalistic supervisory board over themselves. Instead, they swallow the exact pattern of restrictions a legislature happens to pass. Is there really someone who, searching for a group of wise and sensitive persons to regulate him for his own good, would choose that group of people who constitute the membership of both houses of Congress?) Diverse forms of judicial adjudication, differing from the particular package the state provides, certainly could be developed. Nor do the costs of developing and choosing these account for people's use of the state form. For it would be easy to have a large number of preset packages which parties could select. Presumably what drives people to use the state's system of justice is the issue of ultimate enforcement. Only the state can enforce a judgment against the will of one of the parties. For the state does not *allow* anyone else to enforce another system's judgment. So in any dispute in which both parties cannot agree upon a method of settlement, or in any dispute in which one party does not trust another to abide by the decision (if the other contracts to forfeit something of enormous value if he doesn't abide by the decision, by what agency is *that* contract to be enforced?), the parties who wish their claims put into effect will have no recourse permitted by the state's legal sys-

tem other than to use that very legal system. This may present persons greatly opposed to a given state system with particularly poignant and painful choices. (If the state's legal system enforces the results of certain arbitration procedures, people may come to agree—supposing they abide by this agreement—without any actual direct contact with what they perceive to be officers or institutions of the state. But this holds as well if they sign a contract that is enforced only by the state.)

Will protective agencies *require* that their clients renounce exercising their right of private retaliation if they have been wronged by nonclients of the agency? Such retaliation may well lead to counterretaliation by another agency or individual, and a protective agency would not wish *at that late stage* to get drawn into the messy affair by having to defend its client against the counterretaliation. Protective agencies would refuse to protect against counterretaliation unless they had first given permission for the retaliation. (Though might they not merely charge much more for the more extensive protection policy that provides such coverage?) The protective agencies need not even require that as part of his agreement with the agency, a client renounce, by contract, his right of private enforcement of justice against its *other clients*. The agency need only refuse a client *C*, who privately enforces his rights against other clients, any protection against counterretaliation upon him by these other clients. This is similar to what occurs if *C* acts against a nonclient. The additional fact that *C* acts upon a client of the agency means that the agency will act toward *C* as it would toward any nonclient who privately enforced his rights upon any one of its clients (see Chapter 5). This reduces intra-agency private enforcement of rights to minuscule levels.

#### THE DOMINANT PROTECTIVE ASSOCIATION

Initially, several different protective associations or companies will offer their services in the same geographical area. What will occur when there is a conflict between clients of different agencies? Things are relatively simple if the agencies reach the same decision about the disposition of the case. (Though each might want to

exact the penalty.) But what happens if they reach different decisions as to the merits of the case, and one agency attempts to protect its client while the other is attempting to punish him or make him pay compensation? Only three possibilities are worth considering:

1. In such situations the forces of the two agencies do battle. One of the agencies always wins such battles. Since the clients of the losing agency are ill protected in conflicts with clients of the winning agency, they leave their agency to do business with the winner.<sup>6</sup>
2. One agency has its power centered in one geographical area, the other in another. Each wins the battles fought close to its center of power, with some gradient being established.<sup>7</sup> People who deal with one agency but live under the power of the other either move closer to their own agency's home headquarters or shift their patronage to the other protective agency. (The border is about as conflictful as one between states.)

In neither of these two cases does there remain very much geographical interspersal. Only one protective agency operates over a given geographical area.

3. The two agencies fight evenly and often. They win and lose about equally, and their interspersed members have frequent dealings and disputes with each other. Or perhaps without fighting or after only a few skirmishes the agencies realize that such battling will occur continually in the absence of preventive measures. In any case, to avoid frequent, costly, and wasteful battles the two agencies, perhaps through their executives, agree to resolve peacefully those cases about which they reach differing judgments. They agree to set up, and abide by the decisions of, some third judge or court to which they can turn when their respective judgments differ. (Or they might establish rules determining which agency has jurisdiction under which circumstances.)<sup>8</sup> Thus emerges a system of appeals courts and agreed upon rules about jurisdiction and the conflict of laws. Though different agencies operate, there is one unified federal judicial system of which they all are components.

In each of these cases, almost all the persons in a geographical area are under some common system that judges between their competing claims and enforces their rights. Out of anarchy, pressed by spontaneous groupings, mutual-protection associations, division of labor, market pressures, economies of scale, and rational

self-interest there arises something very much resembling a minimal state or a group of geographically distinct minimal states. Why is this market different from all other markets? Why would a virtual monopoly arise in this market without the government intervention that elsewhere creates and maintains it?<sup>9</sup> The worth of the product purchased, protection against others, is *relative*: it depends upon how strong the others are. Yet unlike other goods that are comparatively evaluated, maximal competing protective services cannot coexist; the nature of the service brings different agencies not only into competition for customers' patronage, but also into violent conflict with each other. Also, since the worth of the less than maximal product declines disproportionately with the number who purchase the maximal product, customers will not stably settle for the lesser good, and competing companies are caught in a declining spiral. Hence the three possibilities we have listed.

Our story above assumes that each of the agencies attempts in good faith to act within the limits of Locke's law of nature.<sup>10</sup> But one "protective association" might aggress against other persons. Relative to Locke's law of nature, it would be an outlaw agency. What actual counterweights would there be to its power? (What actual counterweights are there to the power of a state?) Other agencies might unite to act against it. People might refuse to deal with the outlaw agency's clients, boycotting them to reduce the probability of the agency's intervening in their own affairs. This might make it more difficult for the outlaw agency to get clients; but this boycott will seem an effective tool only on very optimistic assumptions about what cannot be kept secret, and about the costs to an individual of partial boycott as compared to the benefits of receiving the more extensive coverage offered by an "outlaw" agency. If the "outlaw" agency simply is an *open* aggressor, pillaging, plundering, and extorting under no plausible claim of justice, it will have a harder time than states. For the state's claim to legitimacy induces its citizens to believe they have some duty to obey its edicts, pay its taxes, fight its battles, and so on; and so some persons cooperate with it voluntarily. An openly aggressive agency could not depend upon, and would not receive, any such voluntary cooperation, since persons would view themselves simply as its victims rather than as its citizens.<sup>11</sup>

## INVISIBLE-HAND EXPLANATIONS

How, if at all, does a dominant protective association differ from the state? Was Locke wrong in imagining a compact necessary to establish civil society? As he was wrong in thinking (sects. 46, 47, 50) that an "agreement," or "mutual consent," was needed to establish the "invention of money." Within a barter system, there is great inconvenience and cost to searching for someone who has what you want and wants what you have, even at a marketplace, which, we should note, needn't become a marketplace by everyone's expressly agreeing to deal there. People will exchange their goods for something they know to be more generally wanted than what they have. For it will be more likely that they can exchange this for what they want. For the same reasons others will be more willing to take in exchange this more generally desired thing. Thus persons will converge in exchanges on the more marketable goods, being willing to exchange their goods for them; the more willing, the more they know others who are also willing to do so, in a mutually reinforcing process. (This process will be reinforced and hastened by middlemen seeking to profit in facilitating exchanges, who themselves will often find it most expedient to offer more marketable goods in exchange.) For obvious reasons, the goods they converge on, via their individual decisions, will have certain properties: initial independent value (else they wouldn't begin as more marketable), physically enduring, non-perishable, divisible, portable, and so forth. No express agreement and no social contract fixing a medium of exchange is necessary.<sup>12</sup>

There is a certain lovely quality to explanations of this sort. They show how some overall pattern or design, which one would have thought had to be produced by an individual's or group's successful attempt to realize the pattern, instead was produced and maintained by a process that in no way had the overall pattern or design "in mind." After Adam Smith, we shall call such explanations *invisible-hand explanations*. ("Every individual intends only his own gain, and he is in this, as in so many other cases, led by an invisible hand to promote an end which was no part of his intention.") The specially satisfying quality of invisible-hand explanations (a quality I hope is possessed by this book's account of the

state) is partially explained by its connection with the notion of fundamental explanation adumbrated in Chapter 1. Fundamental explanations of a realm are explanations of the realm in other terms; they make no use of any of the notions of the realm. Only via such explanations can we explain and hence understand everything about a realm; the less our explanations use notions constituting what is to be explained, the more (*ceteris paribus*) we understand. Consider now complicated patterns which one would have thought could arise only through intelligent design, only through some attempt to realize the pattern. One might attempt straightforwardly to explain such patterns in terms of the desires, wants, beliefs, and so on, of individuals, directed toward realizing the pattern. But within such explanations will appear descriptions of the pattern, *at least within quotation marks*, as objects of belief and desire. The explanation itself will say that some individuals desire to bring about something with (some of) the pattern-features, that some individuals believe that the only (or the best, or the . . . ,) way to bring about the realization of the pattern features is to . . . , and so on. Invisible-hand explanations minimize the use of notions constituting the phenomena to be explained; in contrast to the straightforward explanations, they don't explain complicated patterns by including the full-blown pattern-notions as objects of people's desires or beliefs. Invisible-hand explanations of phenomena thus yield greater understanding than do explanations of them as brought about by design as the object of people's intentions. It therefore is no surprise that they are more satisfying.

