

position which allows that as a matter of metaphysical fact we might not be free and responsible beings—even if this gives us no reason at all to regard ourselves as such—can properly be called optimistic. The pessimist will only give up his pessimism if the possibility of this state of affairs is directly refuted. In other words, for the pessimist, who asks for a justification of the fact that we treat ourselves as free and responsible beings, only one kind of justification will do—a justification, in particular, which relies on the fact that we *are* free and responsible beings.

## 4

## Responsibility and the Limits of Evil: Variations on a Strawsonian Theme

GARY WATSON

Responsibility is . . . one aspect of the identity of character and conduct. We are responsible for our conduct because that conduct is ourselves objectified in actions.

—John Dewey, "Outlines of a Critical Theory of Ethics"

There is nothing regrettable about finding oneself, in the last analysis, left with something which one cannot choose to accept or reject. What one is left with is probably just oneself, a core without which there could be no choice belonging to the person at all. Some unchosen restrictions on choice are among the conditions of its possibility.

—Thomas Nagel, *The Possibility of Altruism*

Our practices do not merely exploit our natures, they express them.

—Peter Strawson, "Freedom and Resentment"

### INTRODUCTION

REGARDING people as responsible agents is evidently not just a matter of belief. So regarding them means something in practice. It is shown in an embrace or a thank you, in an act of reprisal or obscene gesture, in a

Reprinted by permission of Cambridge University Press from Ferdinand David Schoeman, ed., *Responsibility, Character and the Emotions* (1987), pp. 256–286.

To Sally Haslanger and Brian Skyrms, I am grateful for discussing bits and pieces of this material with me; to Ferdinand Schoeman, for comments on an earlier draft.

feeling of resentment or sense of obligation, in an apology or demand for an apology. To regard people as responsible agents is to be ready to treat them in certain ways.

In "Freedom and Resentment,"<sup>1</sup> Peter Strawson is concerned to describe these forms of treatment and their presuppositions. As his title suggests, Strawson's focus is on such attitudes and responses as gratitude and resentment, indignation, approbation, guilt, shame, (some kinds of) pride, hurt feeling, (asking and giving) forgiveness, and (some kinds of) love. All traditional theories of moral responsibility acknowledge connections between these attitudes and holding one another responsible. What is original to Strawson is the way in which they are linked. Whereas traditional views have taken these attitudes to be secondary to seeing others as responsible, to be practical corollaries or emotional side effects of some independently comprehensible belief in responsibility, Strawson's radical claim is that these "reactive attitudes" (as he calls them) are *constitutive* of moral responsibility; to regard oneself or another as responsible just is the proneness to react to them in these kinds of ways under certain conditions. There is no more basic belief which provides the justification or rationale for these reactions. The practice does not rest on a theory at all, but rather on certain needs and aversions that are basic to our conception of being human. The idea that there is or needs to be such an independent basis is where traditional views, in Strawson's opinion, have gone badly astray.

For a long time, I have found Strawson's approach salutary and appealing. Here my aim is not to defend it as superior to its alternatives, but to do something more preliminary. A comparative assessment is not possible without a better grasp of what Strawson's theory (or a Strawsonian theory)<sup>2</sup> is. As Strawson presents it, the theory is incomplete in important respects. I will investigate whether and how the incompleteness can be remedied in Strawsonian ways. In the end, I find that certain features of our practice of holding responsible are rather resistant to such remedies, and that the practice is less philosophically innocent than Strawson supposes. I hope that the issues uncovered by this investigation will be of sufficient importance to interest even those who are not as initially sympathetic to Strawson's approach as I am.<sup>3</sup>

1. This volume. Hereafter, page references are given in the text.

2. My interpretation of Strawson's essay will be in many places very conjectural; and I will sometimes signal this fact by speaking of a "Strawsonian" theory.

3. I have learned much from the penetrating exploration of Strawson's essay by Jonathan Bennett: "Accountability," in *Philosophical Subjects*, edited by Zak van Straaten, Oxford: Clarendon Press, 1980, pp. 14-47.

### Strawson's Theory

Strawson presents the rivals to his view as responses to a *prima facie* problem posed by determinism. One rival—consequentialism—holds that blaming and praising judgments and acts are to be understood, and justified, as forms of social regulation. Apart from the question of its extensional adequacy, consequentialism seems to many to leave out something vital to our practice. By emphasizing their instrumental efficacy, it distorts the fact that our responses are typically personal reactions to the individuals in question that we sometimes think of as eminently appropriate reactions quite aside from concern for effects. Rightly "recoiling" from the consequentialist picture, some philosophers have supposed that responsibility requires a libertarian foundation, that to bring the "vital thing" back in, we must embrace a certain metaphysics of human agency. This is the other rival.

What these otherwise very different views share is the assumption that our reactive attitudes commit us to the truth of some independently apprehensible proposition which gives the content of the belief in responsibility; and so either the search is on for the formulation of this proposition, or we must rest content with an intuition of its content. For the social-regulation theorist, this is a proposition about the standard effects of having and expressing reactive attitudes. For the libertarian, it is a proposition concerning metaphysical freedom. Since the truth of the former is consistent with the thesis of determinism, the consequentialist is a compatibilist; since the truth of the latter is shown or seen not to be, the libertarian is an incompatibilist.

In Strawson's view, there is no such independent notion of responsibility that explains the propriety of the reactive attitudes. The explanatory priority is the other way around: It is not that we hold people responsible because they *are* responsible; rather, the idea (*our* idea) that we are responsible is to be understood by the practice, which itself is not a matter of holding some propositions to be true, but of expressing our concerns and demands about our treatment of one another. These stances and responses are expressions of certain rudimentary needs and aversions: "It matters to us whether the actions of other people . . . reflect attitudes toward us of good will, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other." Accordingly, the reactive attitudes are "natural human reactions to the good or ill will or indifference of others toward us [or toward those we care about] as displayed in *their* attitudes and actions" (p. 53). Taken together, they express "the demand for the manifestation of a reasonable degree of good will or regard, on the part of others, not simply towards oneself, but towards all those on whose behalf moral indignation may be felt. . ." (p. 57).

Hence, Strawson accuses rival conceptions of "overintellectualizing" our practices. In their emphasis on social regulation, consequentialists lose sight of sentiments these practices directly express, without which the notion of moral responsibility cannot be understood. Libertarians see the gaping hole in the consequentialist account, but rather than acknowledging that "it is just these attitudes themselves which fill the gap" (p. 64), they seek to ground these attitudes in a metaphysical intuition—"a pitiful intellectualist trinket for a philosopher to wear as a charm against the recognition of his own humanity" (p. 64). Holding responsible is as natural and primitive in human life as friendship and animosity, sympathy and antipathy. It rests on needs and concerns that are not so much to be justified as acknowledged.

### *Excusing and Exempting*

To say that holding responsible is to be explained by the range of reactive attitudes, rather than by a commitment to some independently comprehensible proposition about responsibility, is not to deny that these reactions depend on a context of belief and perceptions in particular contexts. They are not mere effusions of feeling, unaffected by facts. In one way, Strawson is anxious to insist that these attitudes have no "rationale," that they neither require nor permit a "rational justification" of some general sort. Nevertheless, Strawson has a good deal to say about the particular perceptions that elicit and inhibit them. Reactive attitudes do have internal criteria, since they are reactions to the moral qualities exemplified by an individual's attitudes and conduct.<sup>4</sup>

Thus, reactive attitudes depend upon an interpretation of conduct. If you are resentful when jostled in a crowd, you will see the other's behavior as rude, contemptuous, disrespectful, self-preoccupied, or heedless: in short, as manifesting attitudes contrary to the basic demand for reasonable regard. Your resentment might be inhibited if you are too tired, or busy, or fearful, or simply inured to life in the big city. These are causal inhibitors. In contrast, you might think the other was pushed, didn't realize, didn't mean to. . . . These thoughts would provide reasons

4. Reactive attitudes thus permit a threefold classification. Personal reactive attitudes regarding others' treatment of one (resentment, gratitude, etc.); vicarious analogues of these, regarding others' treatment of others (indignation and approbation); self-reactive attitudes regarding one's own treatment of others (and oneself?) (guilt, shame, moral self-esteem, feeling obligated). Many of the reactive attitudes reflect the basic demand (on oneself and others, for oneself and others), whereas others (for example, gratitude) directly express the basic concern.

