ANARCHY, STATE, AND UTOPIA

ROBERT NOZICK

BLACKWELL
Oxford UK & Cambridge USA

Reprinted 1980, 1984, 1986, 1988, 1990, 1991, 1992, 1993, 1995, 1996, 1997, 1998, 1999

Blackwell Publishers Ltd
108 Cowley Road, Oxford, OX4 1JF, UK

ISBN 0–631–19780–X

Printed in Great Britain by T J International Ltd, Padstow, Cornwall

This book is printed on acid-free paper

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.[1] 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.[2] 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? [3] 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." [4]

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,[5] 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 [6] 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.[7] 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 sore 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.[8] (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-

termediate 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.[9]

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 contraint 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.[10] 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 then 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? [11] 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 so 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.