An invisible-hand explanation explains what looks to be the product of someone's intentional design, as not being brought about by anyone's intentions. We might call the *opposite* sort of explanation a "hidden-hand explanation." A hidden-hand explanation explains what looks to be merely a disconnected set of facts that (certainly) is not the product of intentional design, as the product of an individual's or group's intentional design(s). Some persons also find such explanations satisfying, as is evidenced by the popularity of conspiracy theories.

Someone might so prize each type of explanation, invisible hand and hidden hand, that he might attempt the Sisyphean task of explaining each purported nondesigned or coincidental set of isolated facts as the product of intentional design, *and* each purported

product of design as a nondesigned set of facts! It would be quite lovely to continue this iteration for a bit, even through only one complete cycle.

Since I offer no explicit account of invisible-hand explanations,<sup>13</sup> and since the notion plays a role in what follows, I mention some examples to give the reader a clearer idea of what we have in mind when speaking of this type of explanation. (Examples given to illustrate the type of explanation need not be *correct* explanations.)

1. Explanations within evolutionary theory (via random mutation, natural selection, genetic drift, and so on) of traits of organisms and populations. (James Crow and Motoo Kimura survey mathematical formulations in *An Introduction to Population Genetics Theory* (New York: Harper & Row, 1970).
2. Explanations within ecology of the regulation of animal populations. (See Lawrence Slobodkin, *Growth and Regulation of Animal Populations* [New York: Holt, Rinehart & Winston, 1966] for a survey.)
3. Thomas Schelling's explanatory model (*American Economic Review*, May 1969, pp. 488-493) showing how extreme residential segregation patterns are producible by individuals who do not desire this but want, for example, to live in neighborhoods 55 percent of whose population is in their own group, and who switch their place of residence to achieve their goal.
4. Certain operant-conditioning explanations of various complicated patterns of behavior.
5. Richard Herrnstein's discussion of the genetic factors in a society's pattern of class stratification (*I.Q. in the Meritocracy*, Atlantic Monthly Press, 1973).
6. Discussions of how economic calculation is accomplished in markets. (See Ludwig von Mises, *Socialism*, Part II, *Human Action*, Chapters 4, 7-9.)
7. Microeconomic explanations of the effects of outside intervention in a market, and of the establishment and nature of the new equilibria.
8. Jane Jacobs' explanation of what makes some parts of cities safe in *The Death and Life of Great American Cities* (New York: Random House, 1961).
9. The Austrian theory of the trade cycle.
10. Karl Deutsch and William Madow's observation that in an organization with a large number of important decisions (which can later be evaluated for correctness) to be made among few alternatives, if large numbers of people have a chance to say which way the

decision should be made, a number of persons will gain reputations as sage advisers, even if all randomly decide what advice to offer. ("Note on the Appearance of Wisdom in Large Bureaucratic Organizations," *Behavioral Science*, January 1961, pp. 72-78.)

11. The patterns arising through the operation of a modification of Frederick Frey's modification of the Peter Principle: people have risen three levels beyond their level of incompetence by the time their incompetence is detected.
12. Roberta Wohlstetter's explanation (*Pearl Harbor: Warning and Decision* [Stanford: Stanford University Press, 1962]), contra the "conspiracy" theorists, of why the United States didn't act on the evidence it possessed indicating a Japanese attack forthcoming on Pearl Harbor.
13. That explanation of "the intellectual preeminence of the Jews" that focuses on the great number of the most intelligent male Catholics who, for centuries, had no children, in contrast to the encouragement given rabbis to marry and reproduce.
14. The theory of how public goods aren't supplied solely by individual action.
15. Armen Alchian's pointing to a different invisible hand (in our later terminology, a filter) than does Adam Smith ("Uncertainty, Evolution, and Economic Theory," *Journal of Political Economy*, 1950, pp. 211-221).
16. F. A. Hayek's explanation of how social cooperation utilizes more knowledge than any individual possesses, through people adjusting their activities on the basis of how other people's similarly adjusted activities affect their local situations and through following examples they are presented with, and thereby creates new institutional forms, general modes of behavior, and so on (*The Constitution of Liberty*, chap. 2).

A rewarding research activity would be to catalog the different modes (and combinations) of invisible-hand explanations, specifying which types of invisible-hand explanations can explain which types of patterns. We can mention here two types of invisible-hand processes by which a pattern *P* can be produced: filtering processes and equilibrium processes. Through filtering processes can pass only things fitting *P*, because processes or structures filter out all non-*P*'s; in equilibrium processes each component part responds or adjusts to "local" conditions, with each adjustment changing the local environment of others close by, so that the sum of the ripples of the local adjustments constitutes or realizes *P* (Some processes of such rippling local adjustments don't come to

an equilibrium pattern, not even a moving one.) There are different ways an equilibrium process can help maintain a pattern, and there also might be a filter that eliminates deviations from the pattern that are too great to be brought back by the internal equilibrating mechanisms. Perhaps the most elegant form of explanation of this sort involves two equilibrium processes, each internally maintaining its pattern in the face of small deviations, and each being a filter to eliminate the large deviations occurring in the other.

We might note in passing that the notion of filtering processes enables us to understand one way in which the position in the philosophy of the social sciences known as methodological individualism might go wrong. If there is a filter that filters out (destroys) all non-*P* *Q*'s, then the explanation of why all *Q*'s are *P*'s (fit the pattern *P*) will refer to this filter. For each particular *Q*, there may be a particular explanation of why *it* is *P*, how it came to be *P*, what maintains it as *P*. But the explanation of why all *Q*'s are *P* will not be the conjunction of these individual explanations, even though these are all the *Q*'s there are, for that is part of what is to be explained. The explanation will refer to the filter. To make this clear, we might imagine that we have *no* explanation of why the individual *Q*'s are *P*'s. It just is an ultimate statistical law (so far as we can tell at any rate) that some *Q*'s are *P*; we even might be unable to discover any stable statistical regularity at all. In this case we would know why all *Q*'s are *P*'s (and know there are *Q*'s, and perhaps even know why there are *Q*'s) without knowing of any *Q*, why it is *P*! The methodological individualist position requires that there be no basic (unreduced) social filtering processes.

#### IS THE DOMINANT PROTECTIVE ASSOCIATION A STATE?

Have we provided an invisible-hand explanation of the state? There are at least two ways in which the scheme of private protective associations might be thought to differ from a minimal state, might fail to satisfy a minimal conception of a state: (1) it appears

to allow some people to enforce their own rights, and (2) it appears not to protect all individuals within its domain. Writers in the tradition of Max Weber<sup>14</sup> treat having a monopoly on the use of force in a geographical area, a monopoly incompatible with private enforcement of rights, as crucial to the existence of a state. As Marshall Cohen points out in an unpublished essay, a state may exist without *actually* monopolizing the use of force it has not authorized others to use; within the boundaries of a state there may exist groups such as the Mafia, the KKK, White Citizens Councils, striking unionists, and Weathermen that also use force. *Claiming* such a monopoly is not sufficient (if *you* claimed it you would not become the state), nor is being its sole claimant a necessary condition. Nor need everyone grant the legitimacy of the state's claim to such monopoly, either because as pacifists they think no one has the right to use force, or because as revolutionaries they believe that a given state lacks this right, or because they believe they are entitled to join in and help out no matter what the state says. Formulating sufficient conditions for the existence of the state thus turns out to be a difficult and messy task.<sup>15</sup>

For our purposes here we need focus only upon a necessary condition that the system of private protective agencies (or any component agency within it) apparently does not satisfy. A state claims a monopoly on deciding who may use force when; it says that only it may decide who may use force and under what conditions; it reserves to itself the sole right to pass on the legitimacy and permissibility of any use of force within its boundaries; furthermore it claims the right to punish all those who violate its claimed monopoly. The monopoly may be violated in two ways: (1) a person may use force though unauthorized by the state to do so, or (2) though not themselves using force a group or person may set themselves up as an alternative authority (and perhaps even claim to be the sole legitimate one) to decide when and by whom the use of force is proper and legitimate. It is unclear whether a state must claim the right to punish the second sort of violator, and doubtful whether any state actually would refrain from punishing a significant group of them within its boundaries. I glide over the issue of what sort of "may," "legitimacy," and "permissibility" is in question. Moral permissibility isn't a matter of

decision, and the state need not be so egomaniacal as to claim the sole right to decide moral questions. To speak of legal permissibility would require, to avoid circularity, that an account of a legal system be offered that doesn't use the notion of the state.