Contrary to some of Strawson's discussion, responsibility does not concern only other-regarding attitudes. You can hold yourself responsible for failing to live up to an ideal that has no particular bearing on the interests or feelings of others. It may be said that others cannot *blame* you for this failure; but that would be a moral claim.

for the inhibition of resentment. What makes them reasons is, roughly, that they cancel or qualify the appearance of noncompliance with the basic demand.<sup>5</sup>

In this way, Strawson offers a plausible account of many of the "pleas" that in practice inhibit or modify negative reactive attitudes. One type of plea is exemplified by the aforementioned reasons for inhibited sentiments. This type of plea corresponds to standardly acknowledged *excusing* conditions. It works by denying the appearance that the other failed to fulfill the basic demand; when a valid excuse obtains, the internal criteria of the negative reactive attitudes are not satisfied. Of course, justification does this as well, but in a different way. "He realized what he was doing, but it was an emergency." In general, an excuse shows that *one* was not to blame, whereas a justification shows that one was not to *blame*.

Strawson distinguishes a second type of plea. These correspond roughly to standard *exempting* conditions. They show that the agent, temporarily or permanently, globally or locally, is appropriately exempted from the basic demand in the first place. Strawson's examples are being psychotic, being a child, being under great strain, being hypnotized, being a sociopath ("moral idiot"), and being "unfortunate in formative circumstances." His general characterization of pleas of type 2 is that they present the other either as acting uncharacteristically due to extraordinary circumstances, or as psychologically abnormal or morally undeveloped in such a way as to be incapacitated in some or all respects for "ordinary adult interpersonal relationships."

In sum, type-2 pleas bear upon the question of whether the agent is an appropriate "object of that kind of demand for goodwill or regard which is reflected in ordinary reactive attitudes" (p. 51). If so, he or she is seen as a responsible agent, as a potential term in moral relationships, as a member (albeit, perhaps, in less than good standing) of the moral community. Assuming the absence of such exemptions, type-1 pleas bear upon the question of whether the basic demand has been met. These inhibit negative reactive attitudes because they give evidence that their internal criteria are not satisfied. In contrast, type-2 pleas inhibit reactive attitudes because they inhibit the demand those attitudes express (p. 52).

When reactive attitudes are suspended on type-2 grounds, we tend to take what Strawson calls an "objective view." We see individuals not as ones to be resented or esteemed but as ones to be controlled, managed, manipulated, trained. . . . The objective view does not preclude all emotions: "It may include repulsion and fear, it may include pity or even love," though not reciprocal adult love. We have the capacity to adopt an objective view toward capable agents as well; for certain kinds of thera-

5. Below, this remark is qualified significantly.

peutic relationship, or simply to relieve the "strains of involvement," we sometimes call upon this resource.

As we have seen, one of Strawson's concerns is to deny the relevance of any theoretical issue about determinism to moral responsibility. In effect, incompatibilists insist that the truth of determinism would require us to take the objective attitude universally. But in Strawson's view, when we adopt the objective attitude, it is never a result of a theoretical conviction in determinism, but either because one of the exempting pleas is accepted, or for external reasons—fatigue, for example, or relief from the strain of involvement. No coherent thesis of determinism entails that one or more of the pleas is always valid, that disrespect is never meant, or that we are all abnormal or undeveloped in the relevant ways. Holding responsible is an expression of the basic concern and the basic demand, whose "legitimacy" requires neither metaphysical freedom nor efficacy. The practice does not involve a commitment to anything with which determinism could conflict, or which considerations of utility could challenge.

### *Blaming and Finding Fault*

This is the basic view as Strawson presents it. For convenience, we may call it the expressive theory of responsibility. With certain caveats,<sup>6</sup> the expressive theory may be called a nonconsequentialist form of compatibilism; but it is not the only such form. It can be clarified by contrasting it with another.

Consider the following common view of blame and praise: To blame someone morally for something is to attribute it to a moral fault, or "shortcoming," or defect of character, or vice,<sup>7</sup> and similarly for praise. Responsibility could be constructed in terms of the propriety conditions of such judgments: that is, judgments to the effect that an action or attitude manifests a virtue or vice.<sup>8</sup>

As I understand the Strawsonian theory, such judgments are only part of the story. They indicate what reactive attitudes are reactions to

6. The term "compatibilism" denotes the view that determinism is compatible with responsibility. Hence it may presuppose that determinism is an intelligible thesis. Since Strawson seems skeptical about this presupposition, he might refuse this appellation.

7. See Robert Nozick, *Philosophical Explanations* (Harvard University Press, 1981, p. 224).

8. Such a view is hinted at by James Wallace: "Answers to [the question of when an action is fully characteristic of an excellence or a vice] are fundamental for an account of the conditions for the appropriateness of praise, blame, reward and punishment and for an account of the derivative notion of responsibility" (*Virtues and Vices*, Cornell University Press, p. 43). This also seems to be R. Milo's view in *Immorality* (Princeton University Press, 1984). I don't say that such a view is necessarily incompatibilist—it could be insisted that conduct fully exemplifies a virtue or a vice only if determinism is false (this is clearly the Abélardian view, discussed below)—but it is clear how a compatibilist version would go.

(namely, to the quality of the other's moral self as exemplified in action and attitude), but they are not themselves such reactions. Merely to cite such judgments is to leave out something integral to the practice of holding responsible and to the concept of moral responsibility (of being one to whom it is appropriate to respond in certain ways). It is as though in blaming we were mainly moral clerks, recording moral faults, for whatever purposes (the Last Assizes?).<sup>9</sup> In a Strawsonian view, blaming is not merely a fault-finding appraisal—which could be made from a detached and austere "objective" standpoint—but a range of responses to the agent on the basis of such appraisals.<sup>10</sup> These nonpropositional responses are constitutive of the practice of holding responsible.

I will have something to say later about the nature of these responses. Clearly they make up a wide spectrum. Negative reactive attitudes range from bombing Tripoli to thinking poorly of a person. But even those at the more covert and less retributive end of the spectrum involve more than attributions of defects or shortcomings of moral character. Thinking poorly (less well) of a person is a way of regarding him or her in view of those faults. It has subtle implications for one's way of treating and interacting with the other. (Where the other is dead or otherwise out of reach, these implications will be only hypothetical or potential.) It is the sort of attitude that is forsworn by forgiveness, which itself presupposes the attribution of (former) fault.

### SOME CRITICAL QUESTIONS

I turn now to certain hard questions for the expressive theory. It accounts nicely for "excusing conditions," pleas of type 1; but exactly—or even roughly—what is its account of type-2 pleas? The "participant" reactive attitudes are said to be "natural human reactions to the good or ill-will or indifference of others as displayed in their attitudes and actions" (p. 53); but this characterization must be incomplete, for some agents who display such attitudes are nevertheless exempted. A child can be malicious, a psychotic can be hostile, a sociopath indifferent, a person under great strain can be rude, a woman or man "unfortunate in formative circumstances" can be cruel. Evidently reactive attitudes are

9. Consider Jonathan Glover's remark: "Involved in our present practice of blame is a kind of moral accounting, where a person's actions are recorded in an informal balance sheet, with the object of assessing his moral worth." (*Responsibility*, Routledge and Kegan Paul, 1970, p. 44.)

10. "Blaming is a type of response to faults in oneself or in others," Robert Adams, "Involuntary Sin," *Philosophical Review*, January 1985, p. 21. Adams does not tell us what kind of response it is. Since he thinks that thinking poorly of someone is a form of unspoken blame (*ibid.*), he must think that thinking poorly of is more than noting a moral fault. I think this is correct.

sensitive not only to the quality of others' wills, but depend as well upon a background of beliefs about the objects of those attitudes. What are those beliefs, and can they be accommodated without appealing to the rival accounts of responsibility that Strawson sets out to avoid?

Strawson says that type-2 pleas inhibit reactive attitudes not by providing an interpretation which shows that the other does not display the pertinent attitudes, but by "inhibiting" the basic demand. It would seem that many of the exemption conditions involve *explanations* of why the individuals display qualities to which the reactive attitudes are otherwise sensitive. So on the face of it, the reactive attitudes are also affected by these explanations. Strawson's essay does not provide an account of how this works or what kinds of explanations exempt.

The problem is not just that the theory is incomplete, but that what might be necessary to complete it will undermine the theory. Strawsonian rivals will rush to fill the gap with their own notions. So it will be said that what makes some of these explanations exempting is that they are deterministic; or it will be said that these conditions are exempting because they indicate conditions in which making the basic demand is inefficacious. To the extent that some such account seems necessary, our enterprise is doomed.

