

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


The Significance of Choice

Citation

Scanlon, Thomas M. 1986. The significance of choice. The Tanner Lectures on Human Values 7: 149-216

Published version

http://www.tannerlectures.utah.edu/

Link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:3200667

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story

The Significance of Choice

T. M. SCANLON, JR.

THE TANNER LECTURES ON HUMAN VALUES

Delivered at
Brasenose College, Oxford University

May 16, 23, and 28, 1986

T. M. SCANLON is Professor of Philosophy at Harvard University. He was educated at Princeton, Brasenose College, Oxford, and Harvard, and taught philosophy at Princeton from 1966 until 1984. Professor Scanlon is the author of a number of articles in moral and political philosophy and was one of the founding editors of *Philosophy and Public Affairs.*

Lecture 1

1. INTRODUCTION

Choice has obvious and immediate moral significance. The fact that a certain action or outcome resulted from an agent's choice can make a crucial difference both to our moral appraisal of that agent and to our assessment of the rights and obligations of the agent and others after the action has been performed. My aim in these lectures is to investigate the nature and basis of this significance. The explanation which I will offer will be based upon a contractualist account of morality — that is, a theory according to which an act is right if it would be required or allowed by principles which no one, suitably motivated, could reasonably reject as a basis for informed, unforced general agreement.[1]

I believe that it is possible within this general theory of morality to explain the significance of various familiar moral notions such as rights, welfare, and responsibility in a way that preserves their apparent independence rather than reducing all of them to one master concept such as utility. The present lectures are an attempt to carry out this project for the notions of responsibility and choice.

This is a revised version of three lectures presented at Brasenose College, Oxford, on May 16, 23, and 28, 1986. I am grateful to the participants in the seminars following those lectures for their challenging and instructive comments. These lectures are the descendants of a paper, entitled "Freedom of the Will in Political Theory," which I delivered at a meeting of the Washington, D.C., Area Philosophy Club in November 1977. Since that time I have presented many intervening versions to various audiences. I am indebted to members of those audiences and to numerous other friends for comments, criticism, and helpful suggestions.

[1] I have set out my version of contractualism in "Contractualism and Utilitarianism," in Amartya Sen and Bernard Williams, eds., *Utilitarianism and Beyond* (Cambridge: Cambridge University Press, 1982), pp. 103–28. What follows can be seen as an attempt to fulfill, for the case of choice, the promissory remarks made at the end of section III of that paper.

2. The Problems of Free Will

Quite apart from this general theoretical project, however, there is another, more familiar reason for inquiring into the basis of the moral significance of choice. This is the desire to understand and respond to the challenge to that significance which has gone under the heading of the problem of free will. This problem has a number of forms. One form identifies free will with a person's freedom to act otherwise than he or she in fact did or will. The problem, on this view, is the threat to this freedom posed by deterministic conceptions of the universe. A second, related problem is whether determinism, if true, would deprive us of the kind of freedom, whatever it may be, which is presupposed by moral praise and blame. This version of the problem is closer to my present concern in that it has an explicitly moral dimension. In order to address it one needs to find out what the relevant kind of freedom is, and this question can be approached by asking what gives free choice and free action their special moral significance. Given an answer to this question, which is the one I am primarily concerned with, we can then ask how the lack of freedom would threaten this significance and what kinds of unfreedom would do so.

The challenge I have in mind, however, is not posed by determinism but by what I call the Causal Thesis. This is the thesis that the events which are human actions, thoughts, and decisions are linked to antecedent events by causal laws as deterministic as those governing other goings-on in the universe. According to this thesis, given antecedent conditions and the laws of nature, the occurrence of an act of a specific kind follows, either with certainty or with a certain degree of probability, the indeterminacy being due to chance factors of the sort involved in other natural processes. I am concerned with this thesis rather than with determinism because it seems to me that the space opened up by the falsity of determinism would be relevant to morality only if it

were filled by something other than the cumulative effects of indeterministic physical processes, If the actions we perform result from the fact that we have a certain physical constitution and have been subjected to certain outside influences, then an apparent threat to morality remains, even if the links between these causes and their effects are not deterministic.

The idea that there is such a threat is sometimes supported by thought experiments such as the following: Suppose you were to learn that someone's present state of mind, intentions, and actions were produced in him or her a few minutes ago by the action of outside forces, for example by electrical stimulation of the nervous system. You would not think it appropriate to blame that person for what he or she does under such conditions. But if the Causal Thesis is true then all of our actions are like this. The only differences are in the form of outside intervention and the span of time over which it occurs, but surely these are not essential to the freedom of the agent.

How might this challenge be answered? One strategy would be to argue that there are mistakes in the loose and naive idea of causality to which the challenge appeals or in the assumptions it makes about the relation between mental and physical events. There is obviously much to be said on both of these topics. I propose, however, to follow a different (but equally familiar) line. Leaving the concepts of cause and action more or less unanalyzed, I will argue that the apparent force of the challenge rests on mistaken ideas about the nature of moral blame and responsibility.[2]

[2]In his admirably clear and detailed defense of incompatibilism, Peter van Inwagen observes that if one accepts the premises of his argument for the incompatibility of determinism and free will (in the sense required for moral responsibility) then it is "puzzling" how people could have the kind of freedom required for moral responsibility even under indeterministic universal causation. (See *An Essay on Free Will* [Oxford: Oxford University Press, 1983], pp. 149–50.) On the other hand, he takes it to be not merely puzzling but inconceivable that free will should be impossible or that the premises of his arguments for incompatibilism should be false or that the rules of inference which these arguments employ should be invalid. This leads him, after some further argument, to reject determinism: "If incompatibilism is true, then either determinism or the free-will thesis is false.

It has sometimes been maintained that even if the Causal Thesis holds, this does not represent the kind of unfreedom that excuses agents from moral blame. That kind of unfreedom, it is sometimes said, is specified simply by the excusing conditions which we generally recognize: a person is acting unfreely in the relevant sense only if he or she is acting under posthypnotic suggestion, or under duress, is insane, or falls under some other generally recognized excusing condition. Since the Causal Thesis does not imply that people are always acting under one or another of these conditions, it does not imply that moral praise and blame are generally inapplicable.

I am inclined to think that there is something right about this reaffirmation of common sense. But in this simple form it has been rightly rejected as question begging. It begs the question because it does not take account of the claim that commonsense morality itself holds that people cannot be blamed for what they do when their behavior is the result of outside causes, a claim which is supported by our reactions to imaginary cases like the thought experiment mentioned above and by more general reflection on what a world of universal causality would be like.

In order to show that moral praise and blame are compatible with the Causal Thesis, it is necessary to rebut this claim. The most promising strategy for doing so is to look for a general account of the moral significance of choice, an account which, on

—

To deny the free-will thesis is to deny the existence of moral responsibility, which would be absurd. Moreover, there seems to be no good reason to accept determinism (which, it should be recalled, is *not* the same as the Principle of Universal Causation). Therefore, we should reject determinism" (p. 223).

My response is somewhat different. Determinism is a very general empirical thesis. Our convictions about moral responsibility seem to me an odd basis for drawing a conclusion one way or the other about such a claim. In addition, whatever one may decide about determinism, it remains puzzling how moral responsibility could be compatible with Universal Causation. I am thus led to wonder whether our initial assumptions about the kind of freedom required by moral responsibility might not be mistaken. Rather than starting with a reinterpretation of the principle of alternative possibilities (along the lines of the conditional analysis), my strategy is to ask first, Why does the fact of choice matter morally? and then, What kind of freedom is relevant to mattering in that way?

the one hand, explains why the significance of choice is undermined both by commonly recognized excusing conditions and by factors such as those imagined to be at work in the thought experiment described above and, on the other hand, explains why the moral significance of choice will not be undermined everywhere if the Causal Thesis is true. Such an account, if convincing, would provide a basis for arguing that our initial response to the Causal Thesis was mistaken. At the very least, it would shift the burden of argument to the incompatibilist, who would need to explain why the proffered account of the moral significance of choice was inadequate. Before beginning my search for an account of the significance of choice, however, I will take a moment to examine some other forms of the free-will problem.

The problem of free will is most often discussed as a problem about moral responsibility, but essentially the same problem arises in other forms as well. It arises in political philosophy, for example, as a problem about the significance of choice as a legitimating condition. We generally think that the fact that the affected parties chose or assented to an outcome is an important factor in making that outcome legitimate. But we also recognize that there are conditions under which acquiescence does not have this legitimating force. These include conditions like those listed above: hypnosis, brain stimulation, mental incapacity, brainwashing, and so on. To many, at least, it seems plausible to maintain that these conditions deprive choice of its moral significance because they are conditions under which the agent's action is the result of outside causes. But if the Causal Thesis holds, this is true of all actions, and it would follow that choice never has moral significance as a legitimating factor.

Let me turn to a different example, drawn from John Rawls's book, *A Theory of Justice*.[3] (I believe the example involves a misinterpretation of Rawls, albeit a fairly natural one, but I will

[3] *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), pp. 72-74, 104.

try to correct that later.) Replying to an argument for the justice of a purely laissez-faire economy, Rawls observes that in such a system economic rewards would be unacceptably dependent on factors such as innate talents and fortunate family circumstances, which are, as he puts it, "arbitrary from a moral point of view." In particular, he says that even such factors as willingness to exert oneself will depend to a large extent on family circumstances and upbringing. Therefore we cannot say, of those who might have improved their economic position if they had exerted themselves, that because their predicament is their own doing they have no legitimate complaint. Their lack of exertion has no legitimating force because it is the result of "arbitrary factors."

But this argument, if successful, would seem to prove too much. Consider a society satisfying Rawls's Difference Principle. This principle permits some inequalities, such as those resulting from incentives which improve productivity enough to make everyone better off. When such inequalities exist, they will be due to the fact that some people have responded to these incentives while others have not. If the Causal Thesis is correct, however, there will be some causal explanation of these differences in behavior. They will not be due to gross differences in economic status, since, by hypothesis, these do not exist. But they must be due to something, and it seems clear that the factors responsible, whatever they are, are likely to be as "morally arbitrary" in at least one sense of that phrase as the factors at work in the case of the laissez-faire society to which Rawls was objecting. To sustain Rawls's argument, then, we need a better explanation of how "morally arbitrary" background conditions can undermine the legitimating force of choice, an explanation which will not deprive all choice of moral force if the Causal Thesis is correct.

Let me mention a further, slightly different case. We think it important that a political system should, as we say, "leave people free to make up their own minds," especially about important political questions and questions of personal values. We regard

certain conditions as incompatible with this important freedom and therefore to be avoided. Brainwashing is one extreme example, but there are also more moderate, and more common, forms of manipulation, such as strict control of sources of information, bombardment with one-sided information, and the creation of an environment in which people are distracted from certain questions by fear or other competing stimuli. What is it that is bad about these conditions? If they count as conditions of unfreedom simply because they are conditions under which people's opinions are causal products of outside factors, then there is no such thing as "freedom of thought" if the Causal Thesis is correct. It would follow that defenders of "freedom of thought" who accept the Causal Thesis could rightly be accused of ideological blindness: what they advocate as "freedom" is really just determination by a different set of outside factors, factors which are less rational and no more benign than those to which they object. There may be good reasons to favor some determining factors over others, but the issue cannot be one of "freedom." Here again, then, the problem is to show that "determination by outside causes" is not a sufficient condition for unfreedom. To do this we need to come up with some other explanation of what is bad about the conditions which supporters of freedom of thought condemn.[4]

These are versions of what I will call the political problem of free will. As I have said, they have much the same structure as the more frequently discussed problem about moral praise and blame. In addition to these problems there is what might be called the personal problem of free will. If I were to learn that one of my past actions was the result of hypnosis or brain stimulation, I would feel alienated from this act: manipulated, trapped, reduced to the status of a puppet. But why, if the Causal Thesis is correct, should we not feel this way about all of our acts? Why should

[4]I have said more about this version of the problem in section IIB of "Freedom of Expression and Categories of Expression," *University of Pittsburgh Law Review* 40 (1979).

we not feel trapped all the time? This is like the other problems in that what we need in order to answer it is a better explanation of why it is proper to feel trapped and alienated from our own actions in cases like hypnosis, an explanation which goes beyond the mere fact of determination by outside factors. But while this problem is like the others in its form, it differs from them in not being specifically a problem about morality: the significance with which it deals is not *moral* significance. This makes it a particularly difficult problem, much of the difficulty being that of explaining what the desired but threatened form of significance is supposed to be. Since my concern is with moral theory I will not address this problem directly, though the discussion of the value of choice in lecture 2 will have some bearing on it.