We may proceed, for our purposes, by saying that a necessary condition for the existence of a state is that it (some person or organization) announce that, to the best of its ability (taking into account costs of doing so, the feasibility, the more important alternative things it should be doing, and so forth), it will punish everyone whom it discovers to have used force without its express permission. (This permission may be a particular permission or may be granted via some general regulation or authorization.) This still won't quite do: the state may reserve the right to forgive someone, *ex post facto*; in order to punish they may have not only to discover the "unauthorized" use of force but also prove via a certain specified procedure of proof that it occurred, and so forth. But it enables us to proceed. The protective agencies, it seems, do not make such an announcement, either individually or collectively. *Nor does it seem morally legitimate for them to do so.* So the system of private protective associations, if they perform no morally illegitimate action, appears to lack any monopoly element and so appears not to constitute or contain a state. To examine the question of the monopoly element, we shall have to consider the situation of some group of persons (or some one person) living within a system of private protective agencies who refuse to join any protective society; who insist on judging for themselves whether their rights have been violated, and (if they so judge) on personally enforcing their rights by punishing and/or exacting compensation from those who infringed them.

The second reason for thinking the system described is not a state is that, under it (apart from spillover effects) only those paying for protection get protected; furthermore, differing degrees of protection may be purchased. External economies again to the side, no one pays for the protection of others except as they choose to; no one is required to purchase or contribute to the purchasing of protection for others. Protection and enforcement of people's rights is treated as an economic good to be provided by the market, as are other important goods such as food and clothing. However, under the usual conception of a state, each person living

within (or even sometimes traveling outside) its geographical boundaries gets (or at least, is entitled to get) its protection. Unless some private party donated sufficient funds to cover the costs of such protection (to pay for detectives, police to bring criminals into custody, courts, and prisons), or unless the state found some service it could charge for that would cover these costs,\* one would expect that a state which offered protection so broadly would be redistributive. It would be a state in which some persons paid more so that others could be protected. And indeed the most minimal state seriously discussed by the mainstream of political theorists, the night-watchman state of classical liberal theory, appears to be redistributive in this fashion.<sup>16</sup> Yet how can a protection agency, a business, charge some to provide its product to others? (We ignore things like some partially paying for others because it is too costly for the agency to refine its classification of, and charges to, customers to mirror the costs of the services to them.)

Thus it appears that the dominant protective agency in a territory not only lacks the requisite monopoly over the use of force, but also fails to provide protection for all in its territory; and so the dominant agency appears to fall short of being a state. But these appearances are deceptive.

---

\* I have heard it suggested that the state could finance itself by running a lottery. But since it would have no right to forbid private entrepreneurs from doing the same, why think the state will have any more success in attracting customers in this than in any other competitive business?

# CHAPTER

## 3

# Moral Constraints and the State

### THE MINIMAL STATE AND THE ULTRAMINIMAL STATE

THE night-watchman state of classical liberal theory, limited to the functions of protecting all its citizens against violence, theft, and fraud, and to the enforcement of contracts, and so on, appears to be redistributive.<sup>1</sup> We can imagine at least one social arrangement intermediate between the scheme of private protective associations and the night-watchman state. Since the night-watchman state is often called a minimal state, we shall call this other arrangement the *ultraminimal state*. An ultraminimal state maintains a monopoly over all use of force except that necessary in immediate self-defense, and so excludes private (or agency) retaliation for wrong and exaction of compensation; but it provides protection and enforcement services *only* to those who purchase its protection and enforcement policies. People who don't buy a protection contract from the monopoly don't get protected. The minimal (night-watchman) state is equivalent to the ultraminimal state conjoined with a (clearly redistributive) Friedmanesque voucher

plan, financed from tax revenues.\* Under this plan all people, or some (for example, those in need), are given tax-funded vouchers that can be used only for their purchase of a protection policy from the ultraminimal state.

Since the night-watchman state appears redistributive to the extent that it compels some people to pay for the protection of others, its proponents must explain why this redistributive function of the state is unique. If some redistribution is legitimate in order to protect everyone, why is redistribution not legitimate for other attractive and desirable purposes as well? What rationale specifically selects protective services as the sole subject of legitimate redistributive activities? A rationale, once found, may show that this provision of protective services is *not* redistributive. More precisely, the term "redistributive" applies to types of *reasons* for an arrangement, rather than to an arrangement itself. We might elliptically call an arrangement "redistributive" if its major (only possible) supporting reasons are themselves redistributive. ("Paternalistic" functions similarly.) Finding compelling nonredistributive reasons would cause us to drop this label. Whether we say an institution that takes money from some and gives it to others is redistributive will depend upon *why* we think it does so. Returning stolen money or compensating for violations of rights are *not* redistributive reasons. I have spoken until now of the night-watchman state's *appearing* to be redistributive, to leave open the possibility that nonredistributive types of reasons might be found to justify the provision of protective services for some by others (I explore some such reasons in Chapters 4 and 5 of Part I.)

A proponent of the ultraminimal state may seem to occupy an inconsistent position, even though he avoids the question of what makes protection uniquely suitable for redistributive provision. Greatly concerned to protect rights against violation, he makes this the sole legitimate function of the state; and he protests that all other functions are illegitimate because they themselves involve the violation of rights. Since he accords paramount place to the

\* Milton Friedman, *Capitalism and Freedom* (Chicago: University of Chicago Press, 1962), chap. 6. Friedman's school vouchers, of course, allow a choice about who is to supply the product, and so differ from the protection vouchers imagined here.

protection and nonviolation of rights, how can he support the ultraminimal state, which would seem to leave some persons' rights unprotected or illprotected? How can he support this *in the name of* the nonviolation of rights?

#### MORAL CONSTRAINTS AND MORAL GOALS

This question assumes that a moral concern can function only as a moral *goal*, as an end state for some activities to achieve as their result. It may, indeed, seem to be a necessary truth that "right," "ought," "should," and so on, are to be explained in terms of what is, or is intended to be, productive of the greatest good, with all goals built into the good.<sup>2</sup> Thus it is often thought that what is wrong with utilitarianism (which *is* of this form) is its too narrow conception of good. Utilitarianism doesn't, it is said, properly take rights and their nonviolation into account; it instead leaves them a derivative status. Many of the counterexample cases to utilitarianism fit under this objection, for example, punishing an innocent man to save a neighborhood from a vengeful rampage. But a theory may include in a primary way the nonviolation of rights, yet include it in the wrong place and the wrong manner. For suppose some condition about minimizing the total (weighted) amount of violations of rights is built into the desirable end state to be achieved. We then would have something like a "utilitarianism of rights"; violations of rights (to be *minimized*) merely would replace the total happiness as the relevant end state in the utilitarian structure. (Note that we do not hold the nonviolation of our rights as our sole greatest good or even rank it first lexicographically to exclude trade-offs, if there is some desirable society we would choose to inhabit even though in it some rights of ours sometimes are violated, rather than move to a desert island where we could survive alone.) This still would require us to violate someone's rights when doing so minimizes the total (weighted) amount of the violation of rights in the society. For example, violating someone's rights might deflect others from *their* intended action of gravely violating rights, or might remove their motive for doing so, or might divert their attention, and so on. A

mob rampaging through a part of town killing and burning *will* violate the rights of those living there. Therefore, someone might try to justify his punishing another *he* knows to be innocent of a crime that enraged a mob, on the grounds that punishing this innocent person would help to avoid even greater violations of rights by others, and so would lead to a minimum weighted score for rights violations in the society.

In contrast to incorporating rights into the end state to be achieved, one might place them as side constraints upon the actions to be done: don't violate constraints *C*. The rights of others determine the constraints upon your actions. (A *goal-directed* view with constraints added would be: among those acts available to you that don't violate constraints *C*, act so as to maximize goal *G*. Here, the rights of others would constrain your goal-directed behavior. I do not mean to imply that the correct moral view includes mandatory goals that must be pursued, even within the constraints.) This view differs from one that tries to build the side constraints *C* into the goal *G*. The side-constraint view forbids you to violate these moral constraints in the pursuit of your goals; whereas the view whose objective is to minimize the violation of these rights allows you to violate the rights (the constraints) in order to lessen their total violation in the society.\*

---

\* Unfortunately, too few models of the structure of moral views have been specified heretofore, though there are surely other interesting structures. Hence an argument for a side-constraint structure that consists largely in arguing against an end-state maximization structure is inconclusive, for these alternatives are not exhaustive. (On page 46 we describe a view which fits neither structure happily.) An array of structures must be precisely formulated and investigated; perhaps some novel structure then will seem most appropriate.