In the following sections, I investigate a Strawsonian alternative. Following Strawson's idea that type-2 pleas inhibit reactive attitudes by inhibiting the basic demand, I propose to construe the exempting conditions as indications of the constraints on intelligible moral demand or, put another way, of the constraints on moral address.

I shall not attempt anything like a comprehensive treatment of the type-2 pleas mentioned by Strawson. I discuss, first and rather briefly, the cases of being a child and being under great strain. I then turn to a more extended discussion of "being unfortunate in formative circumstances," for this looks to be entirely beyond the resources of the expressive theory.

### *Demanding and Understanding*

As Strawson is fully aware, being a child is not simply exempting. Children "are potentially and increasingly capable both of holding, and being objects of, the full range of human and moral attitudes, but are not yet fully capable of either" (pp. 60–61). Children are gradually becoming responsible agents; but in virtue of what are they potentially and increasingly these things? A plausible partial answer to this question is "moral understanding." They do not yet (fully) grasp the moral concepts in such a way that they can (fully) engage in moral communication, and so be unqualified members of the moral community.

The relevance of moral understanding to the expressive theory is this: The negative reactive attitudes express a *moral* demand, a demand for reasonable regard. Now a very young child does not even have a clear sense of the reality of others; but even with this cognitive capacity, children may lack an understanding of the effects of their behavior on others. Even when they understand what it is to hurt another physically, they may lack a sense of what it is to hurt another's feelings, or of the various subtle ways in which that may be done; and even when these things are more or less mastered, they may lack the notion of *reasonable* regard, or of justification. The basic demand is, once more, a moral demand, a demand for reasonable regard, a demand addressed to a moral agent, to one who is capable of understanding the demand. Since the negative reactive attitudes involve this demand, they are not (as fully) appropriately directed to those who do not fully grasp the terms of the demand.

To be intelligible, demanding requires understanding on the part of the object of the demand. The reactive attitudes are incipient forms of communication, which make sense only on the assumption that the other can comprehend the message.

No doubt common views about the moral capacities of children are open to challenge, and the appeal to the notion of understanding itself raises important issues.<sup>11</sup> However, what is important here 'is whether these views can be understood by the Strawsonian theory, and it seems the ordinary view that reactive attitudes make less sense in the case of children is intelligible in Strawsonian terms; this exemption condition reflects constraints arising from the notion of moral demand.

In a certain sense, blaming and praising those with diminished moral understanding loses its "point." This way of putting it smacks of consequentialism, but our discussion suggests a different construction. The reactive attitudes are incipient forms of communication, not in the sense that resentment et al. are usually communicated; very often, in fact, they are not. Rather, the most appropriate and direct expression of resentment is to address the other with a complaint and a demand. Being a child exempts, when it does, not because expressing resentment has no

11. Do *we adults* fully comprehend the notions of justification and reasonable regard? Does understanding presuppose a disputable cognitive view of morality? Certainly conceptions of children are subject to cultural variation. William Blackstone discusses the case of an 8-year old boy who was tried for setting fire to some barns. Because he was found to exhibit "malice, revenge, and cunning, he was found guilty, condemned and hanged accordingly." (In *Commentaries on the Laws of England* (1765–7), as quoted by Jennifer Radden. *Madness and Reason*, George Allen and Unwin, 1985, p. 136.)

It is doubtful that diminished moral understanding is the only relevant factor here. Surely various capacities of concentration and "volitional" control are relevant as well. I do not know how the expressive theory could take these into account.

desirable effects; in fact, it often does. Rather the reactive attitudes lose their point as forms of moral address.<sup>12</sup>

### *Not Being Oneself*

Let's consider whether this kind of explanation can be extended to another of Strawson's type-2 pleas: "being under great strain." Strawson includes this plea in a subgroup of exemptions that include "he wasn't himself" and "he was acting under posthypnotic suggestion." His statement of the rationale in the case of stress is somewhat cryptic:

We shall not feel resentment against the man he is for the action done by the man he is not; or at least we shall feel less. We normally have to deal with him under normal stresses; so we shall not feel towards him, when he acts under abnormal stresses, as we should have felt towards him had he acted as he did under normal stresses. (p. 51)

I take it that what leads Strawson to group these cases together is that in each case the agent, due to special circumstances, acts *uncharacteristically*.

When you learn that someone who has treated you extremely rudely has been under great strain lately, has lost a job, say, or is going through a divorce, you may reinterpret the behavior in such a way that your erstwhile resentment or hurt feelings are inhibited and now seem inappropriate. How does this reinterpretation work? Notice, again, that unlike type-1 pleas, the new interpretation does not contradict the *judgment* that the person treated you rudely; rather, it provides an explanation of the rudeness.

What Strawson says about this case seems plausible. What seems to affect your reactive attitudes is the thought that she's not herself, that the behavior does not reflect or fully reflect the person's moral "personality." The following remark indicates the same phenomena: "He was drunk when he said that; I wouldn't hold it against him." (There is room here for disagreement about the bounds of the moral self. Some parts of folk wisdom have it that one's "true self" is revealed when drunk. To my

12. Reactive attitudes are even more clearly pointless in the case of a radically disintegrated personality, one that has no coherent moral self to be addressed. The case of the sociopath is much more complicated, but arguably something similar may be said here. Those who deal with sociopaths often lose the sense that such characters have a moral self at all; despite appearances, there is "no one home."

For case studies and psychiatric commentary, see Hervey Cleckley, *The Mask of Sanity*, C. V. Mosby, 1941. For philosophical discussion, see Herbert Fingarette, *On Responsibility*, Chap. 2; Vinit Haksar, "The Responsibility of Psychopaths," *The Philosophical Quarterly*, Vol. 15 (1965); M. S. Pritchard, "Responsibility, Understanding, and Psychopathology," *The Monist*, Vol. 58 (1974); Antony Duff, "Psychopathy and Moral Understanding," *American Philosophical Quarterly*, Vol. 14 (1977); and Jeffrie Murphy, "Moral Death: A Kantian Essay on Psychopathy," *Ethics*, Vol. 82 (1972).

knowledge, this has never been claimed about stress.) Again, what is the Strawsonian rationale?

Perhaps this type of case can also be understood in terms of the conditions of intelligible moral address. Insofar as resentment is a form of reproach addressed to an agent, such an attitude loses much of its point here—not, as before, because the other does not fully understand the reproach, but because *he* or *she* (the true self) repudiates such conduct as well. Unlike the case in which the agent acts rudely in the absence of "strain," here the target of your resentment is not one who "really" endorses the behavior you are opposing. You see the behavior as not issuing from that person's moral self, and yet it is the person, qua moral self, that your resentment would address.

The point can be put more generally in this way: Insofar as the negative reactive attitudes express demands (or in some cases appeals) addressed to another moral self, they are conceptually conditioned in various ways. One condition is that, to be fully a moral self, the other must possess sufficient (for what?) moral understanding; another is that the conduct in question be seen as reflecting the moral self. Insofar as the person is subject to great stress, his or her conduct and attitudes fail to meet this latter condition.

I am unsure to what extent these remarks accord with Strawson's own views. They are in any case exceedingly sketchy, and raise problems I am unable to take up here. For one thing, the notion of moral address seems essentially interpersonal, and so would be unavailing in the self-reflexive case. We have negative reactive attitudes toward and make moral demands upon ourselves. To determine whether this is a fatal asymmetry, we would have to investigate the reflexive cases in detail. For another thing, the notion of moral self is certainly not altogether transparent. Why are our responses under stress not reflections of our moral selves—namely, reflections of the moral self under stress? Clearly then, the explanation requires development.

It will be recalled, however, that I am not trying to determine whether a Strawsonian account of the exemption conditions is the *best* account, but to indicate what such an account might be. It will be enough for my purposes here if we can be satisfied that a Strawsonian theory has the resources to provide *some* explanation.

To recapitulate, then, the thesis is this: First, type-2 pleas indicate in different ways limiting conditions on moral address. These are relevant to reactive attitudes because those attitudes are incipiently forms of moral address. This thesis makes sense of Strawson's remark that pleas of this type inhibit reactive attitudes by inhibiting moral demand. Second, given that those conditions are satisfied, type-1 pleas indicate that the basic demand has not been flouted, contrary to appearances (though here again, we must distinguish excuse from justification).

On this account, the practice of holding responsible does indeed seem metaphysically modest, in that it involves no commitments to which issues about determinism are relevant. In a subsequent section I will consider some more bothersome features of our practice; but first I want to call attention to some general issues raised by the account given so far.