I will be concerned in these lectures with the first two of these problems and with the relation between them: to what degree can the "better explanation" that each calls for be provided within the compass of a single, reasonably unified theory? My strategy is to put forward two theories which attempt to explain why the conditions which we commonly recognize as undermining the moral significance of choice in various contexts should have this effect. These theories, which I will refer to as the Quality of Will theory and the Value of Choice theory, are similar to the theories put forward in two famous articles, P. F. Strawson's "Freedom and Resentment,"[5] and H. L. A. Hart's "Legal Responsibility and Excuses."[6] My aim is to see whether versions of these two approaches — extended in some respects and modified in others to fit within the contractualist theory I espouse — can be put together into a single coherent account. We can then see how far this combined theory takes us toward providing a satisfactory account of the moral significance of choice across the range of cases I have listed above.

[5]In Strawson, ed., *Studies in the Philosophy of Thought and Action* (Oxford: Oxford University Press, 1968), pp. 71–96.

[6]Chapter 2 of Hart, *Punishment and Responsibility* (Oxford: Oxford University Press, 1968).

3. THE INFLUENCEABILITY THEORY

Before presenting the Quality of Will theory, it will be helpful to consider briefly an older view which serves as a useful benchmark. This view, which I will call the Influenceability theory, employs a familiar strategy for explaining conditions which excuse a person from moral blame.[7] This strategy is first to identify the purpose or rationale of moral praise and blame and then to show that this rationale fails when the standard excusing conditions are present. According to the Influenceability theory, the purpose of moral praise and blame is to influence people's behavior. There is thus no point in praising or blaming agents who are not (or were not) susceptible to being influenced by moral suasion, and it is this fact which is reflected in the commonly recognized excusing conditions.

The difficulties with this theory are, I think, well known.[8] I will not go into them here except to make two brief points. The first is that the theory appears to conflate the question of whether moral judgment is applicable and the question of whether it should be *expressed* (in particular, expressed to the agent). The second point is that difficulties arise for the theory when it is asked whether what matters is influenceability at or shortly before the time of action or influenceability at the (later) time when moral judgment is being expressed. The utilitarian rationale for praise and blame supports the latter interpretation, but it is the former which retains a tie with commonsense notions of responsibility.

[7]See J. J. C. Smart, "Freewill, Praise, and Blame," *Mind* 70 (1961) : 291-306; reprinted in G. Dworkin, ed., *Determinism, Free Will, and Moral Responsibility* (Englewood Cliffs, N.J.: Prentice-Hall, 1970; page references will be to this edition). The theory was stated earlier by Moritz Schlick in chapter 7 of *The Problems of Ethics,* trans. D. Rynin (New York: Prentice-Hall, 1939), reprinted as "When Is a Man Responsible?" in B. Berofsky, ed., *Free Will and Determinism* (New York: Harper and Row, 1966; page references will be to this edition).

[8]Some are set forth by Jonathan Bennett in section 6 of "Accountability," in Zak van Staaten, ed., *Philosophical Subjects* (Oxford: Oxford University Press, 1980).

The Influenceability theory might explain why a utilitarian system of behavior control would include something like what we now recognize as excusing conditions. What some proponents of the theory have had in mind is that commonsense notions of responsibility should be given up and replaced by such a utilitarian practice. Whatever the merits of this proposal, however, it is clear that the Influenceability theory does not provide a satisfactory account of the notions of moral praiseworthiness and blameworthiness as we now understand them. The usefulness of administering praise or blame depends on too many factors other than the nature of the act in question for there ever to be a good fit between the idea of influenceability and the idea of responsibility which we now employ.[9]

4. Quality of Will: Strawson's Account

The view which Strawson presents in "Freedom and Resentment" is clearly superior to the Influenceability theory. Like that theory, however, it focuses less on the cognitive content of moral judgments than on what people are doing in making them. The centerpiece of Strawson's analysis is the idea of a reactive attitude. It is the nature of these attitudes that they are reactions not simply to what happens to us or to others but rather to the attitudes toward ourselves or others which are revealed in an agent's actions. For example, when you tread on my blistered toes, I may feel excruciating pain and greatly regret that my toes were stepped on. In addition, however, I am likely to resent the malevolence or callousness or indifference to my pain which your action indicates. This resentment is what Strawson calls a "personal reactive attitude": it is my attitudinal reaction to the attitude toward me which is revealed in your action. Moral indignation, on the other

[9]Broadening the theory to take into account the possibility of influencing people other than the agent will produce a better fit in some cases, but at the price of introducing even more considerations which are intuitively irrelevant to the question of responsibility.

hand, is what he calls a "vicarious attitude": a reaction to the attitude toward others in general (e.g., lack of concern about their pain) which your action shows you to have. All of these are what Strawson calls "participant attitudes." They "belong to involvement or participation with others in inter-personal human relationships."[10] This is in contrast to "objective attitudes,'' which involve seeing a person "as an object of social policy; as an object for what in a wide range of senses might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained."[11]

It follows from this characterization that the discovery of new facts about an action or an agent can lead to the modification or withdrawal of a reactive attitude in at least three ways: (a) by showing that the action was not, after all, indicative of the agent's attitude toward ourselves or others; (b) by showing that the attitude indicated in the act was not one which makes a certain reactive attitude appropriate; (c) by leading us to see the agent as someone toward whom objective, rather than participant, attitudes are appropriate.

Commonly recognized excusing conditions work in these ways. The most extreme excusing conditions sever any connection between an action (or movement) and the attitudes of the agent. If your stepping on my toes was a mere bodily movement resulting from an epileptic seizure, then it shows nothing at all about your concern or lack of concern about my pain. It would therefore be inappropriate for me to resent your action or for someone else, taking a more impartial view, to feel moral disapproval of you on that account.

Other excusing conditions have the less extreme effect of modifying the quality of will which an action can be taken to indicate, thus modifying the reactive attitudes which are appropri-

[10]Strawson, "Freedom and Resentment," p. 79.

[11]Ibid.

ate. If I learn, for example, that you stepped on my foot by accident, then I can no longer resent your callousness or malevolence, but I may still, if conditions are right, resent your carelessness. If I learn that you (reasonably) believed that the toy spider on my boot was real, and that you were saving my life by killing it before it could bite me, then I can no longer *resent* your action at all, although it remains indicative of a particular quality of will on your part.

Actions produced by posthypnotic suggestion are a less clear case. Much depends on what we take the hypnosis to do. Hypnosis might lead you to perform the intentional act of stamping your foot on mine but without any malice or even any thought that you are causing me harm. In this case a criticizable attitude is indicated by your act: a kind of complacency toward touching other people's bodies in ways that you have reason to believe are unwanted. But this attitude is not really attributable to *you*. *You* may not lack any inhibition in this regard: it is just that your normal inhibition has been inhibited by the hypnotist. The case is similar if the hypnotist implants in you a passing hatred for me and a fleeting but intense desire to cause me pain. Here again there is a criticizable attitude — more serious this time — but it is not yours. It is "just visiting," so to speak.

Strawson's account of why conditions such as insanity and extreme immaturity excuse people from moral blame is less satisfactory. The central idea is that these conditions lead us to take an "objective attitude" toward a person rather than to see him or her as a participant in those interpersonal human relationships of which the reactive attitudes are a part. Strawson's claim here can be understood on two levels. On the one hand there is the empirical claim that when we see someone as "warped or deranged, neurotic or just a child . . . all our reactive attitudes tend to be profoundly modified."[12] In addition to this, however, there is the

[12]Ibid. My appreciation of this straightforwardly factual reading of Strawson's argument was aided by Jonathan Bennett's perceptive analysis in "Accountability."

suggestion that these factors render reactive attitudes such as resentment and indignation *inappropriate.* But Strawson's theory does not explain the grounds of this form of inappropriateness as clearly as it explained the grounds of the other excusing conditions. In fact, aside from the references to interpersonal relationships, which are left unspecified, nothing is said on this point.

In other cases, however, Strawson's theory succeeds in giving a better explanation of commonly recognized excusing conditions than that offered by the idea that a person is not to be blamed for an action which is the result of outside causes. The mere fact of causal determination seems to have little to do with the most common forms of excuse, such as accident and mistake of fact. It is a distinct advantage of Strawson's analysis that it accounts for the force of more extreme excuses such as hypnosis and brain stimulation in a way that is continuous with a natural explanation of these less extreme cases as well. Moreover, his theory can explain the relevance of "inability to do otherwise" in several senses of that phrase. Sometimes, as in the case of brain stimulation, the factors which underlie this inability sever any connection between an action and the agent's attitudes. In other cases, "inability to do otherwise" in the different sense of lack of *eligible* alternatives can modify the quality of will indicated by an agent's willingness to choose a particular course of action. For example, if you stamp on my toes because my archenemy, who is holding your child hostage next door, has ordered you to do so, this does not make you less *responsible* for your act. The act is still fully yours, but the quality of will which it indicates on your part is not blameworthy.

As Strawson observes, these appeals to "inability to do otherwise'' do not generalize. The truth of the Causal Thesis would not mean that either of these forms of inability obtained generally or that actions never indicated the presence in the agent of those attitudes or qualities of will which make resentment or moral indignation appropriate.

Like the unsuccessful defense of common sense mentioned above, Strawson's analysis is internal to our moral concepts as we now understand them. Its explanation of the conditions which negate or modify moral responsibility rests on a claim that, given the kind of thing that moral indignation is, it is an appropriate response only to actions which manifest certain attitudes on the part of the agent. This internal character may be thought to be a weakness in Strawson's account, and he himself considers an objection of this sort. The objection might be put as follows: You have shown what is and is not appropriate given the moral notions we now have; but the question is whether, if the Causal Thesis is correct, it would not be irrational to go on using those concepts and holding the attitudes they describe. Strawson's direct response to this objection is to say that the change proposed is "practically inconceivable."

> The human commitment to participation in ordinary interpersonal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them; and being involved in interpersonal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question.[13]

But there is another reply which is suggested by something that Strawson goes on to say and which seems to me much stronger.[14] This reply points out that the principle "If your action was a causal consequence of prior factors outside your control then you cannot properly be praised or blamed for performing it" derives its strength from its claim to be supported by commonsense morality. Consequently, if an analysis such as Strawson's succeeds

[13] Strawson, "Freedom and Resentment," p. 82.

[14] Ibid., p. 83,

in giving a convincing account of the requirements of freedom implicit in our ordinary moral views — in particular, giving a systematic explanation of why commonly recognized excusing conditions should excuse — then this is success enough. Succeeding this far undermines the incompatibilist challenge by striking at its supposed basis in everyday moral thought.[15]

Plausible and appealing though it is, there are several respects in which Strawson's analysis is not fully satisfactory. One of these has already been mentioned in connection with insanity. Strawson suggests that the attitudes which moral judgments express are appropriately held only toward people who are participants in certain interpersonal relationships and that these attitudes are therefore inhibited when we become aware of conditions which render a person unfit for these relationships. But one needs to know more about what these relationships are, about why moral reactive attitudes depend on them, and about how these relationships are undermined or ruled out by factors such as insanity.

A second problem is more general. Strawson explains why certain kinds of unfreedom make moral praise and blame inapplicable by appealing to a fact about interpersonal reactive attitudes in general (and moral ones in particular), namely the fact that they are attitudes toward the attitudes of others, as manifested in their actions. But one may wonder whether anything further can be said about why attitudes of moral approval and disapproval are of this general type. Moreover, it is not clear that moral judgments need always involve the *expression* of any par-

[15]Compare Thomas Nagel's comments on Strawson's theory in *The View from Nowhere* (Oxford: Oxford University Press, 1986), pp, 124-26. The response I am advocating here does not deny the possibility of what Nagel has called "external" criticism of our practices of moral evaluation. It tries only to deny the incompatibilist critique a foothold in our ordinary ideas of moral responsibility. It claims that a commitment to freedom which is incompatible with the Causal Thesis is not embedded in our ordinary moral practices in the way in which a commitment to objectivity which outruns our experience is embedded in the content of our ordinary empirical beliefs. The incompatibilist response, obviously, is to deny this claim. My point is that the ensuing argument, which I am trying to advance one side of, is internal to the system of our ordinary moral beliefs.

ticular reactive attitude. For example, I may believe that an action of a friend, to whom many horrible things have recently happened, is morally blameworthy. But need this belief, or its expression, involve a feeling or expression of moral indignation or disapproval on my part? Might I not agree that what he did was wrong but be incapable of feeling disapproval toward him?