The issue of whether a side-constraint view can be put in the form of the goal-without-side-constraint view is a tricky one. One might think, for example, that each person could distinguish in his goal between *his* violating rights and someone else's doing it. Give the former infinite (negative) weight in his goal, and no amount of stopping others from violating rights can outweigh his violating someone's rights. In addition to a component of a goal receiving infinite weight, indexical expressions also appear, for example, "*my* doing something." A careful statement delimiting "constraint views" would exclude these gimmicky ways of transforming side constraints into the form of an end-state view as sufficient to constitute a view as end state. Mathematical methods of transforming a constrained minimization problem into a sequence of unconstrained minimizations of an auxiliary function are presented in Anthony Fiacco and Garth McCormick, *Nonlinear Programming: Sequential Unconstrained Minimi-*



The claim that the proponent of the ultraminimal state is inconsistent, we now can see, assumes that he is a "utilitarian of rights." It assumes that his goal is, for example, to minimize the weighted amount of the violation of rights in the society, and that he should pursue this goal even through means that themselves violate people's rights. Instead, he may place the nonviolation of rights as a constraint upon action, rather than (or in addition to) building it into the end state to be realized. The position held by this proponent of the ultraminimal state will be a consistent one if his conception of rights holds that your being *forced* to contribute to another's welfare violates your rights, whereas someone else's not providing you with things you need greatly, including things essential to the protection of your rights, does not *itself* violate your rights, even though it avoids making it more difficult for someone else to violate them. (That conception will be consistent provided it does not construe the monopoly element of the ultraminimal state as itself a violation of rights.) That it is a consistent position does not, of course, show that it is an acceptable one.

#### WHY SIDE CONSTRAINTS?

Isn't it *irrational* to accept a side constraint *C*, rather than a view that directs minimizing the violations of *C*? (The latter view treats *C* as a condition rather than a constraint.) If nonviolation of *C* is so important, shouldn't that be the goal? How can a concern for the nonviolation of *C* lead to the refusal to violate *C* even when this would prevent other more extensive violations of *C*? What is the rationale for placing the nonviolation of rights as a side constraint upon action instead of including it solely as a goal of one's actions?

Side constraints upon action reflect the underlying Kantian

---

*zation Techniques* (New York: Wiley, 1968). The book is interesting both for its methods and for their limitations in illuminating our area of concern; note the way in which the penalty functions include the constraints, the variation in weights of penalty functions (sec. 7.1), and so on.

The question of whether these side constraints are absolute, or whether they may be violated in order to avoid catastrophic moral horror, and if the latter, what the resulting structure might look like, is one I hope largely to avoid.

principle that individuals are ends and not merely means; they may not be sacrificed or used for the achieving of other ends without their consent. Individuals are inviolable. More should be said to illuminate this talk of ends and means. Consider a prime example of a means, a tool. There is no side constraint on how we may use a tool, other than the moral constraints on how we may use it upon others. There are procedures to be followed to preserve it for future use ("don't leave it out in the rain"), and there are more and less efficient ways of using it. But there is no limit on what we may do to it to best achieve our goals. Now imagine that there was an overrideable constraint *C* on some tool's use. For example, the tool might have been lent to you only on the condition that *C* not be violated unless the gain from doing so was above a certain specified amount, or unless it was necessary to achieve a certain specified goal. Here the object is not *completely* your tool, for use according to your wish or whim. But it is a tool nevertheless, even with regard to the overrideable constraint. If we add constraints on its use that may not be overridden, then the object may not be used as a tool *in those ways*. *In those respects*, it is not a tool at all. Can one add enough constraints so that an object cannot be used as a tool at all, in *any* respect?

Can behavior toward a person be constrained so that he is not to be used for any end except as he chooses? This is an impossibly stringent condition if it requires everyone who provides us with a good to approve positively of every use to which we wish to put it. Even the requirement that he merely should not object to any use we plan would seriously curtail bilateral exchange, not to mention sequences of such exchanges. It is sufficient that the other party stands to gain enough from the exchange so that he is willing to go through with it, even though he objects to one or more of the uses to which you shall put the good. Under such conditions, the other party is not being used solely as a means, in that respect. Another party, however, who would not choose to interact with you if he knew of the uses to which you *intend* to put his actions or good, *is* being used as a means, even if he receives enough to choose (in his ignorance) to interact with you. ("All along, you were just *using* me" can be said by someone who chose to interact only because he was ignorant of another's goals and of the uses to which he himself would be put.) Is it morally incumbent upon

someone to reveal his intended uses of an interaction if he has good reason to believe the other would refuse to interact if he knew? Is he *using* the other person, if he does not reveal this? And what of the cases where the other does not choose to be of use at all? In getting pleasure from seeing an attractive person go by, does one use the other solely as a means? <sup>3</sup> Does someone so use an object of sexual fantasies? These and related questions raise very interesting issues for moral philosophy; but not, I think, for political philosophy.

Political philosophy is concerned only with *certain* ways that persons may not use others; primarily, physically aggressing against them. A specific side constraint upon action toward others expresses the fact that others may not be used in the specific ways the side constraint excludes. Side constraints express the inviolability of others, in the ways they specify. These modes of inviolability are expressed by the following injunction: "Don't use people in specified ways." An end-state view, on the other hand, would express the view that people are ends and not merely means (if it chooses to express this view at all), by a different injunction: "Minimize the use in specified ways of persons as means." Following this precept itself may involve using someone as a means in one of the ways specified. Had Kant held this view, he would have given the second formula of the categorical imperative as, "So act as to minimize the use of humanity simply as a means," rather than the one he actually used: "Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end." <sup>4</sup>

Side constraints express the inviolability of other persons. But why may not one violate persons for the greater social good? Individually, we each sometimes choose to undergo some pain or sacrifice for a greater benefit or to avoid a greater harm: we go to the dentist to avoid worse suffering later; we do some unpleasant work for its results; some persons diet to improve their health or looks; some save money to support themselves when they are older. In each case, some cost is borne for the sake of the greater overall good. Why not, *similarly*, hold that some persons have to bear some costs that benefit other persons more, for the sake of the overall social good? But there is no *social entity* with a good that

undergoes some sacrifice for its own good. There are only individual people, different individual people, with their own individual lives. Using one of these people for the benefit of others, uses him and benefits the others. Nothing more. What happens is that something is done to him for the sake of others. Talk of an overall social good covers this up. (Intentionally?) To use a person in this way does not sufficiently respect and take account of the fact that he is a separate person,<sup>5</sup> that his is the only life he has. *He* does not get some overbalancing good from his sacrifice, and no one is entitled to force this upon him—least of all a state or government that claims his allegiance (as other individuals do not) and that therefore scrupulously must be *neutral* between its citizens.

#### LIBERTARIAN CONSTRAINTS

The moral side constraints upon what we may do, I claim, reflect the fact of our separate existences. They reflect the fact that no moral balancing act can take place among us; there is no moral outweighing of one of our lives by others so as to lead to a greater overall *social* good. There is no justified sacrifice of some of us for others. This root idea, namely, that there are different individuals with separate lives and so no one may be sacrificed for others, underlies the existence of moral side constraints, but it also, I believe, leads to a libertarian side constraint that prohibits aggression against another.

The stronger the force of an end-state maximizing view, the more powerful must be the root idea capable of resisting it that underlies the existence of moral side constraints. Hence the more seriously must be taken the existence of distinct individuals who are not resources for others. An underlying notion sufficiently powerful to support moral side constraints against the powerful intuitive force of the end-state maximizing view will suffice to derive a libertarian constraint on aggression against another. Anyone who rejects *that particular* side constraint has three alternatives: (1) he must reject *all* side constraints; (2) he must produce a different explanation of why there are moral side constraints rather than simply a goal-directed maximizing structure, an explanation

that doesn't itself entail the libertarian side constraint; or (3) he must accept the strongly put root idea about the separateness of individuals and yet claim that initiating aggression against another is compatible with this root idea. Thus we have a promising sketch of an argument from moral form to moral content: the form of morality includes *F* (moral side constraints); the best explanation<sup>6</sup> of morality's being *F* is *p* (a strong statement of the distinctness of individuals); and from *p* follows a particular moral content, namely, the libertarian constraint. The particular moral content gotten by this argument, which focuses upon the fact that there are distinct individuals each with his *own* life to lead, will not be the *full* libertarian constraint. It will prohibit sacrificing one person to benefit another. Further steps would be needed to reach a prohibition on paternalistic aggression: using or threatening force for the benefit of the person against whom it is wielded. For this, one must focus upon the fact that there are distinct individuals, each with his own life *to lead*.

