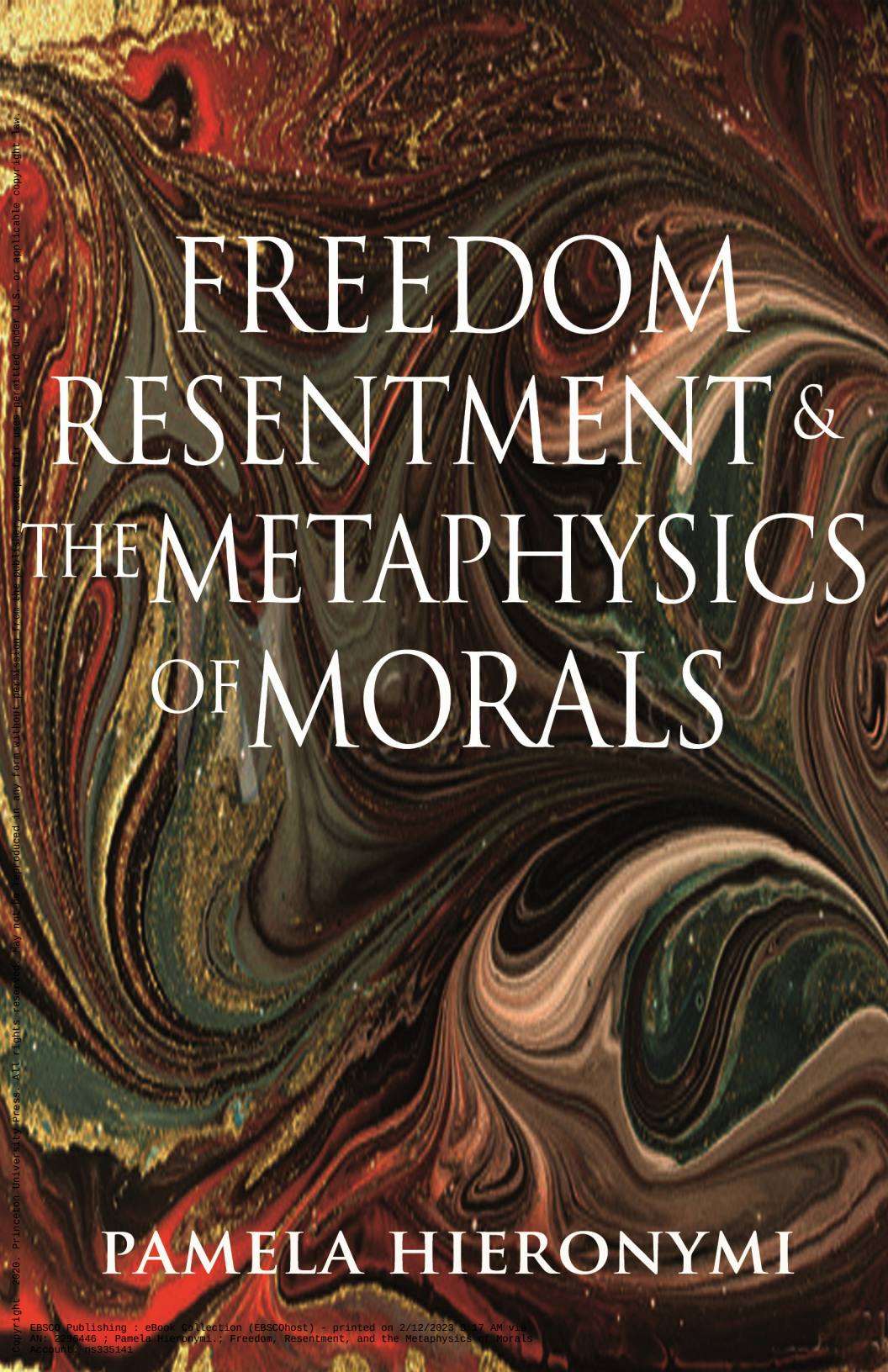

FREEDOM, RESENTMENT, AND THE METAPHYSICS OF MORALS

Freedom, Resentment, and the Metaphysics of Morals

PAMELA HIERONYMI

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Published by Princeton University Press
41 William Street, Princeton, New Jersey 08540
6 Oxford Street, Woodstock, Oxfordshire OX20 1TR

press.princeton.edu

ISBN 978-0-691-19403-5
ISBN (e-book) 978-0-691-20097-2

British Library Cataloging-in-Publication Data is available

Editorial: Matt Rohal
Production Editorial: Jenny Wolkowicki
Jacket design: Lorraine Doneker
Production: Erin Suydam
Publicity: Alyssa Sanford and Katie Lewis
Copyeditor: Joseph Dahm

This book has been composed in Arno

Printed on acid-free paper. ∞

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To David Kaplan, Herbert Morris,
and, of course, Harry Frankfurt

CONTENTS

PREFACE

LUDWIG WITTGENSTEIN reportedly once suggested that philosophers greet one another with "Take your time." In contemporary academia, we might do better with "Give one another time." Or, though it is less pithy, "Work to undermine the bureaucratic and professional pressures that make it inadvisable to take your time."

I have been unusually fortunate to have had time for this project.

I vividly remember sitting in my sister's bedroom while home from school, sometime in the 1990s, reading Peter Strawson's "Freedom and Resentment." I was ready to raise exactly the objection that Strawson claimed, with unexpected impoliteness, would be raised only by someone who "utterly failed to grasp the purport of the preceding answer." I stood guilty as charged. With work, and given some familiarity with Strawson's philosophical milieu, I eventually arrived at what I here call the "broadly Wittgensteinian" interpretation of Strawson's text. I then taught that interpretation to undergraduates almost every year for over a decade.

Still, I was not fully satisfied that I could follow Strawson's thought from sentence to sentence on what were clearly the crucial two pages, and this bothered me. For example, on the first of these pages, Strawson gives an exceedingly quick dismissal of the possible relevance of determinism. He then immediately admits that the dismissal is "too facile"—but, he says, only "in a sense." Why present a too-facile dismissal? And how is it facile only "in a sense"? He next attempts to reassure the reader, saying that "whatever is too quickly dismissed in this dismissal is allowed for in the only possible form of affirmative answer that remains." What is the possible form of affirmative answer? Two sentences later, he says, "And our question reduces to this . . ."—but what follows seemed to me simply a *restatement*, not a

reduction, of the question at hand. The next page brought further mysteries. Then, when Strawson finally raises the objection that he claims has "utterly failed to grasp the purport" of his answer, he poses it in a way that it should not be posed, on the broadly Wittgensteinian interpretation: he has his objector ask not whether it would be *moral* or *just* to continue holding one another responsible if determinism is true, but instead whether it would be *rational* to do so. The pieces did not quite fit.

In 2011–12, with the benefit of a research leave, I began drafting my own manuscript about free will and moral responsibility (with which I am still taking my time). After turning out a couple of chapters, I thought I ought to take the time I had been given to sort out how, exactly, Strawson moves from sentence to sentence across those two crucial pages. I did almost nothing else for the remainder of my leave, and then for a few more years. I was provided crucial guidance and hints from people who knew Strawson's broader corpus, and I was helped by a conference in honor of the fiftieth anniversary of the publication of "Freedom and Resentment." I also had the incredible privilege—and great fun—of teaching some early and midcentury analytic philosophy with David Kaplan (or, better, of assisting David Kaplan as he did the teaching). Eventually, I needed to turn my attention to the manuscript that I had started, along with a few other smaller projects, but I let this one simmer on the back burner, occasionally stopping to give it a stir, add some ingredients, and, especially, filter out some of the accumulated muck.

I believe—or, hope—the result is a testimony to the benefits of time. May we work to give it to one another.

PRIMER ON FREE WILL AND MORAL RESPONSIBILITY

THIS SHORT PRIMER is meant for those unfamiliar with the philosophical literature on the problem of free will and moral responsibility. It sketches the basic, intuitive problem and explains some terms, such as *determinism*, *compatibilism*, and *incompatibilism*, that will be important in what follows. It moves extremely quickly through difficult terrain. I hope it is nonetheless of use to relative newcomers. Old-timers should feel free to skip ahead.

A disclaimer: The philosophical literature on this topic is vast; what follows is not an overview of that literature (there are no citations). It is not even a survey of the different philosophical positions taken in the literature. It is, rather, an opinionated introduction, meant only to remind the reader of, or to orient the reader in, some basic ideas and terms that will be presumed in what follows.

Our topic is P. F. Strawson's "Freedom and Resentment."[1] In it, Strawson aims to resolve an age-old dispute about moral responsibility and free will. He addresses the problem in its contemporary garb, in which the threat to moral responsibility and freedom is posed by the scientific thesis of *determinism*.[2]

As Strawson notes at the beginning of his article, the exact details of that thesis can be a matter of disagreement. For our purposes, we can work with this understanding: According to the thesis of determinism, the complete physical state of the world at any one

1. Peter F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962): 1–25. Page numbers refer to the reprint in this volume.

2. The problem may instead be posed by God's power or knowledge.

point in time, together with the laws of nature, determines the complete physical state of the world at all other points in time. Our world is *deterministic* if determinism is true of it.

It seems to many people that if our world is deterministic, then we are not free. After all, human actions are part of the physical world, and, if the world is deterministic, then every physical event—every movement you have ever made, every action you have ever taken or ever will take—was and will be determined by the laws of nature together with the state of the world as it was some time before you were born. But, it seems, if you enjoy free will, then what you do must be in your control or up to you. It must be the case that you could have done otherwise. However, you cannot change the laws of nature, nor can you change the past. And thus it seems that, if determinism is true, then what you do is not really up to you or in your control. If determinism is true, then, once we fix the past and the laws of nature, what you do has been fixed, as well. Thus it seems you are not free. (One might think that, if the world is instead indeterministic, our freedom would be secure. Unfortunately, things are not so simple.)[3]

3. Many now think that the world is, at some level, *indeterministic*—that the laws of nature and a given state of the world will determine only probabilities, or chances, of subsequent (or prior) occurrences. Some see, here, hope for human freedom. But that hope will be fulfilled only if humans can somehow, by making decisions, influence the chances established by the laws of nature. An immaterial soul or spirit somehow endowed with the power to affect the physical world might be able to intervene. It is much more difficult to see how a physically embodied mind, psychology, or personality could do so—because such a mind will, itself, be subject to the probabilistic laws.

Once we allow that our mind or personality is itself a part of the physical world, it will seem that indeterminism is no more hospitable to human freedom than determinism: if indeterminism is true, then, although your choices are not *determined*, they are left to chance. But things that are left to chance are not things that are *up to you*. The problem can therefore seem to be posed simply by any suitably robust form of scientific explanation.