### EVIL AND THE LIMITS OF MORAL COMMUNITY

To understand certain exempting and extenuating considerations, I have appealed to the notion of the conditions in which it makes sense morally to address another. I suggested that in different ways these conditions are not (fully) satisfied by the child and the person under severe stress. In the case of children, it seemed plausible to speak of a lack of understanding. What is involved in such understanding is a complex question. Obviously we do not want to make *compliance* with the basic demand a condition of moral understanding. (After all, for the most part, children *do* "comply," but without full understanding.) For the negative reactive attitudes come into play only when the basic demand has been flouted or rejected; and flouting and rejecting, strictly speaking, require understanding.

These remarks raise a very general issue about the limits of responsibility and the limits of evil. It is tempting to think that understanding requires a shared framework of values. At any rate, some of Strawson's remarks hint at such a requirement on moral address. He writes that the reactive attitudes essentially involve regarding the other as "a morally responsible agent, as a term of moral relationships, as a member of the moral community" (p. 59). This last phrase suggests shared ends, at some level, or a shared framework for practical reasoning. Thus, comembers of the moral community are potential interlocutors. In his discussion of Strawson's essay, Lawrence Stern suggests this point:

... when one morally disapproves of another person, it is normal to believe that he is susceptible to the appeal of the principles from the standpoint of which one disapproves. He either shares these principles or can come to share them.<sup>13</sup>

Does morally addressing another make sense unless we suppose that the other can see some reason to take us seriously, to acknowledge our claims? Can we be in a moral community with those who reject the basic terms of moral community? Are the enemies of moral community themselves members? If we suppose that moral address requires moral

13. "Freedom, Blame, and Moral Community," *Journal of Philosophy*, February 14, 1974.

community, then some forms of evil will be exempting conditions. If holding responsible requires the intelligibility of moral address, and if a condition of such address is that the other be seen as a potential moral interlocutor, then the paradox results that extreme evil disqualifies one for blame.

Consider the case of Robert Harris.

On the south tier of Death Row, in a section called "Peckerwood Flats" where the white inmates are housed, there will be a small celebration the day Robert Alton Harris dies.

A group of inmates on the row have pledged several dollars for candy, cookies and soda. At the moment they estimate that Harris has been executed, they will eat, drink and toast to his passing.

"The guy's a misery, a total scumbag; we're going to party when he goes," said Richard (Chic) Mroczko, who lived in the cell next to Harris on San Quentin Prison's Death Row for more than a year. "He doesn't care about life, he doesn't care about others, he doesn't care about himself."

We're not a bunch of Boy Scouts around here, and you might think we're pretty cold-blooded about the whole thing. But then, you just don't know that dude."

San Diego County Assistant Dis. Atty. Richard Huffman, who prosecuted Harris, said, "If a person like Harris can't be executed under California law and federal procedure, then we should be honest and say we're incapable of handling capital punishment."

State Deputy Atty. Gen. Michael D. Wellington asked the court during an appeal hearing for Harris, "If this isn't the kind of defendant that justifies the death penalty, is there ever going to be one?"

What crime did Robert Harris commit to be considered the archetypal candidate for the death penalty? And what kind of man provokes such enmity that even those on Death Row ... call for his execution?

On July 5, 1978, John Mayeski and Michael Baker had just driven through [a] fast-food restaurant and were sitting in the parking lot eating lunch. Mayeski and Baker ... lived on the same street and were best friends. They were on their way to a nearby lake for a day of fishing.

At the other end of the parking lot, Robert Harris, 25, and his brother Daniel, 18, were trying to hotwire a [car] when they spotted the two boys. The Harris brothers were planning to rob a bank that afternoon and did not want to use their own car. When Robert Harris could not start the car, he pointed to the [car] where the 16-year-olds were eating and said to Daniel, "We'll take this one."

He pointed a ... Luger at Mayeski, crawled into the back seat, and told him to drive east ...

Daniel Harris followed in the Harrises' car. When they reached a canyon area ... , Robert Harris told the youths he was going to use



their car in a bank robbery and assured them that they would not be hurt. Robert Harris yelled to Daniel to get the .22 caliber rifle out of the back seat of their car.

"When I caught up," Daniel said in a recent interview, Robert was telling them about the bank robbery we were going to do. He was telling them that he would leave them some money in the car and all, for us using it. Both of them said that they would wait on top of this little hill until we were gone, and then walk into town and report the car stolen. Robert Harris agreed.

"Michael turned and went through some bushes. John said, 'Good luck,' and turned to leave."

As the two boys walked away, Harris slowly raised the Luger and shot Mayeski in the back, Daniel said. Mayeski yelled: "Oh, God," and slumped to the ground. Harris chased Baker down a hill into a little valley and shot him four times.

Mayeski was still alive when Harris climbed back up the hill, Daniel said. Harris walked over to the boy, knelt down, put the Luger to his head and fired.

"God, everything started to spin," Daniel said. "It was like slow motion. I saw the gun, and then his head exploded like a balloon, . . . I just started running and running. . . . But I heard Robert and turned around.

"He was swinging the rifle and pistol in the air and laughing. God, that laugh made blood and bone freeze in me."

Harris drove [the] car to a friend's house where he and Daniel were staying. Harris walked into the house, carrying the weapons and the bag [containing] the remainder of the slain youths' lunch. Then, about 15 minutes after he had killed the two 16-year-old boys, Harris took the food out of the bag . . . began eating a hamburger. He offered his brother an apple turnover, and Daniel became nauseated and ran to the bathroom.

"Robert laughed at me," Daniel said. "He said I was weak; he called me a sissy and said I didn't have the stomach for it."

Harris was in an almost lighthearted mood. He smiled and told Daniel that it would be amusing if the two of them were to pose as police officers and inform the parents that their sons were killed. Then, for the first time, he turned serious. He thought that somebody might have heard the shots and that police could be searching for the bodies. He told Daniel that they should begin cruising the street near the bodies, and possibly kill some police in the area.

[Later, as they prepared to rob the bank,] Harris pulled out the Luger, noticed blood stains and remnants of flesh on the barrel as a result of the point-blank shot, and said, "I really blew that guy's brains out." And then, again, he started laughing.

. . . Harris was given the death penalty. He has refused all requests for interviews since the conviction.

He just doesn't see the point of talking," said a sister, . . . who has visited him three times since he has been on Death Row. "He told me that he had his chance, he took the road to hell and there's nothing more to say."

. . . Few of Harris' friends or family were surprised that he ended up on Death row. He had spent seven of the previous 10 years behind bars. Harris, who has an eighth-grade education, was convicted of car theft at 15 and was sentenced to a federal youth center. After being released, he was arrested twice for torturing animals and was convicted of manslaughter for beating a neighbor to death after a dispute.

Barbara Harris, another sister, talked to her brother at a family picnic on July 4, 1978. He had been out of prison less than six months, and his sister had not seen him in several years.

. . . Barbara Harris noticed his eyes, and she began to shudder. . . . "I thought, 'My God, what have they done to him?' He smiled, but his eyes were so cold, totally flat. It was like looking at a rattlesnake or a cobra ready to strike. They were hooded eyes, with nothing but meanness in them.

"He had the eyes of a killer. I told a friend that I knew someone else would die by his hand."

The next day, Robert Harris killed the two youths. Those familiar with the case were as mystified as they were outraged by Harris' actions. Most found it incomprehensible that a man could be so devoid of compassion and conscience that he could kill two youths, laugh about their deaths and then casually eat their hamburgers. . . .

. . . Harris is a dangerous man on the streets and a dangerous man behind bars, said Mroczko, who spent more than a year in the cell next to Harris' . . .

"You don't want to deal with him out there," said Mroczko, . . . "We don't want to deal with him in here."

During his first year on the row, Mroczko said, Harris was involved in several fights on the yard and was caught trying to supply a prisoner in an adjacent yard with a knife. During one fight, Harris was stabbed and the other prisoner was shot by a guard. He grated on people's nerves and one night he kept the whole cell block awake by banging his shoe on a steel water basin and laughing hysterically.

An encounter with Harris always resulted in a confrontation. If an inmate had cigarettes, or something else Harris wanted, and he did not think "you could hold your mud," Mroczko said, he would try to take them.

Harris was a man who just did not know "when to be cool," he said. He was an obnoxious presence in the yard and in his cell, and his behavior precipitated unwanted attention from the guards. . . .