A nonaggression principle is often held to be an appropriate principle to govern relations among nations. What difference is there supposed to be between sovereign individuals and sovereign nations that makes aggression permissible among individuals? Why may individuals jointly, through their government, do to someone what no nation may do to another? If anything, there is a stronger case for nonaggression among individuals; unlike nations, they do not contain as parts individuals that others legitimately might intervene to protect or defend.

I shall not pursue here the details of a principle that prohibits physical aggression, except to note that it does not prohibit the use of force in defense against another party who is a threat, even though he is innocent and deserves no retribution. An *innocent threat* is someone who innocently is a causal agent in a process such that he would be an aggressor had he chosen to become such an agent. If someone picks up a third party and throws him at you down at the bottom of a deep well, the third party is innocent and a threat; had he chosen to launch himself at you in that trajectory he would be an aggressor. Even though the falling person would survive his fall onto you, may you use your ray gun to disintegrate the falling body before it crushes and kills you? Libertarian prohibitions are usually formulated so as to forbid using violence on in-

nocent persons. But innocent threats, I think, are another matter to which different principles must apply.<sup>7</sup> Thus, a full theory in this area also must formulate the *different* constraints on response to innocent threats. Further complications concern *innocent shields of threats*, those innocent persons who themselves are nonthreats but who are so situated that they will be damaged by the only means available for stopping the threat. Innocent persons strapped onto the front of the tanks of aggressors so that the tanks cannot be hit without also hitting them are innocent shields of threats. (Some uses of force on people to get at an aggressor do not act upon innocent shields of threats; for example, an aggressor's innocent child who is tortured in order to get the aggressor to stop wasn't *shielding* the parent.) May one knowingly injure innocent shields? *If* one may attack an aggressor and injure an innocent shield, may the innocent shield fight back in self-defense (supposing that he cannot move against or fight the aggressor)? Do we get two persons battling each other in self-defense? Similarly, if you use force against an innocent threat to you, do you thereby become an innocent threat to him, so that he may now justifiably use additional force against you (supposing that he can do this, yet cannot prevent his original threateningness)? I tiptoe around these incredibly difficult issues here, merely noting that a view that says it makes nonaggression central must resolve them explicitly at some point.

#### CONSTRAINTS AND ANIMALS

We can illuminate the status and implications of moral side constraints by considering living beings for whom such stringent side constraints (or any at all) usually are not considered appropriate: namely, nonhuman animals. Are there any limits to what we may do to animals? Have animals the moral status of mere *objects*? Do some purposes fail to entitle us to impose great costs on animals? What entitles us to use them at all?

Animals count for something. Some higher animals, at least, ought to be given some weight in people's deliberations about what to do. It is difficult to *prove* this. (It is also difficult to prove

that people count for something!) We first shall adduce particular examples, and then arguments. If you felt like snapping your fingers, perhaps to the beat of some music, and you knew that by some strange causal connection your snapping your fingers would cause 10,000 contented, unowned cows to die after great pain and suffering, or even painlessly and instantaneously, would it be perfectly all right to snap your fingers? Is there some reason why it would be morally wrong to do so?

Some say people should not do so because such acts brutalize them and make them more likely to take the lives of *persons*, solely for pleasure. These acts that are morally unobjectionable in themselves, they say, have an undesirable moral spillover. (Things then would be different if there were no possibility of such spillover—for example, for the person who knows himself to be the last person on earth.) But *why should* there be such a spillover? If it is, in itself, perfectly all right to do anything at all to animals for any reason whatsoever, then provided a person realizes the clear line between animals and persons and keeps it in mind as he acts, why should killing animals tend to brutalize him and make him more likely to harm or kill persons? Do butchers commit more murders? (Than other persons who have knives around?) If I enjoy hitting a baseball squarely with a bat, does this significantly increase the danger of my doing the same to someone's head? Am I not capable of understanding that people differ from baseballs, and doesn't this understanding stop the spillover? Why should things be different in the case of animals? To be sure, it is an empirical question whether spillover does take place or not; but there *is* a puzzle as to why it should, at least among readers of this essay, sophisticated people who are capable of drawing distinctions and differentially acting upon them.

If some animals count for something, which animals count, how much do they count, and how can this be determined? Suppose (as I believe the evidence supports) that *eating* animals is not necessary for *health* and is not less expensive than alternate equally healthy diets available to people in the United States. The gain, then, from the eating of animals is pleasures of the palate, gustatory delights, varied tastes. I would not claim that these are not truly pleasant, delightful, and interesting. The question is: do they, or rather does the marginal addition in them gained by eating ani-

mals rather than only nonanimals, *outweigh* the moral weight to be given to animals' lives and pain? Given that animals are to count for *something*, is the *extra* gain obtained by eating them rather than nonanimal products greater than the moral cost? How might these questions be decided?

We might try looking at comparable cases, extending whatever judgments we make on those cases to the one before us. For example, we might look at the case of hunting, where I assume that it's not all right to hunt and kill animals merely for the fun of it. Is hunting a special case, because its *object* and what provides the fun is the chasing and maiming and death of animals? Suppose then that I enjoy swinging a baseball bat. It happens that in front of the only place to swing it stands a cow. Swinging the bat unfortunately would involve smashing the cow's head. But I wouldn't get fun from doing *that*; the pleasure comes from exercising my muscles, swinging well, and so on. It's unfortunate that as a side effect (not a means) of my doing this, the animal's skull gets smashed. To be sure, I could forego swinging the bat, and instead bend down and touch my toes or do some other exercise. But this wouldn't be as enjoyable as swinging the bat; I won't get as much fun, pleasure, or delight out of it. So the question is: would it be all right for me to swing the bat in order to get the *extra* pleasure of swinging it as compared to the best available alternative activity that does not involve harming the animal? Suppose that it is not merely a question of foregoing today's special pleasures of bat swinging; suppose that each day the same situation arises with a different animal. Is there some principle that would allow killing and eating animals for the additional pleasure this brings, yet would not allow swinging the bat for the extra pleasure it brings? What could that principle be like? (Is this a better parallel to eating meat? The animal is killed to get a bone out of which to make the best sort of bat to use; bats made out of other material don't give quite the same pleasure. Is it all right to kill the animal to obtain the *extra* pleasure that using a bat made out of its bone would bring? Would it be morally more permissible if you could hire someone to do the killing for you?)

Such examples and questions might help someone to see what sort of line *he* wishes to draw, what sort of position he wishes to take. They face, however, the usual limitations of consistency

arguments; they do not say, once a conflict is shown, which view to change. After failing to devise a principle to distinguish swinging the bat from killing and eating an animal, you might decide that it's really all right, after all, to swing the bat. Furthermore, such appeal to similar cases does not greatly help us to assign precise moral weight to different sorts of animals. (We further discuss the difficulties in forcing a moral conclusion by appeal to examples in Chapter 9.)

My purpose here in presenting these examples is to pursue the notion of moral side constraints, not the issue of eating animals. Though I should say that in my view the extra benefits Americans today can gain from eating animals do *not* justify doing it. So we shouldn't. One ubiquitous argument, not unconnected with side constraints, deserves mention: because people eat animals, they raise more than otherwise would exist without this practice. To exist for a while is better than never to exist at all. So (the argument concludes) the animals are better off because we have the practice of eating them. Though this is not our object, fortunately it turns out that we really, all along, benefit them! (If tastes changed and people no longer found it enjoyable to eat animals, should those concerned with the welfare of animals steel themselves to an unpleasant task and continue eating them?) I trust I shall not be misunderstood as saying that animals are to be given the same moral weight as people if I note that the parallel argument about people would not look very convincing. We can imagine that population problems lead every couple or group to limit their children to some number fixed in advance. A given couple, having reached the number, proposes to have an additional child and dispose of it at the age of three (or twenty-three) by sacrificing it or using it for some gastronomic purpose. In justification, they note that the child will not exist at all if this is not allowed; and surely it is better for it to exist for some number of years. However, once a person exists, not everything compatible with his overall existence being a net plus can be done, even by those who created him. An existing person has claims, even against those whose purpose in creating him was to violate those claims. It would be worthwhile to pursue moral objections to a system that permits parents to do anything whose permissibility is necessary for their choosing to have the child, that also leaves the child better off than if it hadn't

been born.<sup>8</sup> (Some will think the only objections arise from difficulties in accurately administering the permission.) Once they exist, animals too may have claims to certain treatment. These claims may well carry less weight than those of people. But the fact that some animals were brought into existence only because someone wanted to do something that would violate one of these claims does not show that the claim doesn't exist at all.