The person Strawson calls, in his opening paragraph, the "genuine moral sceptic" rejects the possibility of an immaterial soul or spirit and reasons that, if determinism is *either* true *or* false, we are not free. But since determinism must be either true or

I have just sketched the position held by the *incompatibilist* about determinism and free will: human freedom is incompatible with the truth of determinism. The *compatibilist* disagrees. The compatibilist thinks that, even if determinism is true, humans can enjoy free will.

Upon first encountering the issue, most people find the compatibilist position utterly opaque: if determinism is true, how could we be free? We can begin to bring the compatibilist position into view in this way: notice that, if determinism is true, it is already true. But, we already know that we often (not always) affect certain things (not everything) in exactly the way we had meant to affect them—we sometimes bring certain things to be as we would have them to be, and, to that extent, we control them. Thus we already know that there are some things we can, at least in this sense, control—such as our coffee cup, our fork, or whether we will accept the invitation to next week's event—and other things we cannot—such as whether we will win the lottery, or the geopolitical situation, or our teenager's emotional state. Some things are, at least in this way, up to us; other things are not. We also regularly draw distinctions between cases in which someone acted voluntarily—when, for example, they accepted the invitation without coercion or duress—and cases in which they were, instead, pressured or threatened by someone else, or when the only alternatives they faced were clearly worse than what they chose. If we learn that determinism is true, that will not render us suddenly unable to draw these distinctions—not even the incompatibilist thinks so. Rather, the incompatibilist grants *these* distinctions, but claims that, if determinism is true, then we have been misidentifying one side of them as instances of human freedom. If determinism is true, the incompatibilist thinks, then, although we admittedly regularly affect our cups, forks, and dinner invitations in just the way we meant to, we do not *really* control anything; those things are not *really* up to us. And, the incompatibilist will continue, if determinism is true, then, even if we are not coerced or threatened by some other person,

---

false, we can conclude that we are not free. And since moral responsibility requires freedom, we can conclude that we are not morally responsible.

we are nonetheless forced to do what we do by the physical world—we do not *really* make choices, we are not *really* free.

We can begin to bring the compatibilist position into view by saying that the compatibilist does not accept the incompatibilist's "really"—the compatibilist does not think the "really" picks out anything that could be real, so to speak. And so the compatibilist is content with the kind of freedom and control we already know we sometimes enjoy.[4] (Again, this is not meant as a full explication of the compatibilist position. It is only the beginnings of a sketch, an attempt to bring at least the possibility of a compatibilist position into view. Strawson will provide a more robust defense.)

Notice that compatibilism and incompatibilism are views about the *compatibility* of two claims: the claim that determinism is true and the claim that humans enjoy free will. They are not views about the *truth or falsity* of these claims. And so, strictly speaking, one could be either a compatibilist or an incompatibilist while believing determinism is either true or false, and one could be either a compatibilist or an incompatibilist while believing humans either do or do not have free will. In fact, strictly speaking, a compatibilist could believe any combination of these. Only the incompatibilist faces a restriction: the incompatibilist cannot believe *both* that determinism is true *and* that humans enjoy free will—because, of course, as an incompatibilist, they believe these are incompatible.

Notice that, therefore, incompatibilism about determinism and human freedom comes in two varieties: some incompatibilists believe that determinism is true and therefore humans are not free,

4. Compatibilism comes in (at least) two very different varieties. The more conventional sort of compatibilist thinks what one would expect them to think, given the explanation above. The other sort, the Kantian, neo-Kantian, or "two standpoints" compatibilist, thinks that facts about the physical world and facts about human action occupy different realms, "standpoints," or conceptual schemes, and so cannot be brought into conflict with one another. Thus the two claims—that we are free and that the world is deterministic—are, so to speak, *so* incompatible that they cannot even be made to contradict. And *therefore* they are compatible: both can be true. For this second sort of compatibilist, the incompatibilist's "really" does not make good sense because the two kinds of facts cannot, so to speak, be brought into one frame.

while others believe that humans are free and therefore determinism is false.[5] The first sort of incompatibilist is sometimes called a *hard determinist*, while the second sort is called a *libertarian*.[6] The kind of freedom the libertarian thinks we enjoy is called *libertarian freedom*, or *contra-causal freedom.* This is the kind freedom that is incompatible with the truth of determinism; the kind we would enjoy if we were "really" free.

Finally, notice that, because libertarian freedom has been *defined* as the kind that is incompatible with the truth of determinism, the compatibilist is likely to think we do not enjoy *that* kind of freedom.[7] The compatibilist will think our freedom can take a different form—a form that does not require the falsity of determinism. Thus there is real danger of encountering, in this debate, what turns out to be a merely verbal dispute. If a compatibilist claims that we are free and a hard determinist disagrees, they may simply be talking past one another: They may agree that we do not enjoy *libertarian* freedom, and they may also agree that we do enjoy what the *compatibilist* is content to call freedom. They may be disagreeing only about how to use the word "free."

Let us turn, now, to responsibility. In addition to the debate about the compatibility of determinism and *free will*, there is a closely related debate about the compatibility of determinism and *responsibility*. There are, again, both compatibilist and incompatibilist positions on this related question. The incompatibilist thinks that, if determinism is true, then we are not responsible. The compatibilist thinks the truth of determinism has no implications for our responsibility.

Considering responsibility can transform what would otherwise be a merely verbal dispute about freedom into a substantive debate. Typically, everyone agrees that, if we are not free, then we are not

5. There is a third possibility: agnosticism about both claims.

6. The libertarian about human freedom is to be distinguished from the *political* libertarian. The two positions have some affinity, but no necessary connection.

7. I say "likely" because, strictly speaking, the compatibilist could be agnostic on the question. They could think that, if determinism turns out to be false, then we could enjoy libertarian freedom while also maintaining that, whether or not determinism is true, we enjoy what they would call freedom.

responsible.[8] However, the incompatibilist will think that responsibility requires *libertarian* freedom, and is therefore incompatible with determinism, while the compatibilist will think some other kind of freedom is sufficient for responsibility. Thus, the verbal dispute whether determinism is compatible with "freedom" may find substance as a debate about whether determinism is compatible with responsibility.

The compatibilist position about responsibility may seem harder to hold. The compatibilist maintains that the freedom (or control, or self-determination) that we already know we sometimes enjoy is enough (not only to talk about "freedom," but also) to ground responsibility—enough for praise and blame, reward and punishment, guilt, indignation, and condemnation. The incompatibilist objects: if whatever immoral choices we make were already determined, already in the cards, from sometime before we were born, how could it be just, fair, or appropriate to blame, punish, or condemn us for them? It can certainly seem that, if determinism is true, then we are not free in the way that is required for responsibility.

This is the debate that Strawson means to adjudicate. However, in doing so, he imagines two somewhat more specific contenders. He imagines, first, an incompatibilist about both freedom and responsibility who is what he calls a "pessimist": Strawson's pessimist is an incompatibilist who believes that determinism is *most likely* true. The pessimist thus is close to, though not quite, a hard determinist. In his other corner, Strawson imagines a very specific sort of compatibilist, whom he calls an "optimist." I will wrap up this primer by explaining the position of the optimist.

Strawson's optimist is a *consequentialist* compatibilist. *Consequentialism* is a position in moral philosophy that claims that whether an action is right or wrong, morally justified or unjustified, depends in some way on its consequences: wrong actions are those with bad consequences.

Consequentialism, as a moral theory, encounters some serious counter-examples, because, according to consequentialism, the ends

8. The exception is the *semi-compatibilist*. Some semi-compatibilists are compatibilists about responsibility but not about freedom. The semi-compatibilist may simply remain agnostic about whether we are free.

always justify the means. In fact, according to consequentialism, the ends (or, better, the consequences) are *precisely* what justify the means. But, according to ordinary moral intuitions, the ends do not always justify the means—according to ordinary moral intuition, sometimes you have to respect certain people's rights, or keep your promise, or tell the truth, even though it will result in a *worse* outcome than if you had not done so. Such cases present themselves as counter-examples to consequentialism. The consequentialist has strategies for addressing these cases. Rather than consider those strategies, I turn to the optimist.

The optimist is a consequentialist *compatibilist*—the optimist believes that, whether or not determinism is true, our familiar practices of holding others responsible are justified by their consequences. By "our familiar practices of holding others responsible," I mean such facts as, if you insult someone, they will resent you for it; if you show someone especially good will, they will be grateful; if you betray someone, they may confront you about it and others are likely to distrust you; if you steal from someone, you will not only be expected to return what you took, but also be subject to criminal punishment. The optimist points out that these practices serve important social functions—they promote certain sorts of attitudes and behavior while disincentivizing other sorts. Because we praise, blame, reward, and punish one another, things go better. And thus, according to the optimist, these practices are justified by their consequences—they are, as Strawson puts it, justified by their "efficacy . . . in regulating behavior in socially desirable ways" (108).

Importantly, because we already know that these practices effectively serve these functions, the optimist is confident that, whether or not determinism turns out to be true, these practices will be justified. As Strawson puts it in his opening pages, the optimist argues that "the facts as we know them supply an adequate basis for [our] concepts and practices," even though those facts do not show determinism to be false.[9]

9. Strawson himself agrees with the general *form* of the optimist's argument as he sketches it on page 108: Strawson's own position is that "the facts as we know them" provide an adequate basis for our concepts and practices without showing

Before moving on, we should notice how differently the optimist and the incompatibilist treat certain familiar facts about when we do and when we do not hold one another responsible. It is a familiar fact that we excuse people if they were innocently ignorant or if it was an accident or an unforeseeable consequence. Further, we do not hold young children as responsible as adults, nor those who have been drugged, nor those suffering from serious mental incapacities, and so on.

The incompatibilist typically looks at these familiar facts about excuse and exemption and draws, from them, moral principles. That is, the incompatibilist typically takes these facts as data, so to speak, from which to extract moral truths: the incompatibilist might conclude that a person cannot rightly be held morally responsible unless they could have done otherwise, or, perhaps, unless what they did was in their control, or unless it was up to them.[10]

Some such principle, supported in this way by familiar facts, in turn supports the incompatibilist's incompatibilism: the incompatibilist points out that, if determinism is true, then no one could ever do otherwise, or nothing was ever in our control, or nothing was ever up to us. Thus, if determinism is true, no one can rightly be held morally responsible.

In contrast, rather than taking the familiar facts as data from which to draw a moral principle, the optimist's overarching moral theory about the justification of our practices *predicts* these facts about excuse and exemption—it nicely predicts the boundaries between cases when we do and do not hold people responsible.[11] After all, blaming people for what is unforeseeable or an accident will not provide any useful disincentive for future unforeseeable or accidental

---

determinism false. However, Strawson points to a very different set of facts that we know.