Here Strawson's analysis faces a version of one of the objections to the Influenceability theory: it links the content of a moral judgment too closely to *one* of the things that may be done in expressing that judgment. Of course, Strawson need not claim that moral judgment always involves the expression of a reactive attitude. It would be enough to say that such a judgment always makes some attitude (e.g., disapproval) appropriate. But then one wonders what the content of this underlying judgment is and whether the requirement of freedom is not to be explained by appeal to this content rather than to the attitudes which it makes appropriate.

In order to answer these questions one needs a more complete account of moral blameworthiness. A number of different moral theories might be called upon for this purpose, but what I will do is to sketch briefly how a Quality of Will theory might be based on a contractualist account of moral judgment.

5. Quality of Will: A Contractualist Analysis

According to contractualism as I understand it, the basic moral motivation is a desire to regulate one's behavior according to standards that others could not reasonably reject insofar as they, too, were looking for a common set of practical principles. Morality, on this view, is what might be called a system of co-deliberation. Moral reasoning is an attempt to work out principles which each of us could be expected to employ as a basis for deliberation and to accept as a basis for criticism. To believe that one is morally at fault is just to believe that one has not regulated one's behavior in the way that such standards would

require. This can be so either because one has failed to attend to considerations that such standards would require one to take account of or because one has consciously acted contrary to what such standards would require. If one is concerned, as most people are to at least some extent, to be able to justify one's actions to others on grounds they could not reasonably reject, then the realization that one has failed in these ways will normally produce an attitude of serious self-reproach. But this attitude is distinct from the belief which may give rise to it. Similarly, to believe that another person's behavior is morally faulty is, at base, to believe that there is a divergence of this kind between the way that person regulated his or her behavior and the kind of self-regulation that mutually acceptable standards would require. For reasons like those just mentioned, this belief will normally be the basis for attitudes of disapproval and indignation. This view of morality grounds the fact that moral appraisal is essentially concerned with "the quality of an agent's will" in an account of the nature of moral reasoning and moral motivation. The analysis of moral judgment which it supports is essentially cognitivist. It can explain why moral judgments would normally be accompanied by certain attitudes, but these attitudes are not the basis of its account of moral judgment.

Contractualism also gives specific content to the idea, suggested by Strawson, that moral judgments presuppose a form of interpersonal relationship. On this view, moral judgments apply to people considered as possible participants in a system of co-deliberation. Moral praise and blame can thus be rendered inapplicable by abnormalities which make this kind of participation impossible. (The implications of this idea for excusing conditions such as insanity will be discussed below.)

6. THE SPECIAL FORCE OF MORAL JUDGEMENT

Insofar as it goes beyond Strawson's theory in committing itself to a fuller account of the nature of moral blameworthiness, the

contractualist view I have described leaves itself open to the objection that this notion of blameworthiness requires a stronger form of freedom, a form which may be incompatible with the Causal Thesis. In order to assess this objection, it will be helpful to compare the contractualist account of blame with what Smart calls "praise and dispraise." According to Smart, we commonly use the word "praise" in two different ways.[16] On the one hand, praise is the opposite of blame. These terms apply only to what a person does or to aspects of a person's character, and they are supposed to carry a special force of moral approval or condemnation. But we also praise things other than persons and their character: the California climate, the flavor of a melon, or the view from a certain hill. In this sense we also praise features of persons which we see as "gifts" beyond their control: their looks, their coordination, or their mathematical ability. Praise in this sense is not the opposite of blame, and Smart coins the term "dispraise" to denote its negative correlate. Praise and dispraise lack the special force of moral approval or condemnation which praise and blame are supposed to have. To praise or dispraise something is simply to grade it.

Smart takes the view that the kind of moral judgment involved in praise and blame as these terms are normally used must be rejected because it presupposes an unacceptable metaphysics of free will. However, we can praise and dispraise actions and character just as we can grade eyes and skill and mountain peaks. The primary function of praise in this "grading" sense, according to Smart, is just "to tell people what people are like."[17] However, since people like being praised and dislike being dispraised, praise and dispraise also have the important secondary function of serving to encourage or discourage classes of actions. Smart suggests that "clear-headed people," insofar as they use the terminology of praise and blame, will use it only in this "grading" sense and will

[16]Smart, "Freewill, Praise, and Blame," p. 210.

[17]Ibid., p. 211.

restrict its use to cases in which this important secondary function can be fulfilled.

Most people would agree that moral praise and blame of the kind involved when we "hold a person responsible" have a force which goes beyond the merely informational function of "telling people what people are like." The problem for a compatibilist is to show that judgments with this "additional force" can be appropriate even if the Causal Thesis is true. The prior problem for moral theory is to say what this "additional force" is. What is it that an account of moral judgment must capture in order to be successfully "compatibilist"?

As I have said, Smart's analysis is not compatibilist. His aim is to replace ordinary moral judgment, not to analyze it. Strawson, on the other hand, is offering a compatibilist analysis of (at least some kinds of) moral judgment, and his analysis clearly satisfies one-half of the compatibilist test. The expression of interpersonal reactive attitudes is compatible with the Causal Thesis for much the same reason that Smart's notions of praise and dispraise are. These attitudes are reactions to "what people are like," as this is shown in their actions. As long as the people in question really are like this — as long, that is, as their actions really do manifest the attitudes in question —these reactive attitudes are appropriate.

Strawson's theory is more appealing than Smart's because it offers a plausible account of moral judgment as we currently understand it, an account of how moral judgment goes beyond merely "saying what people are like" and of how it differs from mere attempts to influence behavior. But his theory is like Smart's in locating the "special force" of moral judgment in what the moral judge is *doing.* The contractualist account I am offering, on the other hand, locates the origin of this distinctive force in what is claimed about the person judged. It is quite compatible with this analysis that moral judgments should often be intended to influence behavior and that they should often be made as expressions of reactive attitudes; but such reforming or expressive

intent is not essential. What is essential, on this account, is that a judgment of moral blame asserts that the way in which an agent decided what to do was not in accord with standards which that agent either accepts or should accept insofar as he or she is concerned to justify his or her actions to others on grounds that they could not reasonably reject. This is description, but given that most people care about the justifiability of their actions to others, it is not *mere* description.

This account of the special force of moral judgment may still seem inadequate. Given what I have said it may seem that, on the contractualist view, this special force lies simply in the fact that moral judgments attribute to an agent properties which most people are seriously concerned to have or to avoid. In this respect moral judgments are like judgments of beauty or intelligence. But these forms of appraisal, and the pride and shame that can go with accepting them, involve no attribution of responsibility and hence raise no question of freedom. To the extent that moral appraisal is different in this respect, and does raise a special question of freedom, it would seem that this difference is yet to be accounted for.

One way in which freedom is relevant to moral appraisal on the Quality of Will theory (the main way mentioned so far) is this: insofar as we are talking about praising or blaming a person on the basis of a particular action, the freedom or unfreedom of that action is relevant to the question whether the intentions and attitudes seemingly implicit in it are actually present in the agent. This evidential relevance of freedom is not peculiar to moral appraisal, however. Similar questions can arise in regard to assessments of intelligence or skill on the basis of particular pieces of behavior. (We may ask, for example, whether the occasion was a fair test of her skill, or whether there were interfering conditions.) The objection just raised does not dispute the ability of the Quality of Will theory to explain *this* way in which moral judgments may depend on questions of freedom, but it suggests that this is not enough. It assumes that "blameworthy" intentions

and attitudes are correctly attributed to an agent and then asks how, on the analysis I have offered, this attribution goes beyond welcome or unwelcome description. Behind the objection lies the idea that going "beyond description" in the relevant sense would involve holding the agent *responsible* in a way that people are not (normally) responsible for being beautiful or intelligent and that this notion of responsibility brings with it a further condition of freedom which my discussion of the Quality of Will theory has so far ignored.

I do not believe that in order to criticize a person for behaving in a vicious and callous manner we must maintain that he or she is responsible for becoming vicious and callous. Whether a person is so responsible is, in my view, a separate question. Leaving this question aside, however, there is a sense in which we are responsible for — or, I would prefer to say, *accountable for* — our intentions and decisions but not for our looks or intelligence. This is just because, insofar as these intentions and decisions are *ours,* it is appropriate to ask us to justify or explain them —appropriate, that is, for someone to ask, Why do you think you can treat me this way? in a way that it would not be appropriate to ask, in an accusing tone, Why are you so tall? This is not to say that these mental states are the kinds of thing which have reasons *rather than causes* but only that they are states for which requests for reasons are in principle relevant.

Moral criticism and moral argument, on the contractualist view, consist in the exchange of such requests and justifications. Adverse moral judgment therefore differs from mere unwelcome description because it calls for particular kinds of response, such as justification, explanation, or admission of fault. In what way does it "call for" these responses? Here let me make three points. First, the person making an adverse moral judgment is often literally asking for or demanding an explanation, justification, or apology. Second, moral criticism concerns features of the agent for which questions about reasons, raised by the agent him or her-

self, are appropriate. Insofar as I think of a past intention, decision, or action as *mine,* I think of it as something which was sensitive to my assessment, at the time, of relevant reasons. This makes it appropriate for me to ask myself, Why did I think or do that? and Do I still take those reasons to be sufficient? Third, the contractualist account of moral motivation ties these two points together. A person who is concerned to be able to justify him- or herself to others will be moved to respond to the kind of demand I have mentioned, will want to be able to respond positively (i.e., with a justification) and will want to carry out the kind of first-person reflection just described in a way that makes such a response possible. For such a person, moral blame differs from mere unwelcome description not only because of its seriousness but also because it engages in this way with an agent's own process of critical reflection, thus raising the questions Why did I do that? Do I still endorse those reasons? Can I defend the judgment that they were adequate grounds for acting?

Whether one accepts this as an adequate account of the "special force" of moral judgments will depend, of course, on what one thinks that moral judgment in the "ordinary" sense actually entails. Some have held that from the fact that a person is morally blameworthy it follows that it would be a good thing if he or she were to suffer some harm (or, at least, that this would be less bad than if some innocent person were to suffer the same harm).[18] I do not myself regard moral blame as having this implication. So if a compatibilist account of moral judgment must have this consequence, I am content to be offering a revisionist theory. (The problem of how the fact of choice may make harmful consequences more justifiable will, however, come up again in lecture 2.)

[18]This idea was suggested to me by Derek Parfit in the seminar following the presentation of this lecture in Oxford.

7. BLAMEWORTHINESS AND FREEDOM

It remains to say something about how this contractualist version of the Quality of Will theory handles the difficult question of moral appraisal of the insane. Discussion of this matter will also enable me to draw together some of the points that have just been made and to say more about the kind of freedom which is presupposed by moral blameworthiness according to the theory I have been proposing.

As I said earlier, to believe that one's behavior is morally faulty is to believe either that one has failed to attend to considerations which any standards that others could not reasonably reject would require one to attend to or that one has knowingly acted contrary to what such standards would require. Let me focus for a moment on the first disjunct. Something like this is a necessary part of an account of moral blameworthiness, since failure to give any thought at all to what is morally required can certainly be grounds for moral criticism. But the purely negative statement I have given above is too broad. The class of people who simply fail to attend to the relevant considerations includes many who do not seem to be candidates for moral blame: people acting in their sleep, victims of hypnosis, young children, people suffering from mental illness, and so on. We need to find, within the notion of moral blame itself, some basis for a nonarbitrary qualification of the purely negative criterion.

According to contractualism, thought about right and wrong is a search for principles "for the regulation of behavior" which others, similarly motivated, have reason to accept. What kind of "regulation" is intended here? Not regulation "from without" through a system of social sanctions but regulation "from within" through critical reflection on one's own conduct under the pressure provided by the desire to be able to justify one actions to others on grounds they could not reasonably reject. This idea of regulation has two components, one specifically moral, the other not. The

specifically moral component is the ability to reason about what could be justified to others. The nonmoral component is the more general capacity through which the results of such reasoning make a difference to what one does. Let me call this the capacity for critically reflective, rational self-governance — "critically reflective'' because it involves the ability to reflect and pass judgment upon one's actions and the thought processes leading up to them; "rational" in the broad sense of involving sensitivity to reasons and the ability to weigh them; "self-governance" because it is a process which makes a difference to how one acts.

The critical reflection of a person who has this capacity will have a kind of coherence over time. Conclusions reached at one time will be seen as relevant to critical reflection at later times unless specifically overruled. In addition, the results of this reflection will normally make a difference both in how the person acts given a certain perception of a situation and in the features of situations which he or she is on the alert for and tends to notice.

This general capacity for critically reflective, rational self-governance is not specifically moral, and someone could have it who was entirely unconcerned with morality. Morality does not tell one to have this capacity, and failing to have it in general or on a particular occasion is not a moral fault. Rather, morality is addressed to people who are assumed to have this general capacity, and it tells them how the capacity should be exercised. The most general moral demand is that we exercise our capacity for self-governance in ways that others could reasonably be expected to authorize. More specific moral requirements follow from this.