Consider the following (too minimal) position about the treatment of animals. So that we can easily refer to it, let us label this position "utilitarianism for animals, Kantianism for people." It says: (1) maximize the total happiness of all living beings; (2) place stringent side constraints on what one may do to human beings. Human beings may not be used or sacrificed for the benefit of others; animals may be used or sacrificed for the benefit of other people or animals *only if* those benefits are greater than the loss inflicted. (This inexact statement of the utilitarian position is close enough for our purposes, and it can be handled more easily in discussion.) One may proceed only if the total utilitarian benefit is greater than the utilitarian loss inflicted on the animals. This utilitarian view counts animals as much as normal utilitarianism does persons. Following Orwell, we might summarize this view as: *all animals are equal but some are more equal than others.* (None may be sacrificed except for a greater total benefit; but persons may not be sacrificed at all, or only under far more stringent conditions, and never for the benefit of nonhuman animals. I mean (1) above merely to exclude sacrifices which do not meet the utilitarian standard, not to mandate a utilitarian goal. We shall call this position negative utilitarianism.)

We can now direct arguments for animals counting for something to holders of different views. To the "Kantian" moral philosopher who imposes stringent side constraints on what may be done to a person, we can say:

You hold utilitarianism inadequate because it allows an individual to be sacrificed to and for another, and so forth, thereby neglecting the stringent limitations on how one legitimately may behave toward persons. But *could* there be anything morally intermediate between persons and stones, something without such stringent limitations on its treatment, yet not to be treated merely as an object? One would expect that by subtracting or diminishing some features of persons, we would get this in-

intermediate sort of being. (Or perhaps beings of intermediate moral status are gotten by subtracting some of our characteristics and adding others very different from ours.)

Plausibly, animals are the intermediate beings, and utilitarianism is the intermediate position. We may come at the question from a slightly different angle. Utilitarianism assumes both that happiness is all that matters morally and that all beings are interchangeable. This conjunction does not hold true of persons. But isn't (negative) utilitarianism true of whatever beings the conjunction does hold for, and doesn't it hold for animals?

To the utilitarian we may say:

If only the experiences of pleasure, pain, happiness, and so on (and the capacity for these experiences) are morally relevant, then animals must be counted in moral calculations to the extent they *do* have these capacities and experiences. Form a matrix where the rows represent alternative policies or actions, the columns represent different individual organisms, and each entry represents the utility (net pleasure, happiness) the policy will lead to for the organism. The utilitarian theory evaluates each policy by the sum of the entries in its row and directs us to perform an action or adopt a policy whose sum is maximal. Each column is weighted equally and counted once, be it that of a person or a nonhuman animal. Though the structure of the view treats them equally, animals might be less important in the decisions because of facts about them. If animals have less capacity for pleasure, pain, happiness than humans do, the matrix entries in animals' columns will be lower generally than those in people's columns. In this case, they will be less important factors in the ultimate decisions to be made.

A utilitarian would find it difficult to deny animals this kind of equal consideration. On what grounds could he consistently distinguish persons' happiness from that of animals, to count only the former? Even if experiences don't get entered in the utility matrix unless they are above a certain threshold, surely *some* animal experiences are greater than some people's experiences that the utilitarian wishes to count. (Compare an animal's being burned alive unanesthetized with a person's mild annoyance.) Bentham, we may note, *does* count animals' happiness equally in just the way we have explained.<sup>9</sup>

Under "utilitarianism for animals, Kantianism for people," animals will be used for the gain of other animals and persons, but persons will never be used (harmed, sacrificed) against their will, for the gain of animals. Nothing may be inflicted upon persons for

the sake of animals. (Including penalties for violating laws against cruelty to animals?) Is this an acceptable consequence? Can't one save 10,000 animals from excruciating suffering by inflicting some slight discomfort on a person who did not cause the animals' suffering? One may feel the side constraint is not absolute when it is *people* who can be saved from excruciating suffering. So perhaps the side constraint also relaxes, though not as much, when animals' suffering is at stake. The thoroughgoing utilitarian (for animals *and* for people, combined in one group) goes further and holds that, *ceteris paribus*, we may inflict some suffering on a person to avoid a (slightly) greater suffering of an animal. This permissive principle seems to me to be unacceptably strong, even when the purpose is to avoid greater suffering to a *person!*

Utilitarian theory is embarrassed by the possibility of utility monsters who get enormously greater gains in utility from any sacrifice of others than these others lose. For, unacceptably, the theory seems to require that we all be sacrificed in the monster's maw, in order to increase total utility. Similarly if people are utility devourers with respect to animals, always getting greatly counterbalancing utility from each sacrifice of an animal, we may feel that "utilitarianism for animals, Kantianism for people," in requiring (or allowing) that almost always animals be sacrificed, makes animals too subordinate to persons.

Since it counts only the happiness and suffering of animals, would the utilitarian view hold it all right to kill animals painlessly? Would it be all right, on the utilitarian view, to kill *people* painlessly, in the night, provided one didn't first announce it? Utilitarianism is notoriously inept with decisions where the *number* of persons is at issue. (In this area, it must be conceded, eptness is hard to come by.) Maximizing the total happiness requires continuing to add persons so long as their net utility is positive and is sufficient to counterbalance the loss in utility their presence in the world causes others. Maximizing the average utility allows a person to kill everyone else if that would make him ecstatic, and so happier than average. (Don't say he shouldn't because after his death the average would drop lower than if he didn't kill all the others.) Is it all right to kill someone provided you immediately substitute another (by having a child or, in science-fiction fashion, by creating a full-grown person) who will be as happy as the rest

of the life of the person you killed? After all, there would be no net diminution in total utility, or even any change in its profile of distribution. Do we forbid murder only to prevent feelings of *worry* on the part of potential victims? (And how does a utilitarian explain what it is they're worried about, and would he really base a policy on what he must hold to be an irrational fear?) Clearly, a utilitarian needs to supplement his view to handle such issues; perhaps he will find that the supplementary theory becomes the main one, relegating utilitarian considerations to a corner.

But isn't utilitarianism at least adequate for animals? I think not. But if not only the animals' felt experiences are relevant, what else is? Here a tangle of questions arises. How much does an animal's life have to be respected once it's alive, and how can we decide this? Must one also introduce some notion of a nondegraded existence? Would it be all right to use genetic-engineering techniques to breed natural slaves who would be contented with their lots? Natural animal slaves? Was that the domestication of animals? Even for animals, utilitarianism won't do as the whole story, but the thicket of questions daunts us.

#### THE EXPERIENCE MACHINE

There are also substantial puzzles when we ask what matters other than how *people's* experiences feel "from the inside." Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. Should you plug into this machine for life, preprogramming your life's experiences? If you are worried about missing out on desirable experiences, we can suppose that business enterprises have researched thoroughly the lives of many others. You can pick and choose from their large library or smorgasbord of such experiences, selecting your life's experiences for, say, the next two years. After two years have passed, you will have ten minutes or ten hours out of the tank, to select the experiences of your *next*

two years. Of course, while in the tank you won't know that you're there; you'll think it's all actually happening. Others can also plug in to have the experiences they want, so there's no need to stay unplugged to serve them. (Ignore problems such as who will service the machines if everyone plugs in.) Would you plug in? *What else can matter to us, other than how our lives feel from the inside?* Nor should you refrain because of the few moments of distress between the moment you've decided and the moment you're plugged. What's a few moments of distress compared to a lifetime of bliss (if that's what you choose), and why feel any distress at all if your decision *is* the best one?

What does matter to us in addition to our experiences? First, we want to *do* certain things, and not just have the experience of doing them. In the case of certain experiences, it is only because first we want to do the actions that we want the experiences of doing them or thinking we've done them. (But *why* do we want to do the activities rather than merely to experience them?) A second reason for not plugging in is that we want to *be* a certain way, to be a certain sort of person. Someone floating in a tank is an indeterminate blob. There is no answer to the question of what a person is like who has long been in the tank. Is he courageous, kind, intelligent, witty, loving? It's not merely that it's difficult to tell; there's no way he is. Plugging into the machine is a kind of suicide. It will seem to some, trapped by a picture, that nothing about what we are like can matter except as it gets reflected in our experiences. But should it be surprising that what *we are* is important to us? Why should we be concerned only with how our time is filled, but not with what we are?