10. The incompatibilist argues in what one might think of as a "bottom-up" way: They look at the cases as data, and then argue, up from the data, to a principle. The principle will, so to speak, fit the curve of the data.

11. That is to say, the optimist can argue "top down," from their theory about the nature of morality to the principles in light of which, based on that theory, we would expect to excuse or exempt.

cases. Similarly, young children are slowly developing their capacities for self-control, and so holding them responsible will affect their future behavior only within those limits. Likewise for those who suffer from mental illness. The optimist's position thus nicely accounts for the contours of our familiar practices.

Recall, now, that the compatibilist and the hard determinist agree that responsibility requires "freedom" but disagree about how to use the word. We can now see how the optimist understands "freedom": the optimist will understand a person to be "free" whenever the person is responsible. That is to say, for the optimist, if you are rightly held responsible, you thereby *count as* free. As Strawson puts it, "what 'freedom' means here is nothing but the absence of certain conditions the presence of which would make moral condemnation or punishment inappropriate" (108). Thus, the optimist maintains what we might think of as a "list view" of freedom: to act freely just is to be a reasonably capable adult acting without ignorance, accident, coercion, and so on. For the optimist, the conditions that belong on the list are those that show that holding you responsible for your behavior would be inefficacious—it would do no good.

Notice what this implies: for the optimist, so long as you can be effectively given incentives or disincentives for or against similar future behavior, you are rightly held responsible—and *therefore*, according to the optimist, you count as having acted freely. Of course, we already know that, whether or not determinism is true, you can, in certain circumstances, effectively be given incentives and disincentives. Thus, according to the optimist, we already know that both freedom and responsibility are compatible with the truth of determinism.

The pessimist finds the optimist's picture shocking and appalling. Strawson is sympathetic—he, too, feels the shock and appall. However, he is not sympathetic to the pessimists' concern about the truth of determinism. Strawson is, himself, a compatibilist, though of a very different sort than the optimist. Thus he hopes to adjudicate their debate.

At this point, the reader is, I hope, prepared to dive into Strawson's difficult but fascinating text (which is reprinted in this volume,

starting on page 107). With the scene set and the players' positions before us, hopefully Strawson's thought in the opening sections of his article will now be available to the reader. The remainder of this volume examines the central portion of Strawson's text, line by line, unearthing the assumptions that underwrite the underappreciated central argument of this enduring paper.

Introduction

WHEN P. F. STRAWSON'S "Freedom and Resentment" first appeared, nearly sixty years ago, it forced a profound shift in the debate about free will and moral responsibility.[1] For decades since, it has inspired views on wide-ranging topics.[2] Most of the ongoing attention has focused on Strawson's fascinating and fecund notion of "reactive attitudes." In contrast, the central argument of the paper has received relatively little attention.[3]

The central argument claims that, because determinism is an entirely *general* thesis, true of everyone at all times, its truth would not show that we are not, in fact, morally responsible. It is a startling claim. The neglect given to the argument for it would be surprising, if that argument were not so difficult to discern.

1. Peter F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962): 1–25. Page numbers refer to the reprint of the essay in this volume.

2. Recent titles on further-ranging topics include Stephen Darwall, *The Second-Person Standpoint: Morality, Respect, and Accountability* (Cambridge, MA: Harvard University Press, 2006) and Akeel Bilgrami, *Self-Knowledge and Resentment* (Cambridge, MA: Harvard University Press, 2006). Recent Strawson-inspired work on free will and moral responsibility is legion.

3. Some attention has been given to the argument in the (considerable) time this book has been in preparation. See, e.g., the papers collected in David Shoemaker and Neal Tognazzini, eds., *Freedom and Resentment at 50*, Oxford Studies in Agency and Responsibility, vol. 2 (New York: Oxford University Press, 2015).

When the argument is considered, it is often interpreted as relying on a thought about our psychological capacities: we are simply not capable of abandoning the reactive attitudes, across the board, in something like the way we are simply not capable of remembering everything we are told. We do not have the right equipment. Given our psychological limitations, we are stuck treating one another *as if* we are morally responsible—we are incapable of doing otherwise. Therefore, according to this interpretation, we should rest content in the thought that we *are* morally responsible—asking whether we ought to treat one another differently is useless. I will call this "the simple Humean interpretation" and the thought on which it relies "the simple Humean thought."

A different line interprets Strawson as relying on something like a conceptual point: you can neither support nor call into question the whole of a practice using notions that are, themselves, constituted by that practice. Thus, you cannot ask whether our moral practices, taken as a whole, are, themselves, morally just, right, appropriate, or fair. Doing so would be like asking whether the game of baseball is, itself, "fair" or "foul" in the sense of those words established by the game—"fair" or "foul," in *that* sense, can be rightly asked of batted balls or of territory in the baseball field, but the question cannot be sensibly asked of the game itself, taken as a whole. On this second interpretation, Strawson accuses his opponent of a sophisticated kind of confusion. I will call this "the broadly Wittgensteinian interpretation" and the thought on which it relies "the broadly Wittgensteinian thought."

Both the simple Humean thought and the broadly Wittgensteinian thought can be found in Strawson's paper, and he makes use of each. But neither interpretation would lead you to expect what you will find, looking at the central text: Strawson twice accuses his opponent of being caught in some kind of contradiction. So neither interpretation, on its own, is correct.

By providing a close reading of the central text, I will do my best to articulate Strawson's more interesting, and more powerful, argument. The argument depends on an underlying picture of the nature of moral demands and moral relationships—a picture that has gone largely unnoticed, that is naturalistic without being reductionistic,

and that is worthy of careful consideration. Having drawn out this underlying metaethical picture, I will begin to subject it to some philosophical scrutiny. I hope to show that it can withstand the objections that are both the most obvious and the most serious, leaving it a worthy contender.

1

Strawson's Strategy

LET US BEGIN at the article's beginning. Strawson's stated aim is to adjudicate and reconcile the debate between a gloomy incompatibilist about moral responsibility and determinism, whom he calls "the pessimist," and a consequentialist compatibilist, whom he calls "the optimist."

The optimist argues that our practices of holding one another responsible are justified by their good consequences, whether or not determinism is true. By engaging in these practices, we secure important social goods (we reinforce prosocial behavior, disincentivize antisocial behavior, and so build a well-functioning society).[1]

The pessimist finds this attempted justification appalling. If we justify our practices of blaming and punishing by appeal to their good consequences, the question of whether someone *deserves* blame or punishment becomes the question of whether blaming and punishing in such circumstances (generally) leads to good outcomes. But this, the pessimist thinks, is just to ignore the question of whether anyone *really* deserves blame or punishment—whether anyone is in fact responsible for his or her behavior. And the pessimist is

1. The position was more popular at the time Strawson was writing. See, e.g., P. H. Nowell-Smith, "Freewill and Moral Responsibility," *Mind* 57, no. 225 (1948): 45–61 and J. J. C. Smart, "Free Will, Praise, and Blame," *Mind* 70, no. 279 (1961): 291–306.

pessimistic because he thinks not only that, if determinism is true, then no one is responsible, but also that determinism is very likely true.[2]

Strawson means to adjudicate the dispute. He sees merit in each position. He agrees with the pessimist that the optimist's position distorts our notions of moral blame and guilt beyond recognition. But, with the optimist, Strawson does not think that preserving these notions requires the falsity of determinism.

In fact, Strawson thinks that, by appealing to determinism, the pessimist is making something like the same error the optimist makes by appealing to consequences: they each assume that our practices of holding one another responsible require a justification that those practices do not require. As Strawson puts it, the framework constituted by our practices "neither calls for, nor permits, an external 'rational' justification" (131).

We can already see that, broadly speaking, Strawson would like to adjudicate the dispute by convincing each side to stop talking one step earlier, so to speak: he would like the optimist to stop talking about consequences and he would like the pessimist to stop talking about determinism. He would like them both to stop talking—to stop attempting to justify our practices of holding one another responsible—and instead rest content with what he calls "the facts as we know them" (108). But it is difficult to see how or why one can legitimately rest content right there. It is hard to see why further justification is not required.

By understanding the central argument of the text, we will eventually come to see why Strawson thinks we can stop talking. Again, the central argument reaches the conclusion that, because determinism is a general thesis—true of everything—it will not show our practices of holding others responsible unjustified.

2. The pessimist is not the libertarian (the incompatibilist who thinks that, because we are free, determinism must be false), but nearer, in contemporary terms, to the "hard determinist."

Strawson's Picture of Responsibility

Strawson's central argument depends on a particular picture of what it is to be responsible.

To paint his picture, Strawson first draws our attention to how very much we care about how other people think of us, or, as he puts it,

> the very great importance that we attach to the attitudes and intentions towards us of other human beings, and the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs about those attitudes and intentions. (111)

To illustrate, he points out,

> If someone treads on my hand accidentally, while trying to help me, the pain may be no less acute than if he treads on it in contemptuous disregard for my existence, or with a malevolent wish to injure me. But I shall generally feel in the second case a kind and degree of resentment I shall not feel in the first. (112)

Even if the physical pain is the same in both cases, we care about the motive, or the quality of will, that led to that pain. Justice Holmes makes a similar point when he notes, "Even a dog distinguishes between being stumbled over and being kicked."[3] Perhaps (though only perhaps) the dog simply registers whether a fight-or-flight response is needed. In contrast, we care about how we stand in the other's world.

Strawson next identifies a class of attitudes, which he calls the "reactive" attitudes, such as resentment or gratitude, which we adopt in response to "the quality of others' wills towards us, as manifested in their behavior" (121). These stand in contrast with "objective" attitudes: responses such as frustration or relief, which we might have to events and states of affairs we believe were not willed by anyone. So, while we might be frustrated when our plans are thwarted by a downed tree, we do not resent the tree for lying there (or, if we do,

3. Oliver Wendell Holmes Jr., "Early Forms of Liability," Lecture I from *The Common Law* (1909).

we recognize this as a mistake). If an unsteady board bears your weight in a time of need, you feel relieved, not grateful. Though you may be disappointed when the strap on your old, reliable bag breaks, you do not feel betrayed. If, on the other hand, you believe that the tree was downed, the board supported, or the strap broken by someone, on purpose, with you in mind, then you might resent, feel grateful, or feel betrayed. Resentment, gratitude, and feelings of betrayal are reactive attitudes. When contrasted with these, frustration, relief, and disappointment are objective attitudes.[4]

In addition to these "personal" reactive attitudes, Strawson notes their "impersonal" or "vicarious" analogues. The impersonal reactive attitudes are those we adopt in response to our perception of the quality of another person's will toward *others*. They include indignation (an analogue of resentment) and moral admiration (an analogue of gratitude). Attitudes such as guilt and remorse Strawson identifies as self-directed reactive attitudes—those we have in response to our perception of the quality of *our own* will toward others.