Since moral blameworthiness concerns the exercise of the general capacity of self-governance, our views about the limits of moral blame are sensitive to changes in our views about the limits of this capacity. We normally believe, for example, that very young children lack this capacity and that it does not govern our actions while we are asleep. Nor, according to some assumptions about hypnosis, does it regulate posthypnotic suggestion, and it is

generally believed to be blocked by some forms of mental illness. These assumptions could be wrong, but given that we hold them it is natural that we do not take people in these categories to be morally blameworthy for their actions. (Whether we think it is useful to blame them is of course another question.) It is important to our reactions in such cases, however, that what is impaired or suspended is a *general* capacity for critically reflective, rational self-governance. If what is "lost" is more specifically moral — if, for example, a person lacks any concern for the welfare of others — then the result begins to look more like a species of moral fault.

As a "higher order" capacity, the capacity for critically reflective, rational self-governance has an obvious similarity to the capacities for higher-order desires and judgments which figure in the analyses of personhood and freedom offered by Harry Frankfurt and others.[19] I have been led to this capacity, however, not through an analysis of general notions of freedom and personhood but rather through reflection on the nature of moral argument and moral judgment. Basic to morality as I understand it is an idea of agreement between individuals *qua* critics and regulators of their own actions and deliberative processes. Critically reflective, rational self-governance is a capacity which is required in order for that idea not to be an idle one. It follows that moral criticism is restricted to individuals who have this capacity and to actions which fall within its scope.[20]

[19]See Harry Frankfurt, "Freedom of the Will and the Concept of a Person," *Journal of Philosophy* 68 (1971): 5–20; Wright Neely, "Freedom and Desire," *Philosophical Review* 83 (1974): 32-54; and Gary Watson, "Free Agency," *Journal of Philosophy* 72 (1975): 205–20.

[20]The idea that moral criticism is applicable only to actions which are within the scope of a capacity of self-governance which normally makes a difference in what a person does marks a point of tangency between the Influenceability theory and the analysis I am offering. I am not suggesting, however, that particular acts of moral criticism are aimed at influencing people or that moral criticism is always inappropriate when there is no hope of its making any difference to what people do. Morality as I am describing it is in a general sense "action guiding" — moral argument concerns principles for the general regulation of behavior. But moral

In Frankfurt's terms, these restrictions correspond roughly to a restriction to persons (as opposed to "wantons") and a restriction to actions which are performed freely. In my view, however, this last characterization is not entirely apt. Aside from external impediments to bodily motion, what is required for moral appraisal on the view I am presenting is the "freedom," whatever it may be, which is required by critically reflective, rational self-governance. But this is less appropriately thought of as a kind of freedom than as a kind of intrapersonal responsiveness. What is required is that what we do be importantly dependent on our process of critical reflection, that that process itself be sensitive to reasons, and that later stages of the process be importantly dependent on conclusions reached at earlier stages. But there is no reason, as far as I can see, to require that this process itself not be a causal product of antecedent events and conditions.'' Calling the relevant condition a form of freedom suggests this requirement, but this suggestion is undermined by our investigation into the moral significance of choice.

8. Conclusion

The contractualist version of the Quality of Will theory which I have described seems to me to provide a satisfactory explanation of the significance of choice for the moral appraisal of agents.

—

"ought" judgments need not be intended as action guiding, and insofar as they do guide action they need not do so by being prescriptive in form. Rather, they guide action by calling attention to facts about the justifiability of actions — facts which morally concerned agents care about. In these respects my view differs from R. M. Hare's prescriptivism, though we would say some of the same things about free will. See his "Prediction and Moral Appraisal," in P. French, T. Uehling, and H. Wettstein, eds., *Midwest Studies in Philosophy*, vol. III (Minneapolis: University of Minnesota Press, 1978), pp. 17–27.

[21]For more extended discussion of this issue, see Daniel Dennett's *Elbow Room* (Cambridge, Mass.: MIT Press, 1984), especially chs. 3–5. I make no claim to be advancing beyond what other compatibilists have said about the nature of deliberation and action. My concern is with the question of moral responsibility. Here I differ with Dennett, who goes much further than I would toward accepting the Influenceability theory. See ch. 7 of *Elbow Room* and Gary Watson's criticisms of it in his review in *Journal of Philosophy* 83 (1986): 517–22.

This theory offers a convincing and unified account of familiar excusing conditions, such as mistake of fact and duress, and explains our reactions to questions about moral appraisal of very young children, the insane, and victims of hypnosis. It can explain the special critical force which moral judgments seem to have, and it does this without presupposing a form of freedom incompatible with the Causal Thesis. But the theory applies only to what I called earlier the moral version of the free-will problem. A parallel account may, as I will suggest later, have some relevance to the case of criminal punishment, but it does not offer a promising approach to the other problems I have mentioned. The significance of a person's choices and other subjective responses for questions of economic justice and freedom of thought may have something to do with the fact that these responses reflect what might loosely be called "the quality of the person's will," but this is not because what we are doing in these cases is judging this "quality" or expressing attitudes toward it (since this is not what we are doing.) So, in search of an explanation that might cover these other cases, I will look in a different direction.

Lecture 2

1. THE VALUE OF CHOICE

It would have been natural to call these lectures an investigation into the significance of voluntariness. I have spoken of "choice" instead because this term applies not only to something that an agent does — as in "She made a choice" — but also to what an agent is presented with — as in "She was faced with this choice." It thus encompasses both an action and a situation within which such an action determines what will happen: a set of alternatives, their relative desirabilities, the information available to the agent, and so on. My main concern in these lectures is with the significance of choice in the first of these senses: the moral

significance of the choices people make. In this lecture, however, I will present a theory which exploits the ambiguity just mentioned by seeking to explain one kind of moral significance of the choices people make in terms of the value of the choices they have. I will call this the Value of Choice theory.[22]

This theory starts from the idea that it is often a good thing for a person to have what will happen depend upon how he or she responds when presented with the alternatives under the right conditions. To take a banal example, when I go to a restaurant, it is generally a good thing from my point of view to have what appears on my plate depend on the way in which I respond when presented with the menu. The most obvious reason why choice has value for me in this situation is simply instrumental: I would like what appears on my plate to conform to my preferences at the time it appears, and I believe that if what appears then is made to depend on my response when faced with the menu then the result is likely to coincide with what I want. This reason for valuing choice is both conditional and relative. It is conditional in that the value of my response as a predictor of future satisfaction depends on the nature of the question and the conditions under which my response is elicited. It is relative in that it depends on the reliability of the available alternative means for selecting the outcomes in question. In the restaurant case this value depends on how much I know about the cuisine in question and on my condition at the time the menu arrives: on whether I am drunk or overeager to impress my companions with my knowledge of French

[22]As I have said, the basic idea of this theory was presented by Hart in "Legal Responsibility and Excuses." Since Hart's article others have written in a similar vein, although they have been concerned mainly with the theory of punishment. See, for example, John Mackie, "The Grounds of Responsibility," in P. M. S. Hacker and J. Raz, eds., *Law, Morality, and Society: Essays in Honour of H. L. A. Hart*, (Oxford: Oxford University Press, 1977), and C. S. Nino, "A Consensual Theory of Punishment," *Philosophy and Public Affairs* 12 (1983): 289–306. Like Hart, Nino links the significance of choice (in his terms, consent) as a condition of just punishment with its significance elsewhere in the law, e.g., in contracts and torts. His view of this significance, however, is closer than my own to what I refer to below as the Forfeiture View.

or my ability to swallow highly seasoned food. Thus the same interest which sometimes makes choice valuable — the desire that outcomes should coincide with one's preferences — can at other times provide reasons for wanting outcomes to be determined in some other way, When I go to an exotic restaurant with my sophisticated friends, the chances of getting a meal that accords with my preferences may be increased if someone else does the ordering.

What I have described so far is what might be called the "predictive" or "instrumental" value of choice. In the example I have given, choice is instrumental to my own future enjoyment, but the class of states which one might seek to advance by making outcomes dependent on choices is of course much broader. Aside from such instrumental values, however, there are other ways in which having outcomes depend on my choice can have positive or negative value for me. One of these, which I will call "demonstrative" value, can be illustrated as follows. On our anniversary, I want not only to have a present for my wife but also to have chosen that present myself. This is not because I think this process is the one best calculated to produce a present she will like (for that, it would be better to let her choose the present herself). The reason, rather, is that the gift will have special meaning if I choose it — if it reflects my feelings about her and my thoughts about the occasion. On other occasions, for reasons similar in character but opposite in sign, I might prefer that outcomes *not* be dependent on my choices. For example, I might prefer to have the question of who will get a certain job (my friend or a stranger) not depend on how I respond when presented with the choice: I want it to be clear that the outcome need not reflect my judgment of their respective merits or my balancing of the competing claims of merit and loyalty.

The features of oneself which one may desire to demonstrate or see realized in action are highly varied. They may include the value one attaches to various aims and outcomes, one's knowledge,

awareness, or memory, or one's imagination and skill. Many of these are involved in the example cited: I want to make the choice myself because the result will then indicate the importance I attach to the occasion (my willingness to devote time to choosing a gift); my memory of, attention to, and concern for what she likes; as well as my imagination and skill in coming up with an unusual and amusing gift. The desire to see such features of oneself manifested in actions and outcomes is of course not limited to cases in which one's feelings for another person are at issue. I want to choose the furniture for my own apartment, pick out the pictures for the walls, and even write my own lectures despite the fact that these things might be done better by a decorator, art expert, or talented graduate student. For better or worse, I want these things to be produced by and reflect my own taste, imagination, and powers of discrimination and analysis. I feel the same way, even more strongly, about important decisions affecting my life in larger terms: what career to follow, where to work, how to live.

These last examples, however, may involve not only demonstrative but also what I will call "symbolic" value. In a situation in which people are normally expected to determine outcomes of a certain sort through their own choices unless they are not competent to do so, I may value having a choice because my not having it would reflect a judgment on my own or someone else's part that I fell below the expected standard of competence. Thus, while I might like to have the advantage of my sophisticated friends' expertise when the menu arrives tonight, I might prefer, all things considered, to order for myself, in order to avoid public acknowledgment of my relative ignorance of food, wine, and foreign cultures.

I make no claim that these three categories of value are mutually exclusive or that, taken together, they exhaust the forms of value that choice can have. My aim in distinguishing them is simply to illustrate the value that choice can have and to make clear that this value is not always merely instrumental: the reasons

people have for wanting outcomes to be (or sometimes not to be) dependent on their choices has to do with the significance that choice itself has for them, not merely with its efficacy in promoting outcomes which are desired on other grounds.

The three forms of value which I have distinguished (predictive, demonstrative, and symbolic) would all figure in a full account of the problem of paternalism. Legal restriction of people's freedom "for their own good" is likely to seem justified where (a) people who make a certain choice are likely to suffer very serious loss; (b) the instrumental value of choice as a way of warding off this loss is, given the circumstances under which that choice would be exercised, seriously undermined; (c) the demonstrative value that would be lost by being deprived of this choice is minimal; and (d) the tendency to "make the wrong choice" under the circumstances in question is widely shared, so that no particular group is being held inferior in the argument for legal regulation. The pejorative ring of "paternalism" and the particular bitterness attaching to it stem from cases in which either the seriousness of the loss in question or the foolishness of the choice leading to it is a matter of controversy. Those who are inclined to make a particular choice may not see it as mistaken and may attach demonstrative value to it. Consequently, they may resent paternalistic legislation, which brands them as less than fully competent when, in their view, they merely differ from the majority in the things they value. But this kind of resentment need not properly extend to other kinds of legislation sometimes called "paternalistic," such as wage and hour laws. Whether there is any reason at all for such resentment will depend on the reasons supporting a piece of legislation and also on the reasons people actually have for valuing freedom of choice which they would lose.

As controversies about paternalism illustrate, people can disagree sharply about the value of particular choices. They disagree, for example, about how important it is to have whether one wears

a seat belt depend on how one reacts (in the absence of any coercion) when setting off in a car. Some regard it as a significant loss when some form of coercion or even mild duress (the threat of a fine, or even the monitory presence of a brief buzzer) is introduced. Others, like me, regard this loss as trivial, and see the "constrained" choice as significantly more valuable than the unconstrained one. This disagreement reflects differences in the instrumental, demonstrative, and symbolic value we attach to these choices.