He acted like a man who did not care about anything. His cell was filthy, Mroczko said, and clothes, trash, tobacco and magazines were scattered on the floor. He wore the same clothes every day and had



little interest in showers. Harris spent his days watching television in his cell, occasionally reading a Western novel.<sup>14</sup>

On the face of it, Harris is an "archetypal candidate" for blame. We respond to his heartlessness and viciousness with moral outrage and loathing. Yet if reactive attitudes were implicitly "invitations to dialogue" (as Stern puts it), then Harris would be an inappropriate object of such attitudes. For he is hardly a potential moral interlocutor, "susceptible to the appeal of the principles from the standpoint of which one disapproves." In this instance, an invitation to dialogue would be met with icy silence (he has "nothing more to say") or murderous contempt.

However, not all communication is dialogue. Harris refuses dialogue, and this refusal is meant to make a point. It is in effect a repudiation of the moral community; he thereby declares himself a moral outlaw. Unlike the small child, or in a different way the psychopath, he exhibits an inversion of moral concern, not a lack of understanding. His ears are not deaf, but his heart is frozen. This characteristic, which makes him utterly unsuitable as a moral interlocutor, intensifies rather than inhibits the reactive attitudes. Harris's form of evil *consists* in part in being beyond the boundaries of moral community. Hence, if we are to appeal to the constraints on moral address to explain certain type-2 pleas, we must not include among these constraints comembership in the moral community or the significant possibility of dialogue—unless, that is, evil is to be its own exemption. At these outer limits, our reactive attitudes can be nothing more (or less) than a denunciation forlorn of the hope of an adequate reply.

### *The Roots of Evil*

I said that Harris is an archetypal candidate for blame—so, at least, we react to him. Does it matter to our reactions how he came to be so? Strawson thinks so, for, among type-2 pleas, he includes "being unfortunate in formative circumstances." We must now investigate the relevance of such historical considerations to the reactive attitudes. As it happens, the case of Robert Harris is again a vivid illustration.

[During the interview] Barbara Harris put her palms over her eyes and said softly, "I saw every grain of sweetness, pity and goodness in him destroyed. . . . It was a long and ugly journey before he reached that point."

14. From Miles Corwin, "Icy Killer's Life Steeped in Violence," *Los Angeles Times*, May 16, 1982. Copyright, 1982, *Los Angeles Times*. Reprinted by permission. For the length of this and the next quotation, I ask for the reader's patience. It is very important here to work with realistic and detailed examples.

Robert Harris' 29 years . . . have been dominated by incessant cruelty and profound suffering that he has both experienced and provoked. Violence presaged his birth, and a violent act is expected to end his life.

Harris was born Jan. 15, 1953, several hours after his mother was kicked in the stomach. She was 6½ months pregnant and her husband, an insanely jealous man, . . . came home drunk and accused her of infidelity. He claimed that the child was not his, threw her down and kicked her. She began hemorrhaging, and he took her to the hospital.

Robert was born that night. His heartbeat stopped at one point . . . but labor was induced and he was saved. Because of the premature birth, he was a tiny baby; he was kept alive in an incubator and spent months at the hospital.

His father was an alcoholic who was twice convicted of sexually molesting his daughters. He frequently beat his children . . . and often caused serious injury. Their mother also became an alcoholic and was arrested several times, once for bank robbery.

All of the children had monstrous childhoods. But even in the Harris family, . . . the abuse Robert was subjected to was unusual.

Before their mother died last year, Barbara Harris said, she talked incessantly about Robert's early years. She felt guilty that she was never able to love him; she felt partly responsible that he ended up on Death Row.

When Robert's father visited his wife in the hospital and saw his son for the first time, . . . the first thing he said was, "Who is the father of that bastard?" When his mother picked him up from the hospital . . . she said it was like taking a stranger's baby home.

The pain and permanent injury Robert's mother suffered as a result of the birth, . . . and the constant abuse she was subjected to by her husband, turned her against her son. Money was tight, she was overworked and he was her fifth child in just a few years. She began to blame all of her problems on Robert, and she grew to hate the child.

"I remember one time we were in the car and Mother was in the back seat with Robbie in her arms. He was crying and my father threw a glass bottle at him, but it hit my mother in the face. The glass shattered and Robbie started screaming. I'll never forget it," she said. . . .

"Her face was all pink, from the mixture of blood and milk. She ended up blaming Robbie for all the hurt, all the things like that. She felt helpless and he was someone to vent her anger on."

. . . Harris had a learning disability and a speech problem, but there was no money for therapy. When he was at school he felt stupid and classmates teased him, his sister said, and when he was at home he was abused.

"He was the most beautiful of all my mother's children; he was an angel," she said. "He would just break your heart. He wanted love so bad he would beg for any kind of physical contact."

"He'd come up to my mother and just try to rub his little hands on her leg or her arm. He just never got touched at all. She'd just push him away or kick him. One time she bloodied his nose when he was trying to get close to her."

Barbara Harris put her head in her hands and cried softly. "One killer out of nine kids. . . . The sad thing is he was the most sensitive of all of us. When he was 10 and we all saw 'Bambi,' he cried and cried when Bambi's mother was shot. Everything was pretty to him as a child; he loved animals. But all that changed; it all changed so much."

. . . All nine children are psychologically crippled as a result of their father, she said, but most have been able to lead useful lives. But Robert was too young, and the abuse lasted too long, she said, for him ever to have had a chance to recover.

[At age 14] Harris was sentenced to a federal youth detention center [for car theft]. He was one of the youngest inmates there, Barbara Harris said, and he grew up "hard and fast."

. . . Harris was raped several times, his sister said, and he slashed his wrists twice in suicide attempts. He spent more than four years behind bars as a result of an escape, an attempted escape and a parole violation.

The centers were "gladiator schools," Barbara Harris said, and Harris learned to fight and to be mean. By the time he was released from federal prison at 19, all his problems were accentuated. Everyone in the family knew that he needed psychiatric help.

The child who had cried at the movies when Bambi's mother dies had evolved into a man who was arrested several times for abusing animals. He killed cats and dogs, Daniel said, and laughed while torturing them with mop handles, darts and pellet guns. Once he stabbed a prize pig more than 1,000 times.

"The only way he could vent his feelings was to break or kill something," Barbara Harris said. "He took out all the frustrations of his life on animals. He had no feeling for life, no sense of remorse. He reached the point where there wasn't that much left of him."

. . . Harris' family is ambivalent about his death sentence. [Another sister said that] if she did not know her brother's past so intimately, she would support his execution without hesitation. Barbara has a 16-year-old son; she often imagines the horror of the slain boy's parents.

"If anyone killed my son, I'd try my damndest, no matter what it took, to have my child revenged," Barbara Harris said. "I know how those parents must suffer every day."

"But Robbie in the gas chamber. . . ." She broke off in mid-sentence and stared out a window. "Well, I still remember the little boy who used to beg for love, for just one pat or word of kindness. . . . No I can't say I want my brother to die."

. . . Since Harris has been on Death Row, he has made no demands of time or money on his family. Harris has made only one request; he

wants a dignified and serene ceremony after he dies—a ceremony in marked contrast to his life.

He has asked his oldest brother to take his ashes, to drive to the Sierra, hike to a secluded spot and scatter his remains in the trees.<sup>15</sup>

No doubt this history gives pause to the reactive attitudes. Why does it do so? "No wonder Harris is as he is!" we think. What is the relevance of this thought?

Note, to begin with, that the story in no way undermines the judgments that he is brutal, vicious, heartless, mean.<sup>16</sup> Rather, it provides a kind of explanation for his being so. Can the expressive theory explain why the reactive attitudes should be sensitive to such an explanation?

Strawson's general rubric for type-2 pleas (or the subgroup in which this plea is classified) is "being incapacitated for ordinary interpersonal relationships." Does Harris have some independently identifiable incapacity for which his biography provides evidence? Apparently, he is incapacitated for such relationships—for example, for friendship, for sympathy, for being affected by moral considerations. To be homicidally hateful and callous in Harris's way is to lack moral concern, and to lack moral concern is to be incapacitated for moral community. However, to exempt Harris on these grounds is problematic. For then everyone who is evil in Harris's way will be exempt, independently of facts about their background. But we had ample evidence about *this* incapacity before we learned of his childhood misfortunes, and that did not affect the reactive attitudes. Those misfortunes affect our responses in a special and nonevidential way. The question is why this should be so.

This would seem to be a hard question for compatibilist views generally. What matters is whether, in one version, the practice of holding responsible can be efficacious as a means of social regulation, or whether, using the expressive theory, the conditions of moral address are met. These questions would seem to be settled by how individuals *are*, not by how they came to be. Facts about background would be, at most, evidence that some other plea is satisfied. In themselves, they would not seem to matter.