In general, then, a reactive attitude is *x*'s reaction to *x*'s perception of or beliefs about the quality of *y*'s will toward *z*. In the impersonal reactive attitudes, *x*, *y*, and *z* are different persons. In the case of the personal reactive attitudes, the same person stands in for *x* and *z*. In the case of self-directed reactive attitudes, the same person stands in for *x* and *y*.

4. Strawson does not introduce the reactive attitudes in this way. In fact, he only rarely talks of "objective attitudes," in the plural, and he does not consider responses to inanimate objects or unwilled events. Instead, he usually talks about adopting "*the* objective attitude" or "*an* objective attitude" toward a person or toward their actions. I think the contrast I have drawn is helpful in locating what Strawson has in mind, but it is worth keeping in mind that it goes beyond Strawson's text. I am grateful to an anonymous reviewer for calling attention to the difference.

Strawson lists anger as a reactive attitude. I suspect he has in mind what we might call being "angry with" in contrast to being "angry that." Similarly, he says you cannot "reason with" someone toward whom you have adopted an objective attitude. I suspect he has in mind a kind of "reasoning with" that requires the other to be "reasonable," in the sense of open to and considerate of the interests of others. (He allows that you can both "talk" and "negotiate.")

It is crucial for Strawson's argument that the reactive attitudes are modified or suspended in cases of two distinct kinds. There has been much ink spilled interpreting this distinction. Here is how I believe it should be understood:

In the first kind of case, we learn we were mistaken about the quality of the will in question, and therefore our reactive attitude—our reaction to our *perception of* or our *beliefs about* the quality of that will—must change. We learn, for example, that the actor was innocently ignorant, or that it was an accident, and so we see that they really meant no harm. They only appeared to. Or, now that we know the person was threatened with their life, we can see that their choice showed no disrespect. Or, now that we know that the person was in fact pushed, or that this particular bit of behavior was just a reflexive movement, we can see that the movement displayed no will at all. As an imperfect slogan, we could say these are cases in which we learn that "the will was not ill."[5]

5. Notice that Strawson includes, in his *first* category, "he was pushed." I believe he would include other cases in which we learn that no will was at work: reflex movements, epileptic fits, or having your body used like a puppet.

It will seem to some readers that such cases belong in the second category (it once seemed that way to me). It will seem so because it is tempting to think of this first category as cases in which the person is responsible for their action or attitude and yet showed no ill will, while thinking of the second category as cases in which the person is not responsible. Since reflexes, epileptic fits, and being used like a puppet are cases in which you are not responsible, it can seem they should go into the second category.

However, at this stage of the argument, relying on intuitions about who is responsible would be illicit. Strawson is trying to understand who is responsible by considering ways in which we suspend or modify the reactive attitudes. His first category are cases in which we suspend or modify the reactive attitudes due to *error about the quality of the will*—including error about whether any will was operative at all. If there was no will at work (if he was pushed or used like a puppet), it is still true that we were in error about its quality—in our slogan, that will was not ill. Yet (looking ahead to the second category), *that* will, the will of *that* person in *those* circumstances, might still matter in the usual way. Such cases belong in the first category. Thanks to Sarah Buss for pressing for clarification.

Importantly, the information we receive in this first kind of case, while prompting us to revise our reactive attitudes, does nothing to suggest that the person in question is not an apt target of such attitudes. The quality of that person's will continues to matter to us in the usual way; we were simply mistaken about *which* quality (if any) was manifest, in the case at hand.

Not so, for the second sort of case. In these cases, rather than come to see that we were mistaken about the *quality* of this will, we come to see that it would be a mistake to react to *this* will, in these circumstances, in the usual way, regardless of its quality. This will, in these circumstances, does not call for the usual sort of reactions. It may be that the person really did mean harm, or know what he was doing. It was not an accident. We were in fact shown disregard or malice. Nonetheless, when we come to learn, for example, that the person in question was under extreme strain, or is mentally ill, or is a very young child, we do not react as we otherwise would. As an imperfect slogan, we could say these are cases in which "the ill will does not matter in the usual way." Rather than react with the corresponding reactive attitude, we shift to more objective attitudes—we adopt what Strawson calls "the objective attitude."

Strawson notes that this second type of case comes in two sub-varieties. Sometimes we discount the importance of someone's will temporarily, due to extreme or unusual circumstances, saying, for example, "he wasn't himself." At other times, the discounting is due to some more enduring condition that renders the person incapacitated for tolerably ordinary adult interpersonal relationship, such as disease or immaturity.

Having sketched these two sub-varieties of the second sort of case, Strawson says, "But there is something curious to add to this." He then makes what will be a very important observation: we sometimes shift from the reactive to the more objective attitude even in cases in which the will in question is neither immature, diseased, nor in extreme or unusual circumstances. He says,

> The objective attitude is not only something we naturally tend to fall into in cases [of] abnormalities or immaturity. It is also something which is available as a resource in other cases, too . . . we *can*

> sometimes look with something like the same [objective] eye on the behavior of the normal and mature. We *have* this resource and can sometimes use it—as a refuge, say, from the strains of involvement; or as an aid to policy; or simply out of intellectual curiosity. (116)

This "use" of this "resource" should, I think, be familiar. Sometimes the "strains of involvement"—the emotional effort and expense of engaging with, say, a particular coworker or family member—seems too much. We want to disengage. Sometimes we can—sometimes we can relate to a person simply as an "issue," a "problem" to deal with (as in, "Oh, *him*—he's another issue altogether"). Alternatively, we might similarly disengage, not because the person is too much, but rather because we, ourselves, are exhausted or under great strain. Similarly, when we are thinking through what policies to adopt, or playing the role of therapist, we can step away from what would otherwise be our natural reaction to outrageous or offensive behavior and instead adopt a more objective attitude.

I argue, below, that this available "resource" plays a central role in Strawson's central argument. It provides a *third* sub-variety of case in which ill will does not matter to us in the usual way—a third sub-variety of case in which we suspend the reactive attitudes and shift to a more objective mode.

Strawson seems reluctant, though, simply to class this third sub-variety with the other two; he instead sets it apart as something "curious." It seems he is thinking that, while we "naturally tend to fall into" the objective attitude in the first two sub-varieties of cases, our natural tendency to do this in those cases becomes, for us, a capacity we can make use of—a "resource" we can employ—in other cases, for a variety of reasons, more or less at will.

Having thus surveyed the ways in which the reactive attitudes are modified or suspended, Strawson notes a connection between reactive attitudes, on the one hand, and an expectation of and demand for goodwill or regard, on the other. He says,

> The personal reactive attitudes rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of goodwill or regard on the part of other human beings towards

> ourselves; or at least on the expectation of, and demand for, an absence of the manifestation of active ill will or indifferent disregard. (122)

Likewise,

> The generalized or vicarious analogues of the personal reactive attitudes rest on, and reflect, exactly the same expectation or demand in a generalized form; they rest on, or reflect, that is, 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 behalf of whom indignation may be felt; i.e., as we now think, towards all men. (122)

These expectations of and demands for regard are, broadly speaking, *moral* expectations and demands.

Strikingly, Strawson later claims that "the making of the demand *is* the proneness to such attitudes" (129, emphasis original).

Given this striking identification, notice that, when we shift to a more objective mode (when "the ill will does not matter"), with children, for example, or with those suffering from dementia, we no longer stand ready to respond with reactive attitudes—we are no longer *prone* to such attitudes. Thus, given the identification, we also, thereby, cease to make the associated demands. By shifting to an objective attitude, we lift the corresponding demands. Accordingly, cases of the second kind are often referred to as cases of "exemption": in these cases, the person is exempted from the usual demands.

In contrast, in cases of the first kind (when "the will was not ill"), in which we simply come to see that we were mistaken about the quality of the will, the new information prompts us to modify our attitude "*without* inhibiting, or displacing, the sort of demand of which [the attitude] can be said to be an expression" (123, emphasis original). The demands stay in place. These are often referred to as cases of "excuse."[6]

6. We saw, above and in the previous note, that Strawson includes, in his first category, cases in which I learn that I was mistaken about the quality of your will because I was mistaken about whether your will was active at all (you were

Hereafter I will use the labels "excuse" and "exemption" to refer to Strawson's two types of case.[7]

Turning, now, to responsibility: Strawson identifies those who are rightly subject to the general demand to show good will as those who are morally responsible. Since revisions of the first kind (excuses, in which we were mistaken about the quality of will) do not displace the demand,

> considerations of this group have no tendency to make us see the agent as other than a morally responsible agent; they simply make us see the injury as one for which he was not morally responsible. The offerings and acceptance of such exculpatory pleas as are here in question in no way detracts in our eyes from the agent's status as a term in moral relationships. On the contrary, since things go

---

pushed, or your body was used like a weight). He should do so: these are not cases in which the quality of your will no longer matters to me. However, doing so may seem to generate a strange consequence: one might note that we would suspend or revise our reactive attitude in just the same way if we learned that what we thought was the action of a person was in fact just the movement of an inanimate object, perhaps a mannequin blown by the wind. By using the label "excuse" for Strawson's first category, it seems we are claiming that we sometimes excuse mannequins.

The problem is in our label, not Strawson's theory (he does not use the label). (There will be similar difficulty with the label "exemption," as we will see.) Though "excuse" is a convenient label, we should remember that it is currently just that—a technical term, one that would allow us to "excuse" even mannequins. Thanks to Sarah Buss for raising these issues.

7. There is some delicacy about whether use of the resource should be counted as an exemption, i.e., as a case in which the expectations and demands are no longer in place. On the one hand, if the making of the demand *is* the proneness to the attitude, then one might think that, by using our resource, by stepping away from the attitudes, we are also abandoning our expectation and demand. On the other hand, one might think that, when using our resource, we are still *prone* to the reactive attitudes; we are simply choosing to step away from them temporarily. Thus the demand stays in place; we simply ignore or discount its violation. We will return to this issue. See chapter 2, footnote 15; chapter 5, footnotes 15, 17, and 35.