The existence of such differences raises the question of what is to count as "the value" of a choice as I have been using this phrase. One possibility is what I will call "fully individualized value." This is the value of the choice to a particular individual, taking into account the importance that individual attaches to having particular alternatives available, the difference that it makes to that individual which of these alternatives actually occurs, the importance which the individual attaches to having this be determined by his or her reactions, and the skill and discernment with which that individual will choose under the conditions in question. This fully individualized value may not be the same as the value which the individual actually assigns to the choice in question; rather, it is the *ex ante* value which he or she *should* assign given his or her values and propensities.

Fully individualized value is not what normally figures in moral argument, however. Appeals to the value of choice arise in moral argument chiefly when we are appraising moral principles or social institutions rather than when we are discussing particular choices by specific individuals. In these contexts we have to answer such general questions as How important is it to have the selection among these alternatives depend on one's choice? How bad a thing is it to have to choose under these conditions? When we address these questions, fully individualized values are not known. We argue instead in terms of what might be called the "normalized value" of a choice: a rough assignment of values to

categories of choice which we take to be a fair starting point for justification. Thus, for example, we take it as given for purposes of moral argument that it is very important that what one wears and whom one lives with be dependent on one's choices and much less important that one be able to choose what other people wear, what they eat, and how they live. And we do this despite the fact that there may be some who would not agree with this assignment of values.

This phenomenon — the use in moral argument of nonunanimously held "normalized" standards of value — is familiar and by no means limited to the case of choice. The status and justification of such standards is a difficult problem in moral theory. I will not address the general question here but will mention briefly two points about the case of choice. First, "giving people the choice" — for example, the opportunity to transfer goods through market trading — is one way to deal with the problem of divergent individual preferences. What has just been indicated, however, is that it is at best a partial solution. "Having a choice" among specified alternatives under specified conditions is itself a good which individuals may value differently — as is "having the choice whether to have the choice" and so on.[23] Second, differences in individualized valuations of choices result not only from differences in preference but also from differences in the personal characteristics which make a choice valuable: differences in foresight, in self-control, in self-understanding, and so on. Moral argument commonly refers to "normal" levels of these capacities as well as to "normal" valuations of outcomes and of demonstrative and symbolic values.

Let me turn now to the question of how the value of choice is related to the Quality of Will theory, discussed above. Like

[23]The variability of the value of choice is pointed out clearly by Gerald Dworkin in "Is More Choice Better Than Less?" in P. French, T. Uehling, and H. Wettstein, eds., *Midwest Studies in Philosophy*, vol. VII (Minneapolis: University of Minnesota Press, 1982), pp. 47-62.

what I have here called predictive and demonstrative value, the form of appraisal underlying the Quality of Will theory starts from the obvious fact that subjective responses can indicate or express continuing features of a person and from the equally obvious fact that these responses are better indicators under some conditions than under others. Even in this common starting point, however, there is a difference: the features of the person with which the Quality of Will theory is concerned constitute a narrow subset of those that give choice its value for the agent. For example, I want to choose my own food largely because my choices will be good indicators of what will please me, but my being pleased more by fish than by liver is not part of the quality of my will with which moral judgment is concerned.

Where the two theories differ most importantly, however, is in the way in which they assign moral significance to this indicative aspect of choice. The Quality of Will theory takes the point of view of the moral judge. Variations in the indicative value of subjective responses are significant from this point of view because moral judgment involves an inference from behavior to quality of will. The Value of Choice theory, on the other hand, begins with the value for an agent of having outcomes depend (or not depend) on his or her subjective responses under certain conditions. This (so far purely personal) value takes on moral significance by being the basis for a claim against social institutions (or against other individuals). In my view, to show that a social institution is legitimate one must show that it can be justified to each person affected by it on grounds which that person could not reasonably reject. One thing which people may reasonably demand, however, is the ability to shape their lives and obligations through the exercise of choice under reasonably favorable conditions. Moral principles or social institutions which deny such opportunities when they could easily be provided, or which force one to accept the consequences of choice under extremely unfavorable conditions which could be improved without great cost to others, are likely

to be reasonably rejectable for that reason. Let me illustrate by considering some examples.

2. JUSTICE AND CHOICE

Consider first the economic justice example which I mentioned earlier, Suppose a society, not marked by significant economic inequalities, decides that it needs to have a significant proportion of its workforce work overtime at a particular job. To this end, a bonus is offered to anyone willing to undertake the work, at an amount calculated to elicit the required number of volunteers. The choice between extra pay and extra leisure has obvious instrumental value for the people involved, and giving people this choice makes it overwhelmingly likely that those who prefer additional income (with additional labor) will get it, while those who prefer the opposite will get what *they* prefer. If overtime work was not made dependent on choice the scheme would be very difficult to justify; with this feature, justification is much easier. Nonetheless, whether or not a given worker winds up among those with extra pay will no doubt depend on some "morally arbitrary" facts about his or her background. Why then is this situation any better than the one criticized by Rawls?

The difference does not lie in the "fact" that the choices made in one case have causal antecedents while those made in the other case do not. In the egalitarian case, however, we can say that by placing the people in those circumstances, offering them that choice, and letting the outcome be determined by the choice they make under those conditions, we have done as much for them as could reasonably be required. In the other case it may be argued that we cannot say this: once the people are placed in disadvantageous circumstances, circumstances which themselves make it very unlikely that anyone would make the choices necessary to escape, offering these people the opportunity to exert themselves does little to improve their position.

The background conditions under which choices are made in the laissez-faire system are "arbitrary from a moral point of view" in this sense: they could be almost anything. All we know is that they will be conditions which arose from a series of voluntary transactions, and this does nothing to ensure that they will be good conditions under which to choose. Consequently, there is no assurance that these conditions will have the moral property of being conditions under which choices confer legitimacy on their outcomes.

This interpretation of Rawls's objection to the laissez-faire "system of natural liberty" provides the basis for a reply to one line of criticism raised by Nozick and others. Nozick interprets Rawls as arguing that the fact that some people exert themselves, take risks, and excel while others do not do so cannot by itself justify different economic rewards for the two groups because these differences in motivation may be the result of causal factors outside the control of the agents themselves. He goes on to object that

> this line of argument can succeed in blocking the introduction of a person's autonomous choices and actions (and their results) only by attributing *everything* noteworthy about the person completely to certain sorts of "external" factors. So denigrating a person's autonomy and prime responsibility for his actions is a risky line to take for a theory that founds so much (including a theory of the good) upon persons' choices.[24]

The problem which Nozick raises here is a version of the "political problem of free will" as I presented it in my first lecture. My reply (I do not claim that this was also Rawls's intention) is that it is not mere attributability to "external" factors that undermines the legitimating force of the choices in a "system of natural liberty." The problem, rather, is that such a system

[24]Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974), p. 214.

provides no assurance that these factors will not be ones which undermine the value of choice for many people in the society. Suppose that I exert myself to develop my talents and become wealthy. You, on the other hand, suffering the psychological effects of your unfortunate starting position, fail to exert yourself, and as a consequence remain poor. Can I "claim credit" for my initiative and perseverance, given that they resulted from "fortunate family and social circumstances for which [I] can claim no credit"?[25] If to "claim credit" means simply to consider these traits and actions "mine" in the sense required in order to take pride in them, then the answer is clearly yes. My accomplishments reflect personal qualities which I really do have. If, however, what is meant is that these differences in our behavior can be taken to justify my having more income and your having less, then the answer may be no. This is not because my actions, being caused by outside factors, are not "mine," or because your actions, similarly caused by other factors, are therefore not "yours," but rather because presenting a person with a choice of the kind you had is not doing enough for that person.

Of course, Rawls and Nozick disagree over what constitutes "doing enough" for a person. For Nozick, one has "done enough" as long as the person's Lockean rights have not been violated; for Rawls, the standard is set by the principles which would be accepted behind the Veil of Ignorance. As a result, Rawls's remarks about "factors arbitrary from a moral point of view," as I have interpreted them, may seem not to advance his argument against Notick but merely to restate the disagreement between them. But this restatement seems to me to have several virtues. First, it locates the disagreement in what seems, intuitively, to be the right place — in a question of justice rather than in a separate (and I believe spurious) question of causal determination. Second, framing the argument in terms of the value of choice has

[25]Rawls, *A Theory of Justice,* p. 104; quoted by Nozick, *Anarchy, State, and Utopia*, p. 214.

the effect of disentangling the idea of individual liberty from Nozick's particular system of Lockean rights. This allows opponents of that system to make clear that they, too, value individual choice and liberty and gives them a chance to put forward their alternative interpretations of these values. The argument can then proceed as a debate about the merits of competing interpretations of the moral significance of liberty and choice rather than as a clash between defenders of liberty and proponents of equality or some other pattern of distribution.

The Value of Choice theory represents a general philosophical strategy which is common to Hart's analysis of punishment and Rawls's theory of distributive justice as I have just interpreted it. In approaching the problems of justifying both penal and economic institutions we begin with strong pretheoretical intuitions about the significance of choice: voluntary and intentional commission of a criminal act is a necessary condition of just punishment, and voluntary economic contribution can make an economic reward just and its denial unjust. One way to account for these intuitions is by appeal to a preinstitutional notion of desert: certain acts deserve punishment, certain contributions merit rewards, and institutions are just if they distribute benefits and burdens in accord with these forms of desert.

The strategy I am describing makes a point of avoiding any such appeal. The only notions of desert which it recognizes are internal to institutions and dependent upon a prior notion of justice: if institutions are just then people deserve the rewards and punishments which those institutions assign them. In the justification of institutions, the notion of desert is replaced by an independent notion of justice; in the justification of specific actions and outcomes it is replaced by the idea of legitimate (institutionally defined) expectations.[26]

In order for this strategy to succeed, the conception of justice by which institutions are to be judged must adequately represent

[26]Rawls, *A Theory of Justice*, p. 313.

our intuitions about the significance of choice without falling back on a preinstitutional concept of desert. This is where the idea of the value of choice comes in. Just institutions must make outcomes depend on individuals' choices because of the importance which individuals reasonably attach to this dependence. But there is a serious question whether this strategy can account for the distinctive importance which choice appears to have. Insofar as choice-dependence is merely one form of individual good among others, it may seem that the Value of Choice theory will be unable to explain our intuition that the moral requirement that certain outcomes be made dependent on people's choices is not to be sacrificed for the sake of increases in efficiency, security, or other benefits.

Several defenses can be offered against this charge. The first is to point out the distinctiveness of the value of choice as compared with other elements in a person's welfare. As I have indicated above, the value of choice is not a purely instrumental value. People reasonably attach intrinsic significance to having outcomes depend on their choices. In addition, the moral requirements which this value gives rise to within a contractualist moral theory are not corollaries of a more general duty to look out for people's welfare. In fact, the demand to make outcomes depend on people's choices and the demand to promote their welfare are quite independent, and they can often pull in opposite directions.

A second defense — parallel to Rawls's argument for the priority of liberty — is to argue that in appraising social institutions people would reasonably set a particularly high value on having certain kinds of outcomes be dependent on their choices.[27] A third, more pragmatic defense is to argue that the distinctive significance which choice appears to have is in part an artifact of the position from which we typically view it. This is a position internal to institutions, and one in which choices have special salience because they are the last justifying elements to enter the

[27]See section 82 *of A Theory* of *Justice.*

picture. When the relevant background is in place — when conditions are right, necessary safeguards have been provided, and so on — the fact that a person chooses a certain outcome may make that outcome one that he or she cannot reasonably complain of. But choice has this effect only when these other factors are present. Because they are relatively fixed features of the environment, these background conditions are less noticeable than the actions of the main actors in the drama, but this does not mean that they are less important.

These defenses are most convincing in those cases in which the first argument is strongest — that is, in cases like the economic justice example just discussed, in which people's desire to shape their own lives gives choice an important, positive value. The Value of Choice theory looks weaker in cases where the only reason for wanting to have a choice is that it makes certain unwanted outcomes (such as punishment) less likely. Here choice has no positive value — rather than have the choice, one would prefer to eliminate these outcomes altogether if that were possible — yet the fact of choice seems to retain its special significance as a justifying condition. Let me turn, then, to an example of this kind.

3. Choice and Protection

Suppose that we, the officials of a town, must remove and dispose of some hazardous waste. We need to dig it up from the illegal dump near a residential area where it has lain for years and move it to a safer spot some distance away. Digging it up and moving it will inevitably release dangerous chemicals into the atmosphere, but this is better than leaving it in its present location, where it will in the long run seep into the water supply. Obviously we must take precautions to minimize the risks involved in this operation. We need to find a safe disposal site, far away from where people normally have to go. We should build a high fence around the new site, and another around the old one where

the excavation is to be done, both of them with large signs warning of the danger. We should also arrange for the removal and transportation to be carried out at times when few people are around, in order to minimize the number potentially exposed, and we must be sure to have the material wetted down and transported in covered trucks to minimize the amount of chemicals released into the air. Inevitably, however, enough chemicals will escape to cause lung damage to those who are directly exposed if, because of past exposure or genetic predisposition, they happen to be particularly sensitive, but not enough to pose a threat to anyone who stays indoors and away from the excavation site. Given that this is so, we should be careful to warn people, especially those who know that they are at risk, to stay indoors and away from the relevant area while the chemicals are being moved.