A plea of this kind is, on the other hand, grist for the incompatibilists' mill. For they will insist on an essential historical dimension to the concept of responsibility. Harris's history reveals him to be an inevitable product of his formative circumstances. And seeing him as a product is inconsistent with seeing him as a responsible agent. If his cruel atti-

15. Miles Corwin, op. cit. Copyright, 1982, *Los Angeles Times*. Reprinted by permission.

16. Although significantly, when his past is in focus, we are less inclined to use certain reactive epithets, such as "scumbag." This term is used to express an attitude about the appropriate treatment of the individual (that he is to be thrown in the garbage, flushed down the toilet, etc.). Some other reactive terms are "jerk," "creep," "son of a bitch."

tudes and conduct are the inevitable result of his circumstances, then he is not responsible for them, unless he was responsible for those circumstances. It is this principle that gives the historical dimension of responsibility and of course entails the incompatibility of determinism and responsibility.

In this instance, however, an incompatibilist diagnosis seems doubtful. In the first place, our response to the case is not the simple suspension of reactive attitudes that this diagnosis would lead one to expect, but ambivalence. In the second place, the force of the example does not depend on a belief in the *inevitability* of the upshot. Nothing in the story supports such a belief. The thought is not "It had to be!" but, again, "No wonder!"

### SYMPATHY AND ANTIPATHY

How and why, then, does this larger view of Harris's life in fact affect us? It is too simple to say that it leads us to suspend our reactive attitudes. Our response is too complicated and conflicted for that. What appears to happen is that we are unable to command an overall view of his life that permits the reactive attitudes to be sustained without ambivalence. That is because the biography forces us to see him as a *victim*, and so seeing him does not sit well with the reactive attitudes that are so strongly elicited by Harris's character and conduct. Seeing him as a victim does not totally dispel those attitudes. Rather, in light of the "whole" story, conflicting responses are evoked. The sympathy toward the boy he was is at odds with outrage toward the man he is. These responses conflict not in the way that fear dispels anger, but in the way that sympathy is opposed to antipathy. In fact, each of these responses is appropriate, but taken together they do not enable us to respond overall in a coherent way.

Harris both satisfies and violates the criteria of victimhood. His childhood abuse was a misfortune inflicted upon him against his will. But at the same time (and this is part of his very misfortune) he unambivalently endorses suffering, death, and destruction, and that is what (one form of) evil is. With this in focus, we see him as a victimizer and respond to him accordingly. The ambivalence results from the fact that an overall view simultaneously demands and precludes regarding him as a victim.

What we have here is not exactly a clash between what Thomas Nagel has called the objective and subjective standpoints.<sup>17</sup> It is not that from the more comprehensive viewpoint that reveals Harris as a victim, his responsibility is indiscernible. Rather, the clash occurs within a single point of view that reveals Harris as evil (and hence calling for enmity and moral opposition) and as one who is a victim (calling for sympathy

17. In *The View from Nowhere*, Oxford University Press, 1985.

and understanding). Harris's misfortune is such that scarcely a vestige remains of his earlier sensibilities. Hence, unless one knew Harris as a child or keeps his earlier self vividly in mind, sympathy can scarcely find a purchase.

### Moral Luck and Moral Equality

However, what is arresting about the Harris case is not just the clash between sympathy and antipathy. The case is troubling in a more personal way. The fact that Harris's cruelty is an intelligible response to his circumstances gives a foothold not only for sympathy, but for the thought that if I had been subjected to such circumstances, I might well have become as vile. What is unsettling is the thought that one's moral self is such a fragile thing. One tends to think of one's moral sensibilities as going deeper than that (though it is not clear what this means). This thought induces not only an ontological shudder, but a sense of equality with the other: I too am a potential sinner.<sup>18</sup>

This point is merely the obverse of the point about sympathy. Whereas the point about sympathy focuses on our empathetic response to the other, the thought about moral luck turns one's gaze inward. It makes one feel less in a position to cast blame. The fact that my potential for evil has not been nearly so fully actualized is, for all I know, something for which I cannot take credit. The awareness that, in this respect, the others are or may be like oneself clashes with the distancing effect of enmity.

Admittedly, it is hard to know what to do with this conclusion. Equality of moral potential does not, of course, mean that Harris is not actually a vile man; on the contrary, it means that in similar circumstances I would have become vile as well. Since he is an evil man, we cannot and should not treat him as we would a rabid dog. The awareness of moral luck, however, taints one's own view of one's moral self as an achievement, and infuses one's reactive attitudes with a sense of irony. Only those who have survived circumstances such as those that ravaged Harris are in a good position to know what they would have done. We lucky ones can only wonder. As a product of reflection, this attitude is, of course, easily lost when the knife is at one's own throat.

### DETERMINISM AND IGNORANCE

Nothing in the foregoing reflections is necessarily inconsistent with the expressive theory. The ways in which reactive attitudes are affected by

18. In "Determinism and Moral Perspectives," *Philosophy and Phenomenological Research*, September 1960, Elizabeth Beardsley calls attention to the perspective evoked by such cases as Harris, though she links this perspective too closely, in my opinion, to the notion of determinism.

sympathy and moral luck are intelligible without appealing to any of the conceptions of responsibility that Strawson eschews. Nevertheless, our attitudes remain puzzling in a number of respects.

Earlier we questioned an incompatibilist diagnosis of our example on the grounds that the historical explanation need not be construed as deterministic. Horrid backgrounds do not inevitably give rise to horrid people. Some manage somehow to survive a similar magnitude of misfortune, if not unscathed, at least as minimally decent human beings. Conversely, people are sometimes malicious despite a benign upbringing. What do we suppose makes the difference?

Strictly speaking, no one who is vicious in *just* the way we have interpreted Harris to be could fail to have had an abusive childhood. For our interpretation of who Harris is depends upon his biography, upon our interpretation of his life. Harris's cruelty is a response to the shattering abuse he suffered during the process of socialization. The objects of his hatred were not just the boys he so exultantly murdered, but the "moral order" that mauled and rejected him. (It is significant that Harris wanted to go out and kill some cops after the murder; he wanted not just to reject authority, but to confront it.) He defies the demand for human consideration because he had been denied this consideration himself. The mistreatment he received becomes a ground as well as a cause of the mistreatment he gives. It becomes part of the content of his "project."

Thus, someone who had a supportive and loving environment as a child, but who was devoted to dominating others, who killed for enjoyment, would not be vicious in the way Harris is, since he or she could not be seen as striking back at "society"; but such a person could be just as vicious. In common parlance, we sometimes call such people "bad apples," a phrase that marks a blank in our understanding. In contrast to Harris, whose malice is motivated, the conduct of "bad apples" seems inexplicable. So far, we cannot see them as victims, and there is no application for thoughts about sympathy and moral luck.

However, do we not suppose that *something* must have gone wrong in the developmental histories of these individuals, if not in their socialization, then "in them"—in their genes or brains? (Suppose a certain kind of tumor is such that its onset at an early age is known to be strongly correlated with the development of a malicious character. This supposition is no doubt bad science fiction; that a complex and articulated psychological structure could be caused by gross brain defect seems antecedently implausible.) Whatever "nonenvironmental" factors make the difference, will they not play the same role as Harris's bad upbringing—that is, will they not have victimized these individuals so that thoughts about sympathy and moral luck come into play? Or can evil be the object of unequivocal reactive attitudes only when it is inexplicable?

If determinism is true, then evil is a joint product of nature and nurture. If so, the difference between any evil person and oneself would seem to be a matter of moral luck. For determinism seems to entail that if one had been subjected to the internal and external conditions of some evil person, then one would have been evil as well. If that is so, then the reflections about moral luck seem to entail that the acceptance of determinism should affect our reactive attitudes in the same way as they are affected in Harris's case. In the account we have suggested, then, determinism seems to be relevant to reactive attitudes after all.