> wrong and situations are complicated, it is an essential part of the life of such relationships. (124)

In contrast, in cases of the second sort (exemptions), in which we suspend the reactive attitudes, we cease to regard the person as responsible:

> He is not seen as one on whom the demands and expectations lie in that particular way that we think of them as lying when we speak of moral obligation; he is not, to that extent, seen as a morally responsible agent, as a term of moral relationships, as a member of the moral community. (124)

The person is exempted from responsibility altogether.[8]

We are now ready to piece together Strawson's picture of what it is to be morally responsible. According to Strawson, to be morally responsible is to be a term in moral relationships. (I believe he is thinking of relations on analogy with "to the left of," so that, if *a* is to the left of *b*, *a* and *b* are terms in that relation, which might be written in logical notation as L(*a*,*b*).) To be a term in such relationships is, at least in part, to be the apt target of certain demands for goodwill. To make such a demand of another *is* to be prone to respond to the quality of that person's will with reactive attitudes. So, to be morally responsible is to be such that others are rightly prone to these attitudes, in response to the quality of your will. We might say that, for Strawson, to be responsible is to be such that the quality of your will *matters* to others in this distinctive way.

8. We might think there is room for a third sort of change in the demands and expectations: might we not, in light of new information, modify the demands—ease them, or recalibrate them—rather than lift them entirely? That is, might we not be correct about the quality of will, but change our mind about whether or to what degree a will of that quality, in these circumstances, violates or disappoints reasonable expectations or demands? In such a case, we would shift, rather than lift, the demands, and recalibrate, rather than suspend, the reactive attitudes.

It would have been helpful if Strawson had explicitly addressed this third possibility. We will consider such a case below.

This picture of what it is to be responsible—Strawson's sense of the subject matter—provides much of his contribution to this topic. However, it has not generally been noticed that the picture Strawson paints is not merely a masterly detail of a certain fascinating range of our psychology. It is not even just a picture of what it is to be responsible. He is, rather, sketching what we might call a metaphysics of morals: an underlying picture of the nature of moral and interpersonal demands and requirements—one he paints by observing our actual practices, in his style of "descriptive metaphysics."[9] This underlying metaethical picture is crucial for Strawson's argument—and interesting in its own right. It will emerge more fully as we work to understand his text.

We turn, now, to his central argument.

The Central, and Seemingly Facile, Argument

To introduce his argument, Strawson first poses what he takes to be the question at hand:

> What effect would, or should, the acceptance of the truth of a general thesis of determinism have upon these reactive attitudes? More specifically, would, or should, the acceptance of the truth of the thesis lead to the decay or the repudiation of all such attitudes? (117)

He then immediately considers an objection raised in response to the cavalier way in which he has treated the thesis of determinism: "But how can I answer, or even pose, this question, without knowing *exactly* what the thesis of determinism is?" In response, he announces his argumentative strategy:

> There is one thing we do know: that if there is a coherent thesis of determinism, then there must be a sense of "determined" such that, if that thesis is true, then all behavior whatever is determined in that sense. Remembering this, we can consider at least what

9. See Peter F. Strawson, *Individuals: An Essay in Descriptive Metaphysics* (London: Routledge, 1959).

> possibilities lie formally open; and then perhaps we shall see that the question can be answered *without* knowing exactly what the thesis of determinism is. We can consider what possibilities lie open because we have already before us an account of the ways in which particular reactive attitudes . . . may be, and, sometimes, we judge, should be, inhibited. (117)

Strawson means to use the fact that determinism is a *general* thesis, true of everyone at all times, together with the account he has given of the ways in which the reactive attitudes are and should be modified or suspended, to reach the conclusion that acceptance of the truth of determinism neither would nor should lead to the decay or repudiation of the reactive attitudes.

It is surprising to think Strawson can draw this very strong conclusion from those meager resources. If it works, it is a powerful argument.

Embarking on his task, Strawson first considers the first sort of revision, in which we learn that we were mistaken about the quality of will and so revise which, if any, attitude we hold in response to that will—while continuing to relate to the person in the usual way. He notes that the decay or repudiation of *all* reactive attitudes could not be a revision of this sort: if we lost or repudiated all the reactive attitudes, we would no longer relate to anyone in the usual way.

He makes a second point: We make this first sort of revision when we see we were mistaken about the quality of the other's will. Therefore, to excuse *everyone* in this way would be to come to see we had been mistaken to think that anyone ever meant any harm, or showed any disrespect, or. . . . As Strawson puts it, to excuse everyone in this way would be "not a consequence of the reign of universal determinism, but of the reign of universal goodwill" (117).

Thus, the first sort of revision (in which we come to see that "the will was not ill") is not fit for general application.

Strawson then turns to consider the second sort of revision. In the terms of our slogan, Strawson considers whether, given the truth of determinism, ill will would or should cease to matter to us in the usual way. In his own terms, he considers whether we would or

should come to view everyone objectively, regardless of the quality of anyone's will. He argues we neither would nor should.

Strawson's argument to this conclusion is unexpected, extremely short, and typically overlooked. I believe it contains his main point. He says,

> Next, I remarked that the participant attitude, and personal reactive attitudes in general, tend to give place, and, it is judged by the civilized, should give place, to objective attitudes, just insofar as the agent is seen as excluded from ordinary adult human relationships by deep-rooted psychological abnormality—or simply by being a child. But, it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition. (118)

Strawson immediately follows by saying, "Now this dismissal might seem altogether too facile; and so, in a sense, it is." Strawson's quick follow-up tempts many to dismiss his dismissal. But we need to pause and take seriously Strawson's thought. We need to understand both why Strawson would put forward this argument at all and why he thinks it is facile only "in a sense."

Notice that Strawson's seemingly facile argument starts with a claim—a very strong, questionable claim—about *the reason why* we exempt, why we do and should suspend the reactive attitudes: we do and should "just insofar as [that is, if, only if, and to the extent that] the agent is seen as excluded from ordinary adult human relationships."[10] Strawson claims that we exempt people from these attitudes just in case we believe they are (as he later puts it) "incapacitated" for ordinary adult interpersonal relationships. And, from this single premise, he immediately concludes that his opponent is committed to a contradiction.

The conclusion follows immediately, if we interpret Strawson's "ordinary" as "*statistically* ordinary." On this interpretation, Strawson's first

10. Throughout, Strawson is noncommittal about whether "the reason why" is, as it is sometimes put, "explanatory" or "normative." Note the constant disjunction, "would or should." I will try to remain equally noncommittal. Note, too, the "insofar as": Strawson seems to have in mind degrees. See his later comments about children.

premise claims that we do and should suspend the reactive attitudes just insofar as we believe the agent is excluded from statistically ordinary adult relationships.[11] Strawson's opponent thinks that a general thesis—something true of everyone—will give us reason to suspend the reactive attitudes. So, by Strawson's first premise, the opponent must think this general thesis will give us reason to believe that everyone is excluded from what is statistically ordinary—that everyone is abnormal. But that is a contradiction. And, any thesis which implies a contradiction must itself be contradictory, not a "coherent thesis" (117).

Thus, from his initial premise, Strawson is ready to conclude, straightaway, that nothing true of everyone could give us reason to suspend the reactive attitudes and adopt the objective ones. This is, I believe, Strawson's core idea—though the argument must be amended, for reasons that will emerge.

The opponent will find Strawson's initial premise baffling. Why should whether someone is or should be exempted from the reactive attitudes and associated demands depend on whether that person is ordinary, *statistically speaking*? Strawson owes an explanation.

But Strawson does not addresses the opponent's bafflement. Instead, he sees an entirely different problem with his argument. He had allowed, earlier, for three sub-varieties of cases in which we suspend the reactive attitudes and adopt a more objective stance, but his argument addresses only one of them, the second sub-variety (in which a person is seen as excluded from ordinary relationships due to disease or immaturity). And so this initial argument is, in at least this way, facile: the initial premise is false, by Strawson's own admission. It is not the case that we adopt the objective attitudes *just insofar as* we judge a person to be incapacitated for ordinary adult relationships. We need to examine the other sub-varieties.

We can quickly dispense with the first sub-variety (Strawson claimed we could dispense with it as he introduced it). These were

11. This interpretation of "ordinary" meets much resistance. I do not see how to make sense of the passage without it. Once we make this interpretive move, the rest of the paper follows.

cases in which the reactive attitudes are inhibited temporarily, due to extreme or unusual circumstances ("he is under great strain," "he wasn't himself"). But, just as it cannot be case that everyone is abnormal, so it cannot be the case that everyone is always in an unusual or extreme circumstance (nor that no one is ever themselves).

It is the third sub-variety that troubles Strawson, the use of the resource. He acknowledged that sometimes, to avoid "the strains of involvement," or even just out of intellectual curiosity, we "look with something like the same [objective] eye on the behavior of the normal and mature" (116). Sometimes we use our resource for reasons that have nothing to do with whether the person is in extreme or unusual circumstances or is capable of tolerably ordinary adult interpersonal relationships. Thus, Strawson has not yet ruled out the possibility that the truth of determinism provides a reason to treat everyone in the way in which we now treat the abnormal—because the truth of determinism might somehow give us reason to employ our resource at all times.

Strawson, in fact, says as much. Immediately after giving the seemingly facile argument, he says,

> Now this dismissal might seem altogether too facile; and so, in a sense, it is. But whatever is too quickly dismissed in this dismissal is allowed for in the only possible form of affirmative answer that remains. We can sometimes, and in part, I have remarked, look on the normal (those we rate as "normal") in the objective way in which we have learned to look on certain classified cases of abnormality. [Strawson here refers to our "resource."] And our question reduces to this: could, or should, the acceptance of the determinist thesis lead us always to look on everyone exclusively in this way? [That is, could, or should, acceptance of the deterministic thesis lead us to always employ this resource?] For this is the only condition worth considering under which the acceptance of determinism could lead to the decay or repudiation of participant reactive attitudes [as the rest can be shown incoherent]. (118)

Once one considers the possibility that Strawson is, in this passage, addressing the worry that we might have reason to use our resource, I believe the interpretation becomes irresistible.

Nonetheless, those who remain doubtful can consider the first several paragraphs of the article's sixth section. There Strawson revisits the initial dispute between the optimist and the pessimist. He highlights the fact that the optimist represents our practices "as instruments of policy, as methods of individual treatment and social control" (128). By so representing them, the optimist has suggested (both to Strawson and to the pessimist) that we adopt an objective attitude toward one another, even as we continue to do certain things—impose sanctions, create incentives—that the optimist will incorrectly call "holding responsible."[12] Strawson explicitly connects this pragmatic, policy-minded aspect of the optimist's position to the use of the resource:

> I have remarked that it is possible to cultivate an exclusive objectivity of attitude in some cases, and for some reasons, where the object of that attitude is not set aside from the developed inter-personal attitudes by immaturity or abnormality. And the suggestion which seems to be contained in the optimist's account is that such an attitude should be universally adopted to all offenders. . . . [And the pessimist thinks] if to all offenders than to all mankind. (128–29)

The optimist's pragmatic justification would have us adopt an objective attitude toward the normal in ordinary circumstances—but that would be, in Strawson's terms, using our resource at all times. Strawson needs to address this possibility.[13]

Moving forward, then, we should note the remaining questions, both for us and for Strawson. The most pressing question for us is this: why does Strawson think that whether we would or should exempt depends on what is ordinary, *statistically speaking*? Why should statistics

12. This results in "conceptual shock" to both the pessimist and Strawson.

13. For further support, see Strawson's much later *Skepticism and Naturalism*, where the resource is that which makes doubt about moral responsibility especially pressing. See *Skepticism and Naturalism* (New York: Columbia University Press, 1985), 33–35, 40.

matter? Meanwhile, the remaining question, for Strawson, is very different. He is asking whether the truth of determinism could, would, or should lead us to exercise the resource available to us and react to the normal in the way we characteristically react to the outliers.