Suppose that we do all of these things but that nonetheless some people are exposed. A few of these, who did not know that they were particularly sensitive to the chemical, suffer lung damage. Let me stipulate that with respect to all of these people we did all that we could reasonably be expected to do to warn and protect them. So in that sense they "can't complain" about what happened. The question which concerns me, however, is what role the signs and warnings play in making this the case. These are the factors which make outcomes depend on people's choices. Are they, like the fences, the careful removal techniques, and the remote location of the new site, just further means through which the likelihood of someone's being injured is reduced? This is what the Value of Choice theory seems to imply. For after all, since no one wants to have the opportunity to be exposed to this chemical, the only value which choice can have in this case is that of making exposure less likely. This may be an adequate explanation of why we would want to be warned and hence "given the choice" whether to be exposed or not. But it may not account for the full moral significance of the fact that those who were injured "knew what they were getting into." Consider the following two cases.

Suppose that one person was exposed because, despite the newspaper stories, mailings, posted signs, radio and television announcements, and sound trucks, he never heard about the danger. He simply failed to get the word. So he went for his usual walk with no idea what was going on. A second person, let us suppose, heard the warnings but did not take them seriously. Curious to see how the task was being done, she sneaked past the guards and climbed the fence to get a better look.

There seems to be a clear difference between these two cases. In the first, we have "done enough" to protect the person simply because, given what we have done, it was extremely unlikely that anyone would be directly exposed to contamination, and we could not have made this even more unlikely without inordinate expense. There is, after all, a limit to the lengths to which we must go to protect others. The second person, on the other hand, bears the responsibility for her own injury, and it is this fact, rather than any consideration of the cost to us of doing more, which makes it the case that she has no claim against us. By choosing, in the face of all our warnings, to go to the excavation site, she laid down her right to complain of the harm she suffered as a result.

4. The Forfeiture View

This familiar and intuitively powerful idea about the significance of choice, which I will call "the Forfeiture View," is not captured by either of the theories I have been considering. It is distinct from the Value of Choice theory, since on that theory what matters is the value of the choice a person is presented with: once a person has been placed in a sufficiently good position, the outcome which emerges is legitimate however it may have been produced. On the Forfeiture View, on the other hand, it matters crucially that an outcome actually resulted from an agent's conscious choice, the agent having intentionally passed up specific alternatives. This is why that view accounts so well for our reaction to the person in the second example: not only does she have

no one else to blame for her fate; she has *herself* to blame. We could account for this sense of blame by appealing to a prudential version of the Quality of Will theory: the process of deliberation leading to a decision to climb over the fence "just to see what they are doing" is obviously faulty. But the Quality of Will theory is an account of the moral appraisal of agents, while what we are concerned with here is the justification of outcomes. It may be natural to suppose that a difference in the first translates into or supports a difference in the second, but on reflection it is by no means obvious how this is so.

Moreover, the idea of fault is in fact irrelevant here. The intuition to which the Forfeiture View calls attention concerns the significance of the fact of choice, not the faultiness of that choice. We can imagine a person who, unlike the imprudently curious woman in my example, did not run the risk of contamination foolishly or thoughtlessly. Suppose this third person found, just as the excavation was about to begin, that the day was a perfect one for working on an outdoor project to which she attached great value. Aware of the danger, she considered the matter carefully and decided that taking into account her age and condition it was worth less to her to avoid the risk than to advance her project in the time she was likely to have remaining. Surely this person is as fully "responsible for her fate" as the imprudent woman whom I originally described. But her decision is not a foolish or mistaken one.

This illustrates the fact that what lies behind the Forfeiture View is not an idea of desert. That is, it is not an idea according to which certain choices, because they are foolish, immoral, or otherwise mistaken, positively merit certain outcomes or responses. The idea is rather that a person to whom a certain outcome was available, but who knowingly passed it up, cannot complain about not having it: *volenti non fit iniuria.*

It is important to remember here that the challenge of the Forfeiture View lies in the suggestion that the Value of Choice

theory gives an inadequate account of the significance of choice *in the justification of institutions, policies, and specific moral principles.* Once we have accepted as justified an institution or policy attaching specific consequences to particular choices, there is no disagreement about whether these choices have the kind of special force which the Forfeiture View claims. This force can be accounted for by appeal to the institutions, principles, or policies in question. The disagreement concerns the way in which such institutions, principles, and policies themselves are to be justified. When the Forfeiture View says that people who make certain choices "cannot complain" about the harms they suffer as a result, what is meant is that these harms lack the force in this process of justification which otherwise comparable harms would have.

It may seem that a view of this kind is in fact forced on us by contractualism. According to contractualism the crucial question about a proposed moral principle is whether anyone could reasonably reject it. In order for rejecting a principle to be reasonable it must at least be reasonable from the point of view of the person doing the rejecting, that is, the person who would bear the burden of that principle. It may seem, therefore, that a harm which an agent has the opportunity to avoid (without great sacrifice) could never serve as a ground for reasonable rejection of a moral principle. Consider the following argument. From the point of view of an agent, an action which he has the choice of performing must be seen as available to him. Suppose that an agent will run the risk of suffering a certain harm if he follows one course of action but that he would avoid this harm if he were to follow an alternative course which is available to him and does not involve significant sacrifice. Given, then, that the harm is from his point of view costlessly avoidable, how could the agent appeal to this harm as grounds for objecting, for example, to a principle freeing others from any duty to prevent such harms from occurring? It would seem that such harms can have no weight in moral argument.

But this conclusion is not forced on us. In moral argument we are choosing principles to apply in general to situations in which we may be involved. Even if we know that actions avoiding a certain unwanted outcome will be available to us in a given situation, we also know that our processes of choice are imperfect. We often choose the worse, sometimes even in the knowledge that it is the worse. Therefore, even from the point of view of an agent looking at his own actions over time, situations of choice have to be evaluated not only for what they make "available" but for what they make it likely that one will choose. It is not unreasonable to want to have some protection against the consequences of one's own mistakes.

5. REJECTING THE FORFEITURE VIEW

The appeal of the Forfeiture View can and should be resisted. Note, first, that the Value of Choice theory can account for the apparent difference between the two victims of hazardous waste removal described above. We may have "done enough" to protect the first person, who failed to hear of the danger, in the sense that we have gone to as much effort and expense as could be expected. But because we did not succeed in making him aware of the danger we did not make what happened depend on his choice. Given that this kind of "choice-dependence" is something which we all would want for ourselves — we want such risks to be, as far as possible, "under our control" — we did not make this person as well off as we would reasonably want to be. The second person, on the other hand, did have the benefit of "having the choice," even though this turned out to be worth less to her than it would be to most of us. (There was in this case a divergence between "individualized" and "normalized" value.) Given that she had the choice, however, and was provided with the other protections, it was true of her in a way that it was not of the first person that she was placed in as good a position as one could ask for.

From the fact that a person chose, under good conditions, to take a risk, we may conclude that he alone is responsible for what happens to him as a result. But this conclusion need not be seen as a reflection of the special legitimating force of voluntary action. Rather, the fact that an outcome resulted from a person's choice under good conditions *shows* that he was *given* the choice and provided with good conditions for making it, and it is these facts which make it the case that he alone is responsible. A conscious decision to "take the risk" is not necessary. Consider, here, the case of a person who was informed of the risk of contamination but then simply forgot. As a result, he was out in his yard exercising, breathing hard, when the trucks went by. If enough was done to protect and warn him, then this person is responsible for what happens to him and "cannot complain of it" even though he made no conscious decision to take the risk.

The central element of truth in the Forfeiture View is thus a consequence of the Value of Choice theory rather than an alternative to it. Putting this truth in terms of the Forfeiture View, however, has the distorting effect of suggesting that choice has independent deontic force in the justification of institutions and principles. It also exaggerates the importance of the fact of choice relative to that of the conditions under which the choice was made. The Forfeiture View suggests that these conditions are important only insofar as they bear on the voluntariness of the choice. This is a mistake. The fact that a choice was voluntary does not always establish that we "did enough" for an agent by placing him or her in the position from which the choice was made. Nor does the fact that an agent did not voluntarily choose an outcome, or choose to take a certain risk, establish that what resulted was not his fault. Giving him the *opportunity* to choose may have constituted "doing enough" to protect him. It is thus an important virtue of the Value of Choice theory that it gives the conditions of choice their appropriate independent weight and forces us to keep them clearly in view.

6. RESPONSIBILITY AND THE MORAL DIVISION OF LABOR: BEYOND CHOICE

Within the Value of Choice theory, ideas of responsibility arise as a derived (and often only implicit) moral division of labor. Because most people take themselves to be more actively concerned with the promotion of their own safety and well-being than others are, they want outcomes to be dependent on their choices even when this has only "avoidance value." Given this concern, "giving people the choice" under favorable conditions makes it extremely unlikely that they will suffer easily avoidable harms. We do not want the trouble and expense of supervising others' choices more closely, and do not want them to be supervising us. Therefore, we take the view that giving people the opportunity of avoiding a danger, under favorable conditions, often constitutes "doing enough" for them: the rest is their responsibility. So stated, this is not a principle but only a description of a general tendency in our moral thought. In particular, the idea of "favorable conditions," here left vague, must be filled in before any specific principle of responsibility is obtained, and this filling in will be done differently in the case of different risks and dangers.

This general analysis does, however, shed light on appeals to responsibility in cases in which the notion of choice seems out of place. The idea of freedom of thought, mentioned in my first lecture, is one such case. Another, which I will discuss briefly here, is the idea of responsibility for one's preferences.

This idea arises in the context of debates as to whether, for purposes of assessing claims of justice, people's welfare should be measured in terms of preference satisfaction or in terms of some objective standard of well-being such as what Rawls has called Primary Social Goods. Objective standards of this kind may seem unfair, since the same bundle of objective goods can yield quite different levels of satisfaction for people with different

preferences. Rawls has replied that someone who makes this objection "must argue in addition that it is unreasonable, if not unjust, to hold such persons responsible for their preferences and to require them to make out as best they can." To argue this, he says, "seems to presuppose that citizens' preferences are beyond their control as propensities or cravings which simply happen." The use of an objective standard like primary goods, on the other hand, "relies on a capacity to assume responsibility for our ends." The conception of justice which Rawls advocates thus

> includes what we may call a social division of responsibility: society, the citizens as a collective body, accepts responsibility for maintaining the equal basic liberties and fair equality of opportunity, and for providing a fair share of the other primary goods for everyone within this framework, while citizens (as individuals) and associations accept the responsibility for revising and adjusting their ends and aspirations in view of the all-purpose means they can expect, given their present and foreseeable situation. This division of responsibility relies on the capacity of persons to assume responsibility for their ends and to moderate the claims they make on their social institutions in accordance with the use of primary goods. Citizens' claims to liberties, opportunities and all-purpose means are made secure from the unreasonable demands of others.[28]

I am strongly inclined to agree with Rawls here, and I have defended a similar position myself.[29] Nonetheless, I find this argument somewhat worrisome, because it is easily misinterpreted as involving an appeal to the idea of forfeiture which I argued against above. On this interpretation, the argument is that the imagined objection to objective measures of welfare overlooks the fact that people's preferences are under their control. Given this

[28]John Rawls, "Social Unity and Primary Goods," in Amartya Sen and Bernard Williams, eds., *Utilitarianism and Beyond* (Cambridge: Cambridge University Press, 1982), pp. 168, 169, 170.

[29]In "Preference and Urgency," *Journal of Philosophy* 72 (1975): 655-69. The following discussion concerns issues dealt with in my reply to "the voluntariness objection" on pp. 664–66 of that article.

fact, and in view of the basic moral truth that one cannot complain of harms one could have avoided, the objection is no objection at all: people whose preferences are particularly difficult to satisfy have only themselves to blame.

There are two difficulties with this argument. First, for reasons I have already discussed, the "basic moral truth" to which it appeals seems open to serious doubt. Second, even if this "truth" is correct, the argument appears to exaggerate the degree of control which people have over their preferences. To be sure, the argument does not suggest that people can alter their preferences by simply deciding what to prefer; the kind of control which is envisaged is to be exercised through decisions affecting the development of one's preferences over time. Even so, it is questionable how much control of this kind people can realistically be assumed to exercise.