Actually, this conclusion does not follow without special metaphysical assumptions. For the counterfactuals that underlie thoughts about moral luck must be constrained by the conditions of personal identity. It may be that no one who had been exposed to just the internal and external conditions of some given individual could have been me. To make sense of a counterfactual of the form, "If *i* had been in *C*, then *i* would have become a person of type *t*," *C* must be supposed to be compatible with *i*'s existence as an individual (*i* must exist in the possible world in which *C* obtains). For example, it is widely held that genetic origin is essential to an individual's identity. In that case, the counterfactual, "If I had had Harris's genetic origin and his upbringing, then I would have been as evil as he," will not make sense. Now it might be that Harris's genetic origins are among the determinants of his moral development. Thus, even if this is a deterministic world, there may be no true counterfactual that would support the thought that the difference between Harris and me is a matter of moral luck. There is room for the thought that there is something "in me" by virtue of which I would not have become a vicious person in Harris's circumstances. And if that factor were among my essential properties, so to speak, then that difference between Harris and me would not be a matter of moral luck on my part, but a matter of who we essentially were. That would not, of course, mean that I was essentially good or Harris essentially evil, but that I would not have been corrupted by the same circumstances as those that defeated Harris. To be sure, to suppose that this difference is in itself to my moral credit would be odd. To congratulate me on these grounds would be to congratulate me on being myself. Nevertheless, this difference still might explain what is to my credit, such moral virtues as I may possess. This will seem paradoxical only if we suppose that whatever is a ground of my moral credit must itself be to my credit. But I see no compelling reason to suppose this.

### Historical Responsibility

Libertarians believe that evil is the product neither of nature nor of nurture, but of free will. Do we understand what this might mean?

It is noteworthy that libertarians will be able to agree with much of what we have said about moral luck. Harris's history affects us because it makes us wonder how *we* would have responded, and thus shakes our confidence that we would have avoided a pernicious path in those circumstances. But this effect is perfectly compatible with Harris's responsibility for how he did respond, just as we would have been responsible for how we would have responded. The biography affects us not because it is deterministic, libertarians can say, but because it shakes our confidence that we would have exercised that freedom rightly in more dire straits. We are not, of course, responsible for our formative circumstances—and in this respect we are morally lucky and Harris is unlucky—but those circumstances do not determine our responses to them. It is the individual's own response that distinguishes those who become evil from those who do not.

This idea is nicely captured by Peter Abélard: "Nothing mars the soul except what is of its own nature, namely consent."<sup>19</sup> The idea is that one cannot simply be caused to be morally bad by the environment. So either Harris's soul is not (morally) marred, or he has been a willing accomplice to the malformation of the self. His evil means that he has consented to what he has become—namely, one who consents to cruelty. Thus, Abélardians try to fill the statistical cracks with the will. The development of the moral self, they will say, is mediated by consent.

We should be struck here by the *a priori* character of libertarian convictions. How is Harris's consent to be construed, and why *must* it have occurred? What evidence is there that it occurred? Why couldn't Harris just have become that way? What is the difference between his having acquiesced to what he became and his simply having become that way? The libertarian faces the following difficulty: If there is no such difference, then the view is vacuous, for consent was supposed to explain his becoming that way. If there is a difference, what evidence is there that it obtains in a particular case? Isn't there room for considerable doubt about this, and shouldn't libertarians, or we, insofar as we are libertarians, be very doubtful about Harris's responsibility—and indeed, on the Abélardian thesis, even about whether Harris is an evil man, whether his soul is morally marred? (Notice that the tumor case is *a priori* impossible on that thesis, unless we think of the tumor somehow as merely presenting an occasion for consent—as inclining without necessitating.) One suspects that the libertarian confidence in their attributions of historical responsibility is rooted in a picture according to which the fact that Harris became that way *proves* that he consented. Then, of course, the appeal to consent is explanatorily vacuous.

Epistemology apart, the attempt to trace the evil self to consent at an earlier stage is faced with familiar difficulties. If we suppose (fancifully)

19. From "Intention and Sin," reprinted in Herbert Morris (ed.), *Freedom and Responsibility* (Stanford University Press), p. 169.

that Harris, earlier on, with full knowledge and deliberation, launched himself on his iniquitous career,<sup>20</sup> we would be merely postponing the inquiry, for the will which could fully and deliberately consent to such a career would have to have its roots in a self which is already morally marred—a self, therefore, which cannot itself be seen simply as a product of consent. Are we instead to suppose that at some earlier stage Harris slipped heedlessly or recklessly into patterns of thought and action which he ought to have known would eventuate in an evil character? (This seems to have been Aristotle's view in *Nicomachean Ethics*, Book III.5.) In that case, we would be tracing his present ways to the much less egregious faults of negligence.<sup>21</sup>

#### RESPONSIBILITY FOR THE SELF

Strawson and others often charge libertarians with a metaphysically dubious conception of the self. The foregoing reflections indicate a basis for this charge. Libertarianism combines the Abélardian view about consent (or something like it) with the principle (or something like it) that to be responsible for anything, one must be responsible for (some of) what produces it. If we think of agents as consenting to this or that *because* they are (or have?) selves of a certain character, then it looks as though they are responsible for so consenting only if they are responsible for the self in which that consent is rooted. To establish this in each case, we have to trace the character of the self to earlier acts of consent. This enterprise seems hopeless, since the trace continues interminably or leads to a self to which the individual did not consent. The libertarian seems committed, then, to bearing the unbearable burden of showing how we can be responsible for ourselves. This burden can seem bearable only in a view of the self as an entity that mysteriously both transcends and intervenes in the "causal nexus," because it is both product and author of its actions and attitudes.

Must libertarians try to bear this burden? Perhaps the idea that they must rests upon a view of the self to which libertarians need not be committed. Perhaps the trouble arises in the first place from viewing the self as a thing standing in causal relation to acts of consent. The libertarian might say that to talk about the (moral) self is not to talk about an entity which necessitates specific acts of consent, but to talk about the sorts of things to which an individual tends to consent. To speak of Harris's moral self is not to explain his conduct, but to indicate the way he is morally. What we are responsible for are the particular things we consent to. We need not consider whether we are responsible for the genesis of the

20. If such a thing ever occurred, it must have occurred at a stage when Harris clearly would have fallen under the exemption condition of "being a child."

21. Adams makes this point; *op. cit.*

entity whose characteristics necessitate those acts of consent, for there is no such entity. In a way, of course, one is derivatively responsible for one's self, since one's moral self is constituted by the character of what one consents to, and one is responsible for what one consents to.<sup>22</sup>

The historical dimension of the concept of responsibility results from the principle that one is not responsible for one's conduct if that is necessitated by causes for which one is not responsible. This leads to a problematic requirement that one be responsible for one's self only if one thinks of the self as an entity that causes one's (its) actions and willings. Libertarians can reject this view. What they must affirm is that we are responsible for what we consent to, that consent is not necessitated by causes internal or external to the agent, and that if it were, we could not properly hold the individual responsible for what he or she consents to. These claims are far from self-evident. But they hardly amount to a "panicky metaphysics." (p. 66).<sup>23</sup>

In the end, however, I do not think that libertarianism can be so readily domesticated. The idea that one is responsible for and only for what one consents to is not of course distinctive of libertarianism; that idea has no historical implications. What is distinctive is the further requirement that consent be undetermined. I do not think the idea that consent is undetermined is in itself particularly problematic. The trouble begins only when we ask why this is *required*. The ground of this requirement is the intuition that unless consent were undetermined, we would not truly be *originators* of our deeds. We would be merely products, and not, as it were, producers. It is this intuition to which the libertarian finds it so difficult to give content. "Being an originator" does not mean just "consenting to," for that is already covered by the first thesis. Nor is this notion captured simply by adding the requirement of indeterminism; that is a merely negative condition. Attempts to specify the condition in positive terms either cite something that could obtain in a deterministic world, or something obscurely transcendent.

I suspect, then, that any metaphysically innocuous version of libertarianism must leave its incompatibilist component unmotivated.

## IGNORANCE AND SKEPTICISM

I have been exploring some ways in which the expressive theory might explain the relevance of certain historical considerations. Whatever the

22. It is noteworthy that Harris himself seems to accept responsibility for his life: "He told me he had his chance, he took the road to hell and there's nothing more to say." (From the end of the first extract from the Corwin article.)

23. For an attempt at libertarianism without metaphysics, see David Wiggins, "Towards a Credible Form of Libertarianism," in T. Honderich (ed.), *Essays on Freedom of Action*, Routledge and Kegan Paul, 1973.

best explanation may be, the remarkable fact is that we are, for the most part, quite ignorant of these considerations. Why does our ignorance not give us more pause? If, for whatever reason, reactive attitudes are sensitive to historical considerations, as Strawson acknowledges, and we are largely ignorant of these matters, then it would seem that most of our reactive attitudes are hasty, perhaps even benighted, as skeptics have long maintained. In this respect, our ordinary practices are not as unproblematic as Strawson supposes.