We will, in the next chapter, consider Strawson's answer to his own question, and, in so doing, start to answer our question.

2

The Resource and the Role of Statistics

THE PREVIOUS CHAPTER presented what I take to be Strawson's central idea. He sees cases of exemption—cases in which we suspend our reactive attitudes, shift to objective attitudes, and thereby lift the usual moral and interpersonal demands and expectations—as *outlier* cases. He is therefore tempted to conclude, immediately, that no general thesis could be a reason to exempt—because it could not be the case that we are all outliers.

However, Strawson is not completely satisfied with this argument, because he had earlier noted what he calls our "resource": we can and sometimes do suspend the reactive attitudes even in cases in which both the person and the circumstances are tolerably normal. Thus Strawson believes that he needs to address the question of whether we could, would, or should, in light of the truth of determinism, use our resource all the time (in fact, he and the pessimist take the optimist to suggest that we do so, whether or not determinism is true).

Meanwhile, the reader, I assume, will be baffled by Strawson's appeal to statistical normalcy (so baffled, in fact, that this interpretation of Strawson has not been previously considered).

In this chapter, we will consider Strawson's treatment of what he sees as the remaining question, and, in so doing, we will start to address the reader's bafflement.

Here, again, is the text in which Strawson first makes his seemingly facile (but, I claim, core) argument and then raises his worry about it:

> Next, I remarked that the participant attitude, and personal reactive attitudes in general, tend to give place, and, it is judged by the civilized, should give place, to objective attitudes, just insofar as the agent is seen as excluded from ordinary adult human relationships by deep-rooted psychological abnormality—or simply by being a child. But, it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition.
>
> Now this dismissal might seem altogether too facile; and so, in a sense, it is. But whatever is too quickly dismissed in this dismissal is allowed for in the only possible form of affirmative answer that remains. We can sometimes, and in part, I have remarked, look on the normal (those we rate as "normal") in the objective way in which we have learned to look on certain classified cases of abnormality. [Strawson here refers to our "resource."] And our question reduces to this: could, or should, the acceptance of the determinist thesis lead us always to look on everyone exclusively in this way? [That is, could, or should, acceptance of the deterministic thesis lead us to always employ this resource?] For this is the only condition worth considering under which the acceptance of determinism could lead to the decay or repudiation of participant reactive attitudes [as the rest can be shown incoherent]. (118)

Immediately after posing this remaining question, Strawson gives a complicated answer to its first two parts. Here is the text (which immediately follows the text above):

> It does not seem to be self-contradictory to suppose that this might happen. So I suppose we must say that it is not absolutely inconceivable that it should happen. But I am strongly inclined to think that it is, for us as we are, practically inconceivable. The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical

> 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 inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question. (118)

Strawson first admits, "It does not seem to be self-contradictory" to suppose we would use our resource at all times. That is, he cannot make the same statistical argument, in this case, that he could for the other cases. Therefore he grudgingly admits that it is not "absolutely inconceivable." But abandoning the reactive attitudes, across the board, would entail eliminating interpersonal relationships as we normally understand them. And that, he thinks, is, "for us as we are, practically inconceivable," something we cannot "take seriously."

It is important to avoid misinterpretation. Strawson is *not* making what would be a very conservative, and implausible, claim: that we are stuck with exactly the set of attitudes and demands we (here in the West? in midcentury England?) currently employ. Strawson is not suggesting that we could not be a little kinder and gentler, or more forgiving, or less judgmental or moralistic. Of course we could; we probably should. Rather, what Strawson finds "practically inconceivable" is the abandonment of reactive attitudes, generally, and the adoption of an objective attitude, in all cases—that is, what he finds practically inconceivable is the abandonment of *any such system* of attitudes and demands, any such system of expectations and reactions.[1]

1. Late in the article, Strawson claims one should be "chary of claiming as essential features of morality in general, forms of these attitudes which may have a local and temporary prominence." He continues, "No doubt to some extent my own descriptions of human attitudes have reflected local and temporary features of our own culture. But an awareness of the variety of forms should not prevent us from acknowledging also that in the absence of *any* forms of these attitudes it is doubtful whether we should have anything that *we* could find intelligible as a system of human relationships, as a human society" (132). (And, it would be remarkably uncharitable,

We can elaborate on his thought: To be engaged in anything like interpersonal relationships is to expect some sort of regard or goodwill from others. Further, violations of those expectations will be met with some sort of negative reaction, some reaction *playing the role* of resentment or indignation. (Likewise, superseding our expectations, or meeting them in difficult circumstances, is met with some positive reaction, something playing the role of gratitude or admiration.) Strawson can and should allow both that the expectations might change in their content and that the corresponding reactions might change in their tone, so to speak. We might (in fact, it seems, we might, in light of our increased understanding of our place in nature) become more gracious and less punitive.[2] A different society might react not with resentment, but with a certain shade of disappointment—a reactive attitude of a different stripe.[3] But to imagine a world in which we respond with only *objective* attitudes is to imagine a world in which any framework *of this kind* is absent, in which nothing plays the *role* of expectation-and-reaction—which is to imagine something alien to us. Indeed, it can seem nearly inevitable that nonsolipsistic, language-using, social creatures sharing a world of limited resources should engage in some such system of expectations and reactions.

One might resist this claim. One might think we could imagine a human society in which there were no expectations of goodwill and no distinctive sort of reaction to failures of those expectations. One might try to imagine a society of people who simply did not care about the quality of others' wills—a society of people fitting the popular conception of psychopathy. I find it doubtful that such

---

I think, to ascribe such little imagination to the author of *Individuals*, which contemplates both the "no-space" world [chap. 2] and the social world in which only groups are recognized [Strawson, *Individuals: An Essay in Descriptive Metaphysics*, 112–15].)

2. Strawson makes a similar point: "This is in no way to deny the possibility and desirability of redirection and modification of our human attitudes in the light of these studies" ("Freedom and Resentment," 133).

3. For an interesting exploration of this possibility, see David Goldman, "Modification of the Reactive Attitudes," *Pacific Philosophical Quarterly* 94, no. 4 (2014): 1–22.

people could in fact form a working society.[4] One might try, instead, to imagine a society in which people recognize and value good will (as, say, something beautiful that sometimes appears in the universe), and perhaps hold it up as an ideal to promote, but do not expect or demand it of others—something like a society of stoics.[5] I find it easier to imagine something along these lines, but, when I do so, it seems I am still imagining a society in which a system of expectations and reactions is in place—just a very subtle, serene one. One might instead suggest a society with expectations, but without *reactions* to violations of those demands or expectations. But recall Strawson's claim that the making of the demand *is* the proneness to the reaction. One might hope for a society with only *positive* reactions—we promote the ideal, but we do not demand conformity to it. I suspect, though, that even the simple absence of a positive reaction, in such a context, could be construed as a negative reaction.

In any case, I think it clear that Strawson thinks that some or another such system of expectations and reactions, demands and attitudes, is, as he puts it, "given with the fact of human society" (131).[6] Support for this interpretative claim can be found not only in "Freedom and Resentment," but also in a lesser-known paper, "Social Morality and Individual Ideal" (published a year earlier).[7]

As suggested by its title, the latter paper examines the tension between, on the one hand, Strawson's own (politically liberal) wish to allow and encourage the flourishing of a multitude of contrasting and even conflicting human ideals—ideals of human life and community one can find lived out in different societies or compellingly portrayed

4. Or, at least, that they could do so without employing a lot of surveillance technology of the sort we are now developing.

5. I am grateful to Sharon Street for pressing this objection.

6. In his landmark book, R. Jay Wallace argues for understanding the reactive attitudes as a narrower class, a class we might abandon without abandoning society. See Wallace, *Responsibility and the Moral Sentiments* (Cambridge, MA: Harvard University Press, 1996), §2.2. This is not Strawson's position (nor does Wallace think it is).

7. Peter F. Strawson, "Social Morality and Individual Ideal," *Philosophy* 36, no. 136 (1961): 1–17. I am grateful to Lucy Allais for drawing my attention to this paper.

in literature and art—and, on the other hand, the need for individuals pursuing these ideals to exist together in a functioning society and so to find a unifying social morality. The tension leads Strawson to what he calls a "minimal" conception of morality—a conception he admits is inadequate and incomplete, but which he insists is also fundamental and useful:

> Now it is a condition of the existence of any social organization, any human community, that certain expectations on the part of its members should be pretty regularly fulfilled; that some duties, one might say, should be performed, some obligations acknowledged, some rules observed. We might begin by locating the sphere of morality here. It is the sphere of observation of rules, such that the observance of some such set of rules is the condition of the existence of society. This is a minimal interpretation of morality. It represents it as what might literally be called a kind of public convenience: of first importance as a condition of everything that matters, but only as a condition of everything that matters, not as something that matters in itself. (5)

Note that Strawson here claims that a minimal morality is "a condition on the existence of any social organization." Moreover, he thinks that the demands of such a system will be "pretty regularly *fulfilled*"—that "the *observance* of some such set of rules is the condition of the existence of society" (emphases added). We have, here, the ingredients for a transcendental argument moving from the existence of society to the satisfaction of the conditions required for it—the typical observance of a minimal set of rules.[8]

I must stress that Strawson recognizes (indeed emphasizes) the extreme poverty of this minimal conception. It is definitely not a conception of ideal human life. Indeed, he contrasts it with such conceptions. Before Strawson is willing even to call it "moral," he insists that it involve reciprocal demands and so, in some way, serve the interests of each. (He stops far short of insisting on equality; this

8. One might compare this to Donald Davidson's arguments about the holism of the mental or W. V. O. Quine's about radical interpretation.

is not even an ideal moral system.) But, interestingly, Strawson also thinks that facts about human nature will ensure that some demands will be reciprocal. He says it is "a fact of human nature which can probably be explained in a number of ways, that quite thoroughgoing egotism [of the kind that would ignore or deny reciprocal demands] is rare" (10).[9]

So, Strawson believes that some or another system of demands and reactions will be given with the fact of society and that a system of *reciprocal* demands will be given with the fact of *human* society. Its absence is, "for us as we are, practically inconceivable."