Thirdly, plugging into an experience machine limits us to a man-made reality, to a world no deeper or more important than that which people can construct.<sup>10</sup> There is no *actual* contact with any deeper reality, though the experience of it can be simulated. Many persons desire to leave themselves open to such contact and to a plumbing of deeper significance.\* This clarifies the intensity

\* Traditional religious views differ on the *point* of contact with a transcendent reality. Some say that contact yields eternal bliss or Nirvana, but they have not distinguished this sufficiently from merely a *very* long run on the experience machine. Others think it is intrinsically desirable to do the will of a higher

of the conflict over psychoactive drugs, which some view as mere local experience machines, and others view as avenues to a deeper reality; what some view as equivalent to surrender to the experience machine, others view as following one of the reasons *not* to surrender!

We learn that something matters to us in addition to experience by imagining an experience machine and then realizing that we would not use it. We can continue to imagine a sequence of machines each designed to fill lacks suggested for the earlier machines. For example, since the experience machine doesn't meet our desire to *be* a certain way, imagine a transformation machine which transforms us into whatever sort of person we'd like to be (compatible with our staying us). Surely one would not use the transformation machine to become as one would wish, and thereupon plug into the experience machine! \* So something matters in addition to one's experiences *and* what one is like. Nor is the reason merely that one's experiences are unconnected with what one is like. For the experience machine might be limited to provide only experiences possible to the sort of person plugged in. Is it that we want to make a difference in the world? Consider then the result machine, which produces in the world any result you would produce and injects your vector input into any joint activity. We shall not pursue here the fascinating details of these or other machines. What is most disturbing about them is their living of our lives for us. Is it misguided to search for *particular* additional

---

being which created us all, though presumably no one would think this if we discovered we had been created as an object of amusement by some superpowerful child from another galaxy or dimension. Still others imagine an eventual merging with a higher reality, leaving unclear its desirability, or where that merging leaves *us*.

\* Some wouldn't use the transformation machine at all; it seems like *cheating*. But the one-time use of the transformation machine would not remove all challenges; there would still be obstacles for the new us to overcome, a new plateau from which to strive even higher. And is this plateau any the less earned or deserved than that provided by genetic endowment and early childhood environment? But if the transformation machine could be used indefinitely often, so that we could accomplish anything by pushing a button to transform ourselves into someone who could do it easily, there would remain no limits we *need* to strain against or try to transcend. Would there be anything left *to do*? Do some theological views place God outside of time because an omniscient omnipotent being couldn't fill up his days?

functions beyond the competence of machines to do for us? Perhaps what we desire is to live (an active verb) ourselves, in contact with reality. (And this, machines cannot do *for us*.) Without elaborating on the implications of this, which I believe connect surprisingly with issues about free will and causal accounts of knowledge, we need merely note the intricacy of the question of what matters *for people* other than their experiences. Until one finds a satisfactory answer, and determines that this answer does not *also* apply to animals, one cannot reasonably claim that only the felt experiences of animals limit what we may do to them.

#### UNDERDETERMINATION OF MORAL THEORY

What about persons distinguishes them from animals, so that stringent constraints apply to how persons may be treated, yet not to how animals may be treated? <sup>11</sup> Could beings from another galaxy stand to *us* as it is usually thought we do to animals, and if so, would they be justified in treating us as means à la utilitarianism? Are organisms arranged on some ascending scale, so that any may be sacrificed or caused to suffer to achieve a greater total benefit for those not lower on the scale? \* Such an elitist hierarchical view would distinguish three moral statuses (forming an interval partition of the scale):

*Status 1:* The being may not be sacrificed, harmed, and so on, for any other organism's sake.

*Status 2:* The being may be sacrificed, harmed, and so on, only for the sake of beings higher on the scale, but not for the sake of beings at the same level.

---

\* We pass over the difficulties about deciding *where* on the scale to place an organism, and about particular interspecies comparisons. How is it to be decided where on the scale a species goes? Is an organism, if defective, to be placed at its species level? Is it an anomaly that it might be impermissible to treat two currently identical organisms similarly (they might even be identical in future and past capacities as well), because one is a normal member of one species and the other is a subnormal member of a species higher on the scale? And the problems of intraspecies interpersonal comparisons pale before those of interspecies comparisons.



*Status 3*: The being may be sacrificed, harmed, and so on, for the sake of other beings at the same or higher levels on the scale.

If animals occupy status 3 and we occupy status 1, what occupies status 2? Perhaps *we* occupy status 2! Is it morally forbidden to use people as means for the benefit of others, or is it only forbidden to use them for the sake of other *people*, that is, for beings at the same level? \* Do ordinary views include the possibility of more than one significant moral divide (like that between persons and animals), and *might one come on the other side of human beings?* Some theological views hold that God is permitted to sacrifice people for his own purposes. We also might imagine people encountering beings from another planet who traverse in their childhood whatever "stages" of moral development our developmental psychologists can identify. These beings claim that they all continue on through fourteen further sequential stages, each being necessary to enter the next one. However, they cannot explain to us (primitive as we are) the content and modes of reasoning of these later stages. These beings claim that we may be sacrificed for their well-being, or at least in order to preserve their higher capacities. They say that they see the truth of this now that they are in their moral maturity, though they didn't as children at what is our highest level of moral development. (A story like this, perhaps, reminds us that a sequence of developmental stages, each a precondition for the next, may after some point deteriorate rather than progress. It would be no recommendation of senility to point out that in order to reach it one must have passed first through other stages.) Do

\* Some would say that here we have a teleological view giving human beings infinite worth relative to other human beings. But a teleological theory that maximizes total value will not prohibit the sacrifice of some people for the sake of other people. Sacrificing some for others wouldn't produce a net gain, but there wouldn't be a net loss either. Since a teleological theory that gives each person's life equal weight excludes only a lowering of total value (to require that each act produce a *gain* in total value would exclude neutral acts), it *would allow* the sacrifice of one person for another. Without gimmicky devices similar to those mentioned earlier, for example, using indexical expressions in the infinitely weighted goals, or giving some goals (representing the constraints) an infinite weight of a *higher* order of infinity than others (even this won't quite do, and the details are very messy), views embodying a status 2 do not seem to be representable as teleological. This illustrates our earlier remark that "teleological" and "side constraint" do not exhaust the possible structures for a moral view.

our moral views permit our sacrifice for the sake of these beings' higher capacities, including their moral ones? This decision is not easily disentangled from the epistemological effects of contemplating the existence of such moral authorities who differ from us, while we admit that, being fallible, we may be wrong. (A similar effect would obtain even if we happened not to know which view of the matter these other beings actually held.)

Beings who occupy the intermediate status 2 will be sacrificeable, but *not* for the sake of beings at the same or lower levels. If they never encounter or know of or affect beings higher in the hierarchy, then *they* will occupy the highest level for every situation they actually encounter and deliberate over. It will be as if an *absolute* side constraint prohibits their being sacrificed for any purpose. Two very different moral theories, the elitist hierarchical theory placing people in status 2 and the absolute-side-constraint theory, yield exactly the same moral judgments for the situations people actually have faced and account equally well for (almost) all of the moral judgments we have made. ("Almost all," because we make judgments about hypothetical situations, and these may include some involving "superbeings" from another planet.) This is not the philosopher's vision of two alternative theories accounting equally well for all of the *possible* data. Nor is it merely the claim that by various gimmicks a side-constraint view can be put into the form of a maximizing view. Rather, the two alternative theories account for all of the actual data, the data about cases we have encountered heretofore; yet they diverge significantly for certain other hypothetical situations.

It would not be surprising if we found it difficult to decide which theory to believe. For we have not been obliged to think about these situations; they are not the situations that shaped our views. Yet the issues do not concern merely whether superior beings may sacrifice us for their sakes. They also concern what *we* ought to do. For if there are other such beings, the elitist hierarchical view does *not* collapse into the "Kantian" side-constraint view, as far as *we* are concerned. A person may not sacrifice one of his fellows for his own benefit or that of another of his fellows, but may he sacrifice one of his fellows for the benefit of the higher beings? (We also will be interested in the question of whether the higher beings may sacrifice us for their own benefit.)