It might be thought that these suspicions about reactive attitudes have no bearing on responsibility, but with the expressive theory, that cannot be easily maintained. As we normally think of the matter, not all considerations that affect reactive attitudes are strictly relevant to responsibility. For example, if one shares a moral fault with another, one may feel it inappropriate to blame the other. Here the point is not that the other is not responsible or blameworthy, but that it is not *one's* business to blame. One should tend to one's own faults first.<sup>24</sup> Thoughts about moral luck seem to be continuous with this ordinary phenomenon. The thought is not that the other is not blameworthy, but that one may be no better, and that indignation on one's part would be self-righteous and indulgent. By calling our attention to our general ignorance of historical considerations, the skepticism we have just been considering is merely an extension of these reflections.

With an expressive theory, however, it is not clear that a general skepticism about the propriety of the reactive attitudes can be separated from skepticism about responsibility. For the latter concept is the concept of the conditions in which it is appropriate to respond to one another in reactive ways. In a Strawsonian view, there is no reason for a wedge between the practices that evince the reactive attitudes and the belief in responsibility. In a particular case, one may believe another to be responsible without actually responding to him or her in reactive ways (due to strains of commitment and so on), because one may regard the other as blameworthy, as an appropriate object of the reactive attitudes by others in the moral community. But if one thinks that *none* of us mortals is in a position to blame, then it is doubtful that any sense can be given to the belief that the other is nonetheless blameworthy. One can still attribute cowardice, thoughtlessness, cruelty, and so on, to others; but as we have seen, these judgments are not sufficient in a Strawsonian view to characterize the practice of holding responsible. We might try to appeal to the reactive attitudes of a select group of actual or hypothetical judges (here is another job

24. Montaigne would not agree: "To censure my own faults in some other person seems to me no more incongruous than to censure, as I often do, another's in myself. They must be denounced everywhere, and be allowed no place of sanctuary." ("On the Education of Children," in *Essays*, Penguin Classics, 1971, p. 51.)

for God to do),<sup>25</sup> but then the connection to reactive attitudes becomes so tenuous or hypothetical that the attitudes lose the central role they are given in "Freedom and Resentment," and the expressive theory loses its distinctive character. It then collapses into the view discussed in the section called "Blaming and Finding Fault."

## OBJECTIVITY AND ISOLATION

It remains unclear to what extent our ordinary practices involve dubious beliefs about ourselves and our histories. To acknowledge the relevance of historical considerations is, on any account, to acknowledge a potential source of skepticism about those practices; moreover, in a Strawsonian account (though not in a libertarian account), such skepticism cannot be readily separated from skepticism about responsibility itself. In this respect, Strawson is inordinately optimistic about our common ways.

However, these practices are vulnerable to a different kind of suspicion. This suspicion is related to Strawson's conception of the place of "retributive" sentiments in those practices, and to his claim that that practice, so conceived, is not something that is optional and open to radical criticism, but rather is part of the "framework" of our conception of human society. One could agree that the expressive theory best gives the basis and content of the practice of holding responsible and still maintain that abandoning this practice is not only conceivable but desirable, for what it expresses is itself destructive of human community. I conclude with some comments on this further issue.

Consider some remarks by Albert Einstein:

I do not at all believe in human freedom in the philosophical sense. Everybody acts not only under external compulsion but also in accordance with inner necessity. Schopenhauer's saying, "A man can do what he wants, but not want what he wants," has been a very real inspiration to me since my youth; it has been a continual consolation in the face of life's hardships, my own and others', and an unfailing well-spring of tolerance. This realization mercifully mitigates the easily paralyzing sense of responsibility and prevents us from taking ourselves and other people all too seriously; it is conducive to a view of life which, in particular, gives humor its due.<sup>26</sup>

Significantly, in the same place Einstein speaks of himself as a "lone traveler," with a "pronounced lack of need for direct contact with other human beings and human communities," who has

25. Just as Berkeley tried to save the thesis that material objects consist in ideas.

26. Albert Einstein, *Ideas and Opinions*, Crown Publishers, 1982, pp. 8–9.

never belonged to my country, my home, my friends, or even my immediate family, with my whole heart; in the face of all these ties, I have never lost a sense of distance and a need for solitude—feelings which increase with the years.

The point that interests me here is not that these remarks confute Strawson's claim that reactive attitudes are never in practice affected by an acceptance of determinism, but that they corroborate his central claim about the alternative to the reactive, participant stance. The "distance" of which Einstein speaks is just an aspect of the "detachment" Strawson thinks characterizes the objective stance. At its extremes, it takes the form of human isolation. What is absent from Einstein's outlook is something that, I suspect, Strawson cherishes: the attachment or commitment to the personal, as it might be called.<sup>27</sup>

Whatever its grounds, Einstein's outlook is not without its appeal. Perhaps part of its appeal can be attributed to a fear of the personal, but it is also appealing precisely on account of its repudiation of the retributive sentiments. In another place, Einstein salutes the person "to whom aggressiveness and resentment are alien."<sup>28</sup> Can such an ideal of the person be pursued only at the cost of the attachment to the personal? Must we choose between isolation and animosity?

Some of Strawson's remarks imply that we must:

Indignation, disapprobation, like resentment, tend to inhibit or at least to limit our goodwill towards the object of these attitudes, tend to promote at least partial and temporary withdrawal of goodwill . . . (These, of course, are not contingent connections.) But these attitudes . . . are precisely the correlates of the moral demand in the case where the demand is felt to be disregarded. The making of the demand is the proneness to such attitudes. The holding of them does not . . . involve . . . viewing their object other than as a member of the moral community. The partial withdrawal of goodwill which these attitudes entail, the modification *they* entail of the general demand that another should if possible be spared suffering, is . . . the consequence of *continuing* to view him as a member of the moral community; only as one who has offended against its demands. So the preparedness to acquiesce in that infliction of suffering on the offender which is an es-

27. To what extent Einstein lived up to this outlook, I am not prepared to say. Some other writings suggest a different view: "External compulsion can . . . reduce but never cancel the responsibility of the individual. In the Nuremberg trials, this idea was considered to be self-evident. . . . Institutions are in a moral sense impotent unless they are supported by the sense of responsibility of living individuals. An effort to arouse and strengthen this sense of responsibility of the individual is an important service to mankind" (op. cit., p. 27). Is Einstein taking a consequentialist stance here?

28. Ibid.



sential part of punishment is all of a piece with this whole range of attitudes. . . . (pp. 62–63)

This passage is troubling. Some have aspired to rid themselves of the readiness to limit goodwill and to acquiesce in the suffering of others not in order to relieve the strains of involvement, nor out of a conviction in determinism, but out of a certain ideal of human relationships, which they see as poisoned by the retributive sentiments. It is an ideal of human fellowship or love which embodies values that are arguably as historically important to our civilization as the notion of moral responsibility itself. The question here is not whether this aspiration is finally commendable, but whether it is compatible with holding one another morally responsible. The passage implies that it is not.

If holding one another responsible involves making the moral demand, and if the making of the demand is the proneness to such attitudes, and if such attitudes involve retributive sentiments and hence<sup>29</sup> a limitation of goodwill, then skepticism about retribution is skepticism about responsibility, and holding one another responsible is at odds with one historically important ideal of love.

Many who have this idea, such as Gandhi or King,<sup>30</sup> do not seem to adopt an objective attitude in Strawson's sense. Unlike Einstein's, their lives do not seem characterized by human isolation: They are often intensely involved in the "fray" of interpersonal relations. Nor does it seem plausible to suppose that they do not hold themselves and others morally responsible: They *stand up* for themselves and others against their oppressors; they *confront* their oppressors with the fact of their misconduct, *urging* and even *demanding* consideration for themselves and others; but they manage, or come much closer than others to managing, to do such things without vindictiveness or malice.

Hence, Strawson's claims about the interpenetration of responsibility and the retributive sentiments must not be confused with the expressive theory itself. As these lives suggest, the retributive sentiments can in principle be stripped away from holding responsible and the demands and appeals in which this consists. What is left are various forms of reaction and appeal to others as moral agents. The boundaries of moral responsibility are the boundaries of intelligible moral address. To regard another as morally responsible is to react to him or her as a moral self.<sup>31</sup>

29. Rather than attempting to separate retribution from responsibility, one might try to harmonize retribution and goodwill. This possibility seems to me worth exploring.

30. For these examples, and the discussion in this section, I am indebted to Stern (op. cit.).

31. We have, of course, seen reasons why these boundaries require further delineation.

## Part II

### HIERARCHY, RATIONALITY, AND THE "REAL SELF"