Once we have in view Strawson's way of thinking of the system of moral and interpersonal expectations and reactions—not as ideals, but as a framework required for, and therefore guaranteed by, the existence of a human society—we can start to see why he thinks that statistics matter: Given the existence of a working society, we know both that some such system is in place and that the minimal expectations and demands of that system are "pretty regularly fulfilled." Now, we can ask, what *sets* those expectations—that is, what determines their particular content? And, what would ensure that they are pretty regularly fulfilled?

The thought, I take it, is that the details of the system will be determined, at least in part, by what is usual or ordinary. If we had a different emotional constitution, we would have different reactions. If we had very different capacities or very different needs, we would, presumably, also have different expectations of one another, and so we would live under a different system of demands. If most of us lacked the capacities required to satisfy certain expectations or demands, those expectations and demands would be

9. I believe this is roughly the same point he makes in "Freedom and Resentment" when he denies wide-ranging moral solipsism (122–23). In his much later reply to Jonathan Bennett, he says, "I freely affirm the central importance of that sense of sympathy, and of a *common* humanity, which underlies not only my indignation on another's behalf but also my indignation on my own." "Replies," in *Philosophical Subjects: Essays Presented to P. F. Strawson*, ed. Zak Van Straaten (Oxford: Clarendon, 1980), 266, emphasis original.

unsustainable—and so would not be part of our system.[10] The system will be attuned to the usual capacities.[11]

But, of course, the range of capacities in the human population is wide, and, at the edges—in the case of small children, for example, or those suffering from dementia or from certain other forms of serious mental illness—we encounter those who lack the capacities required to interact with others in the ordinary way. Rather than continue to make usual demands of them, we can and do, it seems, make exceptions.[12] We shift, to some extent, to a more objective attitude and so, to some extent, lift the associated demands—we overlook and work around instances of what would otherwise be malice, disrespect, or disregard.

Compare what Strawson says about the first sub-variety of exemption, cases such as "he is having a bad day," or "he wasn't himself": After taking seriously the claim that "he wasn't himself" (a claim that *must* be exceptional) Strawson says, "We normally have to deal with him under normal stresses; so we shall not feel towards him, when he acts as he does under abnormal stresses, as we should have felt towards him had he acted as he did under normal stresses" (115). Strikingly, Strawson suggests we sometimes exempt *precisely because* circumstances are not normal; because cases of this kind are unusual,

10. If, on the other hand, we were to gain in our capacities, we might also start to expect more of one another.

11. I have elsewhere suggested morality is more like a hymn than an opera: written for the average congregant, not for the star performer. See my "Reflection and Responsibility," *Philosophy and Public Affairs* 42, no. 1 (2014): 35.

The view I attribute to Strawson is open to an extremely important objection: What if bad behavior is ordinary? Would that render it permissible? I will consider this objection in the final chapter.

12. I suspect Strawson's (now uncivil) references to "the civilized" refer to this fact: "the civilized" drop their demands at the edges, so to speak. (One might compare this to Nietzsche's thought that the strong may show mercy: *On the Genealogy of Morality*, trans. Maudemarie Clarke and Alan J. Swenson [Indianapolis: Hackett, 1998], Treatise II.)

we can overlook them.[13] I believe he would say the same about his second sub-variety of exemption: "We normally have to deal with people of normal capacities; so we shall not feel, towards persons of abnormal capacities, as we would feel towards those of normal capacities." Those who lack the capacities required to fit into the usual system tolerably well, are, for that reason, exempted from it. And so we arrive at Strawson's baffling premise.

Strawson's thought will continue to seem strange and unappealing. To motivate it a bit more, let us consider a society in which, although people are endowed with capacities sufficient to maintain a working society, everyone lacks certain capacities required to satisfy certain of the expectations and demands that we currently impose on one another. Perhaps those suffering from psychopathy could not form a working society, but perhaps active alcoholics could. Currently, I take it, drunkenness can (sometimes) exempt. I have suggested that Strawson thinks that, if so, it exempts in part because it is (relatively) unusual.[14] But now suppose that we all naturally possessed only the

13. Strikingly, the appeal is *not* to fairness toward the offender (and so contrasts sharply with the approach in Wallace). It is about what we, the offended, will feel. The suggestion seems to be that we can simply afford to overlook things once in a while. To help make this plausible, bear in mind that these are not cases in which we simply lower our expectations—cases in which, because of the stressful circumstances, the ordinary expectations would be unreasonable, and so we think that, given the circumstances, no real disrespect or disregard was shown. Rather, these are cases in which even *adjusted* standards have been violated, so that the person really did show disregard or malice. But because of these unusual certain circumstances, we adopt a more objective attitude, saying, e.g., "he wasn't himself." Thanks to Mark C. Johnson for conversation.

14. It is interesting to think about exactly how, and to what extent, drunkenness exempts. It can seem to depend on the circumstances in ways that Strawson would predict: If, while out on the occasional night on the town, your now obviously inebriated roommate says something mean, ill-tempered, or petty, you might simply dismiss it with the thought "he was drunk," much as you would if "he was under great strain" or "he wasn't himself." Such a case falls into Strawson's first subcategory: unusual circumstances. But, as occasional use becomes regular abuse, that attitude will become difficult to maintain. You may have to use your resource "to avoid the strains of involvement." Finally, relations with those we see as active alcoholics

degree of inhibitory control, attention, and memory that we now possess when fairly intoxicated. The system of demands and expectations that would form, in our society, would be sensitive to those limitations. Certain expectations and demands would be unreasonable and unsustainable.

Notice that, in such a society, our relations with one another would, in certain ways, look similar to the relations we now have to those who are drunk: we would not react to certain outbursts or revelations, nor would we expect certain temptations to be avoided or certain occasions or events to be remembered. But there would be this crucial difference: As things now stand, our lack of reaction or expectation amounts to taking up a detached or partially objective attitude; we are exempting the person from the usual expectations and demands. Crucially, in a context in which those particular expectations and demands are simply absent—in which people are simply not expected to avoid those outbursts or remember as much—our lack of reaction would *not* amount to taking up a detached or objective attitude. It would rather be an unremarkable part of ordinary adult relations. So, although no one would be blamed for the outbursts or forgetfulness, that would not be because everyone was exempted from the usual expectations or demands—it would simply be because such outbursts or forgetting would not be disrespectful or wrong.

To put this crucial point a different way: If we had different capacities, we would live under a different system of demands, but a difference in the content of the demands is a difference in what we might call duties, or standards of regard—it is a difference in what *counts* as showing ill-will or disregard; it is not a difference in the conditions on moral responsibility—not a difference in whose ill-will or disregard *matters*. So, if we all had lesser natural capacities, things

---

resemble, in certain ways, our relations with the immature: we hope they will find a way out of their condition, but, as things stand, our relations are partially objective. We have shifted to Strawson's second sub-variety of exemption: this person is not a term in ordinary moral relations, due to disease. The categories here are intriguingly fluid. In the main text I consider *not* widespread active alcoholism, but instead an inherent condition of similarly limited capacities.

would be different—but not because we would uniformly exempt one another from certain moral demands or responsibility. Instead, the system of demands would, itself, adjust to us—it would adjust to what is typical or tolerably ordinary.

Pulling together Strawson's picture: Strawson believes that the existence of a human society requires some or another system of demands and expectations for regard, including reactions to their violation and to their being exceeded. Moreover, we can know, in advance, that certain of these expectations and demands (the minimal ones) will typically be satisfied, and so we can know, in advance, that those to whom those apply will typically have the capacities required to satisfy them. We can now, I hope, start to see why Strawson thinks that nothing true of everyone could provide an exemption. Once we focus not on ideals we might advocate, but rather, as he puts it, on "what it is actually like to be involved in ordinary inter-personal relationships" (113), we are considering a system attuned to the facts on the ground, so to speak.

With this understanding, let us retrace our steps. The initial, seemingly facile argument relied on the claim that we exempt because the person in question is either in extreme or unusual circumstances or else incapacitated for ordinary adult interpersonal relationships—we exempt in the outlier cases. By understanding Strawson's underlying picture, we can at least start to see why he might think this is so. If we grant that we exempt *only* in the outlier cases, it follows that no general thesis could be a reason to exempt—because it cannot be the case that everyone is an outlier. However, Strawson recognized that this initial argument is flawed, in that it overlooks what Strawson himself earlier allowed: we sometimes use our resource to exempt even the normal. As things stand, we do so in only a minority of cases (to avoid the strains of involvement, or as an aid to policy, or out of curiosity, or for therapeutic purposes). Even so, the *reason* we use our resource, in such cases, is not (always) that the person or the circumstance is in some way exceptional. Thus, such cases reveal that we sometimes have reasons to adopt an objective attitude that could, in principle, apply to everyone. And thus Strawson must confront the

remaining question: whether we could, would, or should, in light of the truth of determinism, *make use of our resource* all the time.

In answer to *this* question, Strawson says that, although there is nothing self-contradictory about the hypothesis, and so it is not absolutely inconceivable, our commitment to interpersonal relationships runs too deeply to take it seriously. It is "practically inconceivable." Strawson thus addresses the first and second parts of his remaining question—whether acceptance of the truth of determinism *could* or *would* lead to the decay or repudiation of the reactive attitudes. In some sense it could—it is not absolutely inconceivable—but it would not.

However, it will seem as though Strawson has not yet addressed the final part of his question—whether acceptance of the truth of determinism *should* lead to the decay or repudiation of the reactive attitudes. In particular, it seems as though he has not addressed the question of whether the truth of determinism should lead us to use our resource, repudiate the reactive attitudes, and thus lift all moral and interpersonal demands and expectations, even for the statistically normal.[15] This is not a minor oversight: it seems, in fact, like the crucial question. I will call this "the crucial objection."