## WHAT ARE CONSTRAINTS BASED UPON?

Such questions do not press upon us as practical problems (yet?), but they force us to consider fundamental issues about the foundations of our moral views: first, is our moral view a side-constraint view, or a view of a more complicated hierarchical structure; and second, in virtue of precisely what characteristics of persons are there moral constraints on how they may treat each other or be treated? We also want to understand *why* these characteristics connect with these constraints. (And, perhaps, we want these characteristics not to be had by animals; or not had by them in as high a degree.) It would appear that a person's characteristics, by virtue of which others are constrained in their treatment of him, must themselves be valuable characteristics. How else are we to understand why something so valuable emerges from them? (This natural assumption is worth further scrutiny.)

The traditional proposals for the important individuating characteristic connected with moral constraints are the following: sentient and self-conscious; rational (capable of using abstract concepts, not tied to responses to immediate stimuli); possessing free will; being a moral agent capable of guiding its behavior by moral principles and capable of engaging in mutual limitation of conduct; having a soul. Let us ignore questions about how these notions are precisely to be understood, and whether the characteristics are possessed, and possessed uniquely, by man, and instead seek their connection with moral constraints on others. Leaving aside the last on the list, each of them seems insufficient to forge the requisite connection. Why is the fact that a being is very smart or foresightful or has an I.Q. above a certain threshold a reason to limit specially how we treat it? Would beings even more intelligent than we have the right not to limit themselves with regard to us? Or, what is the significance of any purported crucial threshold? If a being is capable of choosing autonomously among alternatives, is there some reason to *let it* do so? Are autonomous choices intrinsically good? If a being could make only once an autonomous choice, say between flavors of ice cream on a particular occasion, and would forget immediately afterwards, would there

be strong reasons to allow it to choose? That a being can agree with others to mutual rule-governed limitations on conduct shows that it *can* observe limits. But it does not show which limits should be observed toward it ("no abstaining from murdering it"?), or why any limits should be observed at all.

An intervening variable *M* is needed for which the listed traits are individually necessary, *perhaps* jointly sufficient (at least we should be able to see what needs to be added to obtain *M*), and which has a perspicuous and convincing connection to moral constraints on behavior toward someone with *M*. Also, in the light of *M*, we should be in a position to see why others have concentrated on the traits of rationality, free will, and moral agency. This will be easier if these traits are not merely necessary conditions for *M* but also are important components of *M* or important means to *M*.

But haven't we been unfair in treating rationality, free will, and moral agency individually and separately? In conjunction, don't they add up to something whose significance is clear: a being able to formulate long-term plans for its life, able to consider and decide on the basis of abstract principles or considerations it formulates to itself and hence not merely the plaything of immediate stimuli, a being that limits its own behavior in accordance with some principles or picture it has of what an appropriate life is for itself and others, and so on. However, this exceeds the three listed traits. We can distinguish theoretically between long-term planning and an overall conception of a life that guides particular decisions, and the three traits that are their basis. For a being could possess these three traits and yet also have built into it some particular barrier that prevents it from operating in terms of an overall conception of its life and what it is to add up to. So let us add, as an additional feature, the ability to regulate and guide its life in accordance with some overall conception it chooses to accept. Such an overall conception, and knowing how we are doing in terms of it, is important to the kind of goals we formulate for ourselves and the kind of beings we are. Think how different we would be (and how differently it would be legitimate to treat us) if we all were amnesiacs, forgetting each evening as we slept the happenings of the preceding day. Even if by accident someone were to pick up each day where he left off the previous day, living

in accordance with a coherent conception an aware individual might have chosen, he still would not be leading the other's sort of life. His life would parallel the other life, but it would not be integrated in the same way.

What is the moral importance of this additional ability to form a picture of one's whole life (or at least of significant chunks of it) and to act in terms of some overall conception of the life one wishes to lead? Why not interfere with someone else's shaping of his own life? (And what of those not actively shaping their lives, but drifting with the forces that play upon them?) One might note that anyone might come up with the pattern of life you would wish to adopt. Since one cannot predict in advance that someone won't, it is in your self-interest to allow another to pursue his conception of his life as he sees it; you may learn (to emulate or avoid or modify) from his example. This prudential argument seems insufficient.

I conjecture that the answer is connected with that elusive and difficult notion: the meaning of life. A person's shaping his life in accordance with some overall plan is his way of giving meaning to his life; only a being with the capacity to so shape his life can have or strive for meaningful life. But even supposing that we could elaborate and clarify this notion satisfactorily, we would face many difficult questions. Is the capacity to shape a life itself the capacity to have (or strive for?) a life with meaning, or is something else required? (For ethics, might the content of the attribute of having a soul simply be that the being strives, or is capable of striving, to give meaning to its life?) Why are there constraints on how we may treat beings shaping their lives? Are certain modes of treatment incompatible with their having meaningful lives? And even if so, why not destroy meaningful lives? Or, why not replace "happiness" with "meaningfulness" within utilitarian theory, and maximize the total "meaningfulness" score of the persons of the world? Or does the notion of the meaningfulness of a life enter into ethics in a different fashion? This notion, we should note, has the right "feel" as something that might help to bridge an "is-ought" gap; it appropriately seems to straddle the two. Suppose, for example, that one could show that if a person acted in certain ways his life would be meaningless. Would this be a hypothetical or

a categorical imperative? Would one need to answer the further question: "But why shouldn't my life be meaningless?" Or, suppose that acting in a certain way toward others was itself a way of granting that one's own life (and those very actions) was meaningless. Mightn't this, resembling a pragmatic contradiction, lead at least to a status 2 conclusion of side constraints in behavior to all other human beings? I hope to grapple with these and related issues on another occasion.

#### THE INDIVIDUALIST ANARCHIST

We have surveyed the important issues underlying the view that moral side constraints limit how people may behave to each other, and we may return now to the private protection scheme. A system of private protection, even when one protective agency is dominant in a geographical territory, appears to fall short of a state. It apparently does not provide protection for everyone in its territory, as does a state, and it apparently does not possess or claim the sort of monopoly over the use of force necessary to a state. In our earlier terminology, it apparently does not constitute a minimal state, and it apparently does not even constitute an ultraminimal state.

These very ways in which the dominant protective agency or association in a territory apparently falls short of being a state provide the focus of the individualist anarchist's complaint *against* the state. For he holds that when the state monopolizes the use of force in a territory and punishes others who violate its monopoly, and when the state provides protection for everyone by forcing some to purchase protection for others, it violates moral side constraints on how individuals may be treated. Hence, he concludes, the state itself is intrinsically immoral. The state grants that under some circumstances it is legitimate to punish persons who violate the rights of others, for it itself does so. How then does it arrogate to itself the right to forbid private exaction of justice by other nonaggressive individuals whose rights have been violated? *What* right does the private exacter of justice violate that is not violated

also by the state when it punishes? When a group of persons constitute themselves as the state and begin to punish, *and forbid others from doing likewise*, is there some right these others would violate that they themselves do not? By what right, then, can the state and its officials claim a unique right (a privilege) with regard to force and enforce this monopoly? If the private exacter of justice violates no one's rights, then punishing him for his actions (actions state officials also perform) violates his rights and hence violates moral side constraints. Monopolizing the use of force then, on this view, is itself immoral, as is redistribution through the compulsory tax apparatus of the state. Peaceful individuals minding their own business are not violating the rights of others. It does not constitute a violation of someone's rights to refrain from purchasing something for him (that you have not entered specifically into an obligation to buy). Hence, so the argument continues, when the state threatens someone with punishment if he does not contribute to the protection of another, it violates (and its officials violate) his rights. In threatening him with something that would be a violation of his rights if done by a private citizen, they violate moral constraints.

To get to something recognizable as a state we must show (1) how an ultraminimal state arises out of the system of private protective associations; and (2) how the ultraminimal state is transformed into the minimal state, how it gives rise to that "redistribution" for the general provision of protective services that constitutes it as the minimal state. To show that the minimal state is morally legitimate, to show it is not immoral itself, we must show also that these transitions in (1) and (2) *each* are morally legitimate. In the rest of Part I of this work we show how each of these transitions occurs and is morally permissible. We argue that the first transition, from a system of private protective agencies to an ultraminimal state, will occur by an invisible-hand process in a morally permissible way that violates no one's rights. Secondly, we argue that the transition from an ultraminimal state to a minimal state morally must occur. It would be morally impermissible for persons to maintain the monopoly in the ultraminimal state without providing protective services for all, even if this requires specific "redistribution." The operators of the ultraminimal state are morally obligated to produce the minimal state. The remainder of

Part I, then, attempts to justify the minimal state. In Part II, we argue that no state *more* powerful or extensive than the minimal state is legitimate or justifiable; hence that Part I justifies all that can be justified. In Part III, we argue that the conclusion of Part II is not an unhappy one; that in addition to being uniquely right, the minimal state is not uninspiring.