This leads me to look for an alternative interpretation under which the argument avoids these difficulties while still retaining its force. Following the general strategy which I have been advocating in this lecture, this alternative interpretation takes the idea of responsibility for one's preferences to be part of the view being defended rather than an independent moral premise. As Rawls says, the conception of justice which he is defending *includes* "what we may call a social division of responsibility." The question is how this combination — an objective standard of welfare and the idea of responsibility which it entails — can be defended without appeal to anything like the notion of forfeiture.

The issue here is the choice between two types of public standards of justice, objective standards of the sort just described, according to which institutions are judged on the degree to which they provide their citizens with good objective conditions for the development and satisfaction of their preferences, and subjective standards, under which institutions are also judged on the basis of the levels of preference satisfaction which actually result from their policies. In our earlier discussion of individual choice, the

argument for a "moral division of labor" rested on three claims: the value which we attach to having outcomes depend on our own choices (even when this is only "avoidance value"), our reluctance to have our choices supervised by others, and our reluctance to bear the costs of protecting others beyond a certain point. The case for the "social division of responsibility" entailed by objective standards of welfare rests on three analogous claims. We reasonably attach a high value to forming our own preferences under favorable conditions, and one reason for this is our expectation that we will to *some* extent be steered away from forming preferences when we can see that they will be difficult to satisfy and will lead mainly to frustration. Second, we do not want others to be taking an active role in determining what we will prefer. And third, we do not want to be burdened with the costs of satisfying other people's preferences when these are much more costly than our own.

The first of these claims accounts for the (limited) force of the idea, to which Rawls appeals, that people can to some extent avoid "costly" preferences. But it does this without invoking a preinstitutional notion of forfeiture, and without assuming the degree of conscious and deliberate control which the Forfeiture View would require.

The second claim is especially important. Particularly in a society marked by sharp disagreements about what is worth preferring, a public standard of justice requiring government policy to be aimed at raising individual levels of satisfaction is an open invitation to unwelcome governmental intervention in the formation of individuals' values and preferences. The "social division of responsibility" which goes with an objective standard of welfare is therefore an attractive alternative.

The case for an objective standard of welfare is thus largely defensive. Giving up the claim to a greater share of resources in the event that one's preferences turn out to be particularly difficult to satisfy is the price one pays for greater security against

governmental interference and greater freedom from the possibly burdensome demands of other people's preferences. The role of the possibility of modifying one's preferences (or of avoiding the formation of preferences which are difficult to satisfy) is just to make this price smaller and not, as the Forfeiture View would have it, to license the result.

7. CONCLUSION

In this lecture I have presented the idea of the Value of Choice as part of a general strategy explaining the moral significance of choice in the justification of social institutions and policies. As compared with its main rival, the Forfeiture View, this strategy has the advantage of assigning choice an important positive value without exaggerating its role and significance in justification. It remains to be seen what kind of freedom the Value of Choice theory presupposes and how it fits together with the Quality of Will theory to account for the significance of choice across a range of cases. These questions will be addressed in my next lecture.

Lecture 3

1. PUNISHMENT AND PROTECTION

Let me begin with a schematic comparison of the institution of punishment and the policy of hazardous waste disposal which I discussed in my last lecture. In each case we have the following elements. First, there is an important social goal: protecting the water supply in the one case; protecting ourselves and our possessions in the other. Second, there is a strategy for promoting that goal which involves the creation of another risk: the risk of contamination in the one case, the risk of punishment in the other. Third, the effect of this strategy is to make it the case that there is, literally or metaphorically, a certain affected area which one

can no longer enter without danger. In the one case this is the area of excavation, transport, and disposal, in the other the "area" of activities which have been declared illegal. Fourth, although we introduce certain safeguards to reduce exposure to the risk created, it remains the case that many of those who choose to enter the affected area, and perhaps a few others, will suffer harm. Some of these safeguards (such as requirements of due process, and careful methods of excavation and transport) have the effect of protecting those who choose to stay out of the affected area. Other safeguards enhance the value of choice as a protection by making it less likely that people will choose to enter. In the hazardous waste case these include signs, warnings, and publicity to inform people about the nature of the risk, as well as fences, guards, and the choice of an obscure disposal site where no one has reason to go. Analogous features in the case of punishment are education, including moral education, the dissemination of basic information about the law, and the maintenance of social and economic conditions which reduce the incentive to commit crime by offering the possibility of a satisfactory life within the law. Restrictions on "entrapment" by law enforcement officers also belong in this category of safeguards which make it less likely that one will choose badly. Without such safeguards the value of choice as a protection would be reduced to an unacceptable level.

In each case, in order to defend the institution in question we need to claim that the importance of the social goal justifies creating the risk and making the affected area unusable and that, given the prevailing conditions and the safeguards we have put in place, we have done enough to protect people against suffering harm from the threat that has been created.

Now let me turn to some of the differences between the two cases. First, insofar as the activities which make up "the affected area" in the case of punishment are ones which it is morally wrong to engage in, being deprived of the ability to "enter this area" without risk cannot be counted as a morally cognizable loss. This

makes the task of justification easier than in the example of hazardous waste.

A second difference makes this task more difficult, however. In neither case is it our aim that people should suffer the new harm, though in both cases the possibility of their doing so is created by our policy. But in the case of punishment this harm, when it occurs, is intentionally inflicted on particular people. It is an essential part of that institution that people who run afoul of the law should be punished; but it is no part of our waste-removal policy that those who enter the affected area should suffer contamination. If, as I believe, intentionally inflicting harm is in most cases more difficult to justify than merely failing to prevent harm, it follows that an institution of punishment carries a heavier burden of justification.

When such an institution *is* justified, however, this justification entails the kind of "forfeiture" which we looked for but did not find in the hazardous waste case. A person who intentionally commits a crime lays down his or her right not to suffer the prescribed punishment. This forfeiture is a consequence of the justification of the institution of punishment, however, not an element in that justification. It is a consequence, specifically, of the "heavier justificatory burden" just mentioned: because the institution assigns punishment to those who fulfill certain conditions, justifying the institution involves justifying the infliction of these penalties. If the conditions for punishment include having made a certain kind of choice, then a justification for the institution justifies making that choice a necessary and, when the other conditions are fulfilled, sufficient condition for punishment. No such assignment and hence no such forfeiture is involved in the justification of the policy of hazardous waste removal. A person who recklessly chooses to enter the affected area does not lay down a right to further protection against contamination: she has already received all the protection she is entitled to. She does not lay down her right to treatment (or rescue) unless this has been pre-

scribed and the policy including this prescription is justified. Forfeiture, like economic desert, is the creature of particular social institutions and relatively specific moral principles (such as those governing promising). It is not a moral feature of choice in general. As I argued in my last lecture, the moral aspect of choice which figures in the justification and criticism of such institutions and principles is not forfeiture but the less-sharp-edged notion of the value of choice.

I have been assuming that "the affected area" is so defined that one can "enter" it only by conscious choice. This will be so if we identify "entering" that area with committing a crime whose definition involves conditions of voluntariness and intent. But a system of criminal law incorporating elements of strict liability could also fit the abstract model I have described, If a legal penalty is attached to selling adulterated milk (not merely to doing so knowingly, recklessly, or negligently), then one "enters the affected area" simply by going into the milk business, and if such a law is justified then doing this involves laying down one's right not to be penalized if the milk one sells turns out to be impure. This enlargement of the affected area is one reason (perhaps not the only one) why such laws are more difficult to justify, especially since the newly affected area includes activities, such as conscientious engagement in the milk business, which people are morally entitled to engage in. Having them entail forfeiture of the right not to be punished is a morally cognizable loss.

2. Excuses and the Value of Choice

I said in my first lecture that an acceptable account of the significance of choice should be able to explain standardly recognized excusing conditions in a way that will not generalize to undermine the moral significance of all choice if the Causal Thesis is true. Let me now say something about how the Value of Choice theory fulfills this assignment. My aim here is not to derive particular excusing conditions or to define the notion of voluntariness

appropriate to particular social institutions and moral principles. This would be an extremely time-consuming task, since it is reasonable to suppose that these conditions will vary in detail from case to case. My present purpose is merely to point out in a more general way how the Value of Choice theory would account for these conditions and for their variation.

The general point is obvious. If the justification for a principle or institution depends in part on the value of the choices it presents people with, and if the value of these choices in turn can vary greatly depending on the presence or absence of certain conditions, then in order to be justifiable the institution will have to qualify the consequences it attaches to choices by explicitly requiring the presence or absence of the most important of these conditions.

Lack of knowledge of the nature of the alternatives available, lack of time to consider them, and the disruptive effects of fear or emotional distress can all weaken the connection between a person's reaction at a given time and his or her more stable preferences, values, and sensitivities, thus undermining both the predictive and demonstrative value of choice. Coercion and duress can have similar disrupting effects on the process of choice, but also and more often they diminish the value of choice simply by contracting or altering the set of alternatives between which one can choose. Diminishing the set of alternatives or weighting some with penalties can sometimes increase the value of choice — or so those of us must believe who sign up to give lectures we have not yet written and buy automobiles with seat belt buzzers. But this is not usually the case.

Even when duress, false belief, or other conditions clearly diminish the value of choice, however, it does not immediately follow that these conditions must be recognized as negating a particular obligation or liability. Whether it does or not will depend on, among other things, the costs to others of introducing such an exception into the principle or institution in question. This is a

further reason why, on the present theory, it is possible for excusing conditions to vary from principle to principle and institution to institution.

Here there is a clear contrast with the genesis of excusing conditions under the Quality of Will theory. Once we learn that an agent acted under duress or under the influence of a mistaken belief, this immediately alters the "will" attributable to that agent. There is no need to ask what the effect would be of recognizing this "excuse." Of course, such considerations are relevant to the further question of which "qualities of will" should be regarded as morally deficient. But the Quality of Will theory plays no role in answering this question; it is an account only of the process of moral appraisal.

A second contrast between the two theories is this. The Value of Choice theory treats changes in the set of alternatives available to a person and changes in the conditions under which he or she chooses among them as factors contributing to the answer to a single question: how good or bad a thing is it to be presented with that choice? Under the Quality of Will theory, on the other hand, there is an important difference here. Some conditions affect the degree to which a "will" can be imputed to the agent; others modify the nature of that will. This difference may explain Hart's remark that while continental jurisprudence has traditionally distinguished between imputability and fault he sees little to be gained by observing this rigid distinction.[30] This difference is to be expected insofar as Hart is speaking as a Value of Choice theorist while the continental tradition may be more concerned with aspects of the law akin to quality of will.

3. The Value of Choice and the Causal Thesis

I turn now to the question of whether choice will retain the moral significance which the Value of Choice theory assigns it if the Causal Thesis is true. Whether it does so or not will depend on

[30] *Punishment and Responsibility,* p. 218.

whether choice will retain its value for an individual if the Causal Thesis is true. This is at least part of what I called in my first lecture "the personal problem of free will." So it seems that the most that the Value of Choice theory could accomplish would be to reduce the political problem of free will to the personal problem.

The mere truth of the Causal Thesis would not deprive choice of its predictive value: a person's choices could remain indicative of his or her future preferences and satisfactions even if they had a systematic causal explanation. Nor, it seems to me, need the demonstrative value of choice be undermined. A person's choices could still reflect continuing features of his or her personality such as feelings for others, memory, knowledge, skill, taste, and discernment.

This is how things seem to me, perhaps because I am in the grip of a theory. It is difficult to support these intuitions by argument because it is difficult, for me at least, to identify clearly the basis of the intuitions which move one toward the opposite conclusion. It might be claimed that what I have called the demonstrative value of choice would be undermined because the feelings, attitudes, and so on which a person's choices might be taken to "reflect" will no longer "belong" to that person if the Causal Thesis is true, but it is not clear why this should be the case. It is easy to see that particular kinds of causal history might make a belief or desire "alien." This would happen when, as in the "implantation" examples mentioned above, the special causal genesis of a belief meant also that it lacked connection with the person's other conscious states — that it was not all dependent on other beliefs and desires for support and not subject to modification through the agent's process of critical reflection. But it does not seem that this kind of loss of connection need hold generally if the Causal Thesis is correct.

One can certainly imagine a form of causal determination which would make this kind of alienation hold generally and

would make it inappropriate to speak of a person's holding beliefs and attitudes at all. A person's conscious states might be caused to occur in a pattern which made no sense at all "from the inside," following one another in a random and meaningless sequence preserving no continuity of belief or attitude. It might be argued that the "normal case" is more like this than we are inclined to suppose: that our idea of the coherence and regularity of our conscious life is to a large degree an illusion. This might undermine the sense of self on which the value of choice depends. But this, if true, would be the result of a particular substantive claim about the order and coherence of the events that make up our "mental lives." It would not be a consequence of the bare Causal Thesis itself.

4. Freedom and Overdetermination

The kind of freedom required by the Value of Choice theory is in one respect more extensive than that required for moral appraisal of the kind discussed in my first lecture. This difference can be brought out by considering how the ideas of quality of will and value of choice apply to overdetermination cases of the kind introduced by Harry Frankfurt.[31] Frankfurt's central example involves two drug addicts. It is assumed that neither is capable of resisting the pull of his addiction: both will take the drug when it is offered, and neither could do otherwise. But while one, the "unwilling addict," would prefer that the desire to take the drug not be the one which he acts on, the other, "the willing addict," not only has a desire for the drug but also has the "second-order desire" to act on that desire. Frankfurt believes that the latter addict acts freely in the sense required for moral responsibility but that the former does not. What interests me here is the fact that the two theories I have presented appear to give different answers to the question of freedom in cases like that of Frankfurt's willing addict — that is to say, cases in which (for reasons which may or

[31]In "Freedom of the Will and the Concept of a Person."

may not be like those in Frankfurt's particular example) a person has no alternative to doing a certain thing but nonetheless gets what he wants or does what he is inclined to do. If the question is whether the action reflects the agent's quality of will, then cases like that of Frankfurt's willing addict seem to be cases of freedom. (This answer agrees with Frankfurt, which is not surprising given that he is concerned specifically with moral responsibility.) If, on the other hand, the question is whether the agent has been given a fair chance to make outcomes conform to or exhibit his or her preferences and abilities, then the answer seems to be no, and the cases count as instances of unfreedom.

It may seem that this difference is illusory. The question under the Value of Choice theory is whether there was the right kind of opportunity for the person's disposition to choose to be discovered and registered. Insofar as it is predictive value we are concerned with, the assumption is that "we" do not generally know in advance what a person's preference is: we are trying to set up a social mechanism to discover this and react to it. In Frankfurt's cases, however, it is assumed that *we* know the addicts' (first- and second-order) preferences. Indeed, we are assumed to know more about this than agents themselves normally do. The question of how these preferences might be discovered is not at issue in Frankfurt's discussion. But this question can arise with respect to moral responsibility. Administering praise and blame is something we do, and it is relevant to ask whether we have adequate grounds for doing so: whether it is fair to judge a person on the basis we have. This is like the question which arose in application of the Value of Choice theory: whether there was adequate opportunity for the person's preferences, whatever they may have been, to be revealed.

This same question of fairness can also be raised when we are only forming an opinion about an agent's blameworthiness, without intending to express it. But the question whether the agent is blame*worthy* goes beyond these questions of adequate grounds,

and it is the question which is fundamental: if the person's will in doing the action was of the appropriate sort, then a certain moral judgment is in fact applicable, whether or not any particular person is in a position to make it. Insofar as this is the case, the difference between the two theories that was pointed out above still stands.

Of course, parallel to the fact that a person "really was blameworthy" in acting a certain way, there is the fact that a person "really did want X, which was what he got," and this too might be held to be the fundamental fact, on the basis of which we could ask, How can he complain, since he got what he wanted? But this fact of preference is not fundamental in the way that the fact of blameworthiness is: the two facts are differently related to the moral ideas on which the theories in which they figure are based. The Quality of Will theory is based on the idea that the applicability of moral praise and blame depends on what the quality of will expressed in an action actually was. In determining this quality we may need to know what the agent believed the alternatives to be, but the question of which of these were actually available is in at least some cases irrelevant. Under the Value of Choice theory, however, the basic moral idea is not simply that people should get what they want but that things should be set up so that outcomes are made dependent on people's choices. In overdetermination cases this demand may not have been met, even though, as it happens, the person is in certain respects no worse off as a result.

5. The Two Theories Combined

I have described two theories and said something about how they are related to one another. It remains to be seen how these two theories, when combined, cover the territory. I have so far employed the Value of Choice theory mainly to give an account of the significance of choice in "political" cases, and I have relied upon the Quality of Will theory in discussing moral responsibility.

But this division of labor is overly simple. In fact, both analyses are required to account for the significance of choice in morality, and both are required to explain its force in the law.

Let me take the moral case first. Suppose you think that I promised on Monday to pick up your child at school on Tuesday but then failed to do this. There are two ways in which considerations of voluntariness and choice might enter into an assessment of how blameworthy I am on this account. First, such considerations could undermine my blameworthiness by making it the case that I had no obligation to pick up your child in the first place. It could be that I never assented to your request: when I said yes, it was to something else, and I never heard your request at all. Or perhaps I did assent to your request but only because you threatened me or concealed from me the fact that I would have to wait three hours beyond the normal end of the school day. Factors such as these could erase or modify my obligation.

On the other hand, it could be that while I did indeed incur an obligation to you, my not meeting your child was not due to any failure on my part to take my obligation seriously and try to fulfill it. It might be that I was hit over the head and knocked unconscious just before I was to leave, or that my car broke down on the way, leaving me stranded in a deserted spot.

These two kinds of excusing conditions are quite different. Something like the Value of Choice theory seems to provide the best explanation of why moral obligations are qualified by restrictions of the first sort. As Hart suggested, a system for the making of binding agreements, whether moral or legal, is defensible only if it is constrained by restrictions to ensure that the obligations one acquires are obligations one judges to be worth acquiring. The assessment of quality of will has at most a secondary role here.

Things are reversed in a case of involuntary nonfulfillment of a valid obligation. Here the natural value of choice analysis (modeled on that analysis of the choice requirement for criminal

punishment) would be that a morality which held agents liable to blame in such cases would be objectionable because it gave people insufficient "protection" against incurring the sanction of moral blame. This is clearly not the right explanation. It is wrong because it treats moral blame simply as a "sanction" which people would like to avoid, which we attach to certain actions although it could just as well be attached to others (e.g., to things that are done involuntarily). This ignores the distinctive content of moral blame, in virtue of which it is not simply another kind of unpleasant treatment, like being shunned. Morality is, at base, a system of mutually authorizable deliberation. To feel oneself subject to moral blame is to be aware of a gap between the way one in fact decided what to do and the form of decision which others could reasonably demand. The absence of such a gap is by itself a sufficient explanation of why blame is inapplicable in cases like that of the person who, despite his or her best efforts, fails to pick up the child. There is no need to refer to the kind of question which the Value of Choice theory addresses.

This internal connection between the nature of "the moral sanction" and the content of morality — between the nature of blame and the things one can be blamed for — differentiates morality from a social institution set up to serve certain extrinsic purposes. Of course there could be a social practice according to which people would be subject to scolding and shunning in cases for actions involving no faulty willing or deliberation, but what was expressed by this behavior would not be moral blame. Even without such a practice there is a question, distinct from that of blameworthiness, of whether one has good reason to engage in "blaming behavior" toward a given person on a given occasion. As I mentioned in my first lecture, even when people are blameworthy it might be callous to scold them, and the reverse may also be true. For example, even though very young children are not blameworthy it may be important for their moral education to treat them as if they were.

The issues raised here are similar to those which arise in connection with what Hart called the "definitional stop" argument against exemplary or vicarious punishment of persons known to be innocent of any off ense.[32] A utilitarian justification of punishment, insofar as it is a justification of *punishment,* could not justify such practices, this argument ran, because these practices do not count as punishment, which, by definition, must be of an offender for an offense. The obvious response to this argument is that it is not important what we call it; the question is why it would not be permissible to subject people, known to be innocent, to unpleasant treatment (prison, fines, etc.) as part of a scheme to intimidate others ,into obeying the law. As I have said above, I agree with Hart that the Value of Choice theory provides a good (though perhaps not fully satisfying) answer to this question. With respect to moral blame, however, I have responded in effect that it matters a great deal what you call it, because blameworthi*ness,* rather than any form of "blaming behavior," is the central issue. There is also, of course, a question of the desirability and permissibility of expressing or administering blame in a certain way, but this is a separate question and a secondary one.

In the case of criminal punishment this emphasis is reversed: the main question is whether we can justify depriving people of their property, their liberty, or even their lives.[33] Despite the

[32]*Punishment and Responsibility,* pp. 5-6.

[33]In a recent article, R. B. Brandt put forward something like the Quality of Will theory as a limitation on legal punishment. See "A Motivational Theory of Excuses in the Criminal Law," in J. R. Pennock and J. W. Chapman, eds., *NOMOS XXVII*: *Criminal Justice* (New York: New York University Press, 1985), pp. 165–98. Specifically, Brandt defends the principle that a condition should be recognized as excusing a person from legal blame if the presence of that condition "blocks the normal inference" from the fact that the agent performed a certain act to the conclusion that the agent's motivation is defective. His defense of this principle appeals to the value of assuring people that if they lack "defective motivation" they will almost certainly not be punished. This is reminiscent of Hart and the Value of Choice theory, but Brandt's defense is avowedly rule-utilitarian: he sees the value in question merely as a contribution to the general welfare, not as fulfilling a special requirement of fairness to the individual. Moreover, he sees the requirement of "defective motivation" as a replacement for Hart's notion of "capac-

changed emphasis, however, both elements are still present, and consequently it does "matter what you call it" even if this consideration does not settle the crucial question of justification. The law is not just an organized system of threats. It also provides rules and standards which good citizens are supposed to "respect," that is, to employ as a way of deciding what to do — not simply as a way of avoiding sanctions but as a set of norms which they accept as reason-giving. This important feature of law offers a further reason why the Value of Choice theory was not completely satisfying as an explanation of the choice requirement for criminal punishment. Insofar as punishment is in part an expression of "legal blame," as Feinberg and others have pointed out,[34] there is a special inappropriateness in having it fall on persons who have deliberated and acted just as the law says they should. The Value of Choice theory thus fails to be a complete account of the significance of choice in the law for much the same reason that it fails to be a complete account in the case of morality. In each case there is *something* to the "definitional stop."

Something, perhaps, but in the case of the law, how much? Pointing out "the expressive function of punishment" helps us to understand our reactions to punishing particular kinds of people, but what role if any does it have in the justification of punishment? It seems to have no positive role in justifying hard treatment of the legally blameworthy. Insofar as expression is our aim, we could just as well "say it with flowers" or, perhaps more appropriately, with weeds. Nor, it seems, is this idea the central explanation of the apparent wrongfulness of punishing, say, young children or the mentally ill. Assuming that these people lack the

—

ity and fair opportunity" to avoid punishment *(ibid*., p. 180). My analysis is similar to Brandt's in a number of respects, but, unlike him, I see quality of will and the value of choice as two independent (though related) *reasons* for the limits of moral and legal blameworthiness. Since they are related, it is not surprising that these two kinds of reasons often support the same limits. But they do not always do so.

[34]Joel Feinberg, "The Expressive Function of Punishment," in *Doing and Deserving* (Princeton: Princeton University Press, 1970).

capacity for critically reflective, rational self-governance, we could argue, as we did in the case of morality, that they cannot be legally blameworthy. But even in the case of morality, the justification of "blaming behavior" is a separate issue from that of blameworthiness, and here it is a much weightier one in view of the losses that the law can inflict.

The Value of Choice theory offers a more plausible explanation. According to that theory the lack of the normal capacity for critically reflective, rational self-governance is relevant because people who lack it are so unlikely to be deterred. This may or may not make punishment pointless for us, but it certainly makes it unfair to them: we must protect them against punishment just as, in my other example, we must post barriers or guards to keep people with Alzheimer's disease away from the hazardous waste. But within the Value of Choice theory the normal capacity for critically reflective, rational self-governance lacks the *distinctive* importance which it has when moral (or legal) blameworthiness is at issue. There are many people who have this capacity yet will not be deterred. It is easy to say why they are blameworthy, but why should we respond differently to their suffering than to that of the mentally ill? We can say that, because they have this normal capacity for self-governance, deterrence is a plausible strategy for us to use in dealing with them and that the possibility of their being deterred is, from their point of view, *some* measure of protection. If it turns out not to be enough, then the best we can say, if it is true, is that we did as much as we could be expected to do to protect them.

At some moments it seems to me that we must be able to say more — that choice has a further significance not captured by either of the theories I have considered, perhaps something more like what the Forfeiture View is straining toward. At other times, however, it seems to me an advantage of the combined theory I have been defending, and a natural consequence of its aspiration to be compatible with the Causal Thesis, that it leaves us in this

position: moral and (if there is such a thing) legal indignation toward lawbreakers is entirely in order, and the sufferings we inflict upon them may be justified. But in justifying these sufferings, and inflicting them, we have to say not "You asked for this" but "There but for the grace of God go I."