15. Returning to the delicate question of whether use of the resource should count as an "exemption" (see chapter 1, footnote 7): When we use our resource only occasionally, we may continue not only to *believe* that the expectations and demands are in place, but also to be *prone* to the reactive attitudes; we thus maintain the expectations and demands, but we choose to ignore their violation. This seems like an "exemption." However, if we were to use our resource across the board, for everyone at all times—if we were to *uniformly* repudiate the reactive attitudes—it would then be difficult to insist that the expectations or demands remain in place; rather, it would seem we had lifted them. Of course, if the demands are not in place, it would be odd to call this a case of "exempting" *from the demands*—the label we introduced again causes difficulty. The important point is that Strawson will see the *universal* use of the resource as indistinguishable from the lifting of all demands. One might say that, while universal *use of the resource* is, strictly speaking, possible (though "practically inconceivable"), universal "exempting" is impossible—because, once suspension of the reactive attitudes is universal, there are no demands from which to

Strawson does, eventually, explicitly raise and address the crucial objection, but he does not do so right away. Instead, at this point, he turns to consider an additional complication. He notes that there is "a further point implicit in the forgoing, which must be made explicit" (118). In fact, he calls this further, implicit point the second part of what he calls his two-part answer. So, before turning to the crucial objection, we will, in the next chapter, consider this further point.

"exempt." But that is an artifact of our label. (See also chapter 2, footnotes 15, 17, and 35.) Thanks to Sarah Buss and an anonymous reviewer for pressing for clarification.

3

The Further, Implicit Point

IN THE FIRST CHAPTER we considered Strawson's initial argument to the conclusion that no general thesis could provide a reason to exempt everyone from moral demands: we do and should exempt in the outlier cases, and it could not be the case that everyone is an outlier. Strawson saw a problem with his argument: we do not adopt the objective attitude *only* in the outlier cases; sometimes use our "resource" in cases that are not outliers. Thus, Strawson took himself to need to address the question of whether acceptance of the truth of determinism would, could, or should lead us to use our resource to exempt everyone.

While Strawson was worried about the use of our resource, the reader was baffled by the initial premise of Strawson's argument: Why does Strawson think that we do and should exempt a person from responsibility because the person or the circumstance is in some way an outlier? Why should statistics matter? In the previous chapter we considered Strawson's initial answer to *his* question, about the resource, and, in so doing, we began to unearth his underlying picture of the nature of morality and so began to address our own bafflement.

We saw that Strawson believes that some or another framework of moral and interpersonal demands and expectations will be given with the fact of human society. We surmised that the content of those demands and expectations must be set, in part, by the actual abilities of those in the society. Thus it is that statistics matter: Actual abilities help to determine the content of the demands and expectations, and so help to determine what counts as disregard or ill will.

If it were the case that everyone lacked some natural ability, then failing to exercise that ability would not count as showing disregard. Likewise, if satisfying certain demands is generally required for the existence of human society, then the capacity to satisfy such demands will be typical in human society.[1] In certain extreme or unusual circumstances, or in those cases in which a person lacks the usual capacities, we can and do simply suspend the usual expectations and demands and thereby exempt the person. We exempt in the outlier cases.

This underlying picture tempts a quick argument to the conclusion that no general thesis could be a reason to exempt—the argument Strawson makes. However, Strawson also sees that his quick argument will be undermined by his own observation that we sometimes use our "resource" to adopt an objective attitude toward those of normal capacities in circumstances that are not unusual. Thus, he must address the question of whether, in light of the truth of determinism, we could, would, or should use our resource to exempt everyone at all times.

He answers this remaining question by appeal to "our commitment to ordinary interpersonal relationships." Given this commitment, he says, abandoning ordinary interpersonal relations, across the board, is "practically inconceivable" (118), not something we can take seriously.

As we noted in the previous chapter, this claim seems not to provide a complete answer to the question Strawson needs to address: it answers the first two parts—though we *could* abandon ordinary moral and interpersonal relationships, we *would* not—but it seems to leave untouched the question of whether, given the truth of determinism, we *should* do so. I called this the crucial objection. Strawson eventually raises and addresses it. However, before doing so, he turns to what he calls a "further point, implicit in the foregoing, which must be made explicit" (118). This chapter considers Strawson's further point. The next will consider the crucial objection.

1. These claims about the relevance of actual ability require elaboration and defense. Some will be provided in the final chapter.

Returning to Strawson's text: Immediately after making the point that it is "practically inconceivable" that we would give up on ordinary interpersonal relationships, Strawson says,

> This, then, is a part of the reply to our question. A sustained objectivity of the inter-personal attitude, and the human isolation which that would entail, does not seem to be something of which human beings would be capable, even if some general truth were a theoretical ground for it. But this is not all. There is a further point, implicit in the foregoing, which must be made explicit. (113)

Looking ahead, we learn the further point is that we never shift to a more objective attitude because we believe the person's behavior was determined in some sense (forced, caused, fated, out of their control) in which all behavior is determined, if determinism is true. As Strawson puts it, "when we do in fact adopt such an [objective] attitude in a given case, our doing so is not the consequence of a theoretical conviction which might be expressed as 'Determinism in this case'" (120).

This is a strong and striking claim. If true, it would undercut one of the standard, and one of the strongest, arguments for incompatibilism, sometimes called "the generalization strategy."[2] I believe Strawson pauses to make this further claim explicit with an eye to the generalization strategy.[3]

2. I take the term "generalization strategy" from Wallace, *Responsibility and the Moral Sentiments*, 115.

3. One might be tempted to read Strawson's formulation, "our doing so is not the consequence of a theoretical conviction which might be expressed as 'Determinism in this case'" (120), not as making the strong claim that we do not exempt for any reason that would generalize if determinism is true, but rather as making the much weaker claim that, when we exempt, we do not have in mind the scientific thesis of determinism. There are two strong reasons to resist this interpretation. First, so interpreted, the claim is narrower than what follows from the argument Strawson gives. And, second, Strawson means, throughout, to argue without presupposing any specific understanding of the thesis of determinism.

The Generalization Strategy

Someone employing the generalization strategy argues as follows: Sometimes we exempt a person from moral responsibility because we learn that their behavior was caused by forces beyond their control—because, for example, they are subject to an illness or disorder that renders them unable to control their impulses. In such a case, it seems we exempt the person because they were made to do what they did by facts about their physiology. But, if determinism is true, then each of us is *always* made to do what we do by such facts. Therefore, contrary to Strawson's further point, we *do* sometimes shift to an objective attitude because a person's behavior was determined in a way in which all behavior would be, if determinism is true. And thus, if determinism is true, we will have reason to exempt everyone—the very same reason at work in these familiar cases (the reason will "generalize," given the truth of determinism). Therefore, if we learn that determinism is true, then whether or not we would or could, we certainly *should* exempt everyone—we should do so for the same reason for which we now exempt certain cases of disease. The generalization strategy can thus motivate the pessimist's position: if determinism is true, everyone should be exempt for the same reason that we now exempt certain people.

Strawson's further point would undercut the generalization strategy by denying its main premise: the further point claims that we never, in fact, shift to the objective attitudes because a person's behavior was determined in some way in which all behavior would be if determinism is true. Rather, Strawson claims, we exempt because the person's behavior was determined in some way that renders them incapable of ordinary interpersonal relationships. If Strawson is correct about this, there is no reason that will generalize, given the truth of determinism—because nothing could show that we are all incapable of ordinary interpersonal relationships.

Before examining how Strawson makes explicit his further point, let us note that, given Strawson's distinction between (what we called) excuse and exemption, the generalization strategist must choose their cases carefully. If we learn that your body was used like a puppet, for example, then you are, in the terms we have introduced, excused, not

exempted: we have learned that we were mistaken about the quality of your will, and therefore we must modify or suspend any reactive attitude that were based on that mistake, but we need not regard you as someone whose will does not matter—we need not adopt an objective attitude.[4] Likewise if you were held at gunpoint. In fact, many of the typical cases in which someone was "made to do something" or "had no choice" are cases of excuse, not exemption. As Strawson notes, if such cases were universal, that would be a consequence not of universal determinism, but of universal goodwill. Thus, these are not candidates for the generalization strategy. The generalization strategist must instead find cases in which someone is "made to do something," "had no choice," or "lacked control" in a way that renders the person no longer an apt target of the reactive attitudes, at all. Certain forms of mental illness or incapacity seem to provide such cases.

Making Explicit the Further Point

How, then, does Strawson "make explicit" this further point? How does he argue from his claim that "sustained objectivity of the interpersonal attitude . . . [is not] something of which human beings would be capable" (118) to the conclusion that we never in fact exempt anyone because their behavior is determined in some way in which all behavior is, if determinism is true?

He first notes what he calls a "connection," saying, "In the extreme case of the mentally deranged, it is easy to see the connection between the possibility of a wholly objective attitude and the impossibility of what we understand by ordinary inter-personal relationships" (119). This "connection" is, presumably, found in, implicit in, the fact that sustained objectivity is not something of which human beings are capable: We are not capable of relating objectively all the time, and so (as we saw in the previous chapter) the framework of expectations and demands that constitutes what we understand as engaged, nonobjective interpersonal relating will adjust to whatever

4. See chapter 1, footnotes 5 and 6.

is ordinary. Still, in unusual cases, when such ordinary relating is not possible, we might relate in an objective way. Strawson appeals to this connection to support his thought that when we exempt those subject to, for example, an impulse control disorder, we do so because it renders the person incapable of ordinary interpersonal relating (at least in a certain range of circumstances).

But, Strawson will then point out, ordinary adult relating is *actual*—it is, as he puts it, "among the facts as we know them." Thus, we already know that many people are, in fact, capable of ordinary adult relating. But, if determinism is true, it is already true—if determinism is true, then everyone's behavior is already determined in whatever way the thesis of determinism specifies. And so we know that, if determinism is true, then being determined in whatever way *it* specifies does not render us incapable of ordinary adult relating. To think otherwise would flatly contradict the facts as we know them. And so Strawson concludes that, if determinism is true, then, when we exempt those who are incapable of ordinary relating, we do not do so because the behavior in question was determined in whatever sense is at issue in that thesis. This is the further point, which Strawson sees as implicit in the claim that sustained objectivity is not something of which we are capable, now made explicit.

Here is how Strawson makes the point, in his text: He first reminds us of the possibility of objective relating, "Exceptionally, I have said, we can have direct dealings with human beings without any degree of personal involvement, treating them simply as creatures to be handled in our own interests, or our side's, or society's—or even theirs" (118–19). He then notes the "connection" between the possibility of objective relating and the impossibility of ordinary relating:

> In the extreme case of the mentally deranged, it is easy to see the connection between the possibility of a wholly objective attitude and the impossibility of what we understand by ordinary inter-personal relationships. Given this latter impossibility, no other civilized attitude is available. . . . [T]o view him as outside the reach of personal relationships is already, for the civilized, to view him [in] this [objective] way. For reasons of policy or

> self-protection we may have occasion, perhaps temporary, to adopt a fundamentally similar attitude towards a "normal" human being. (119)

Having noted this connection, and having also reminded us of the possibility of using our resource for reasons of "policy or self-protection," he launches into an argument:

> Now it is certainly true that in the case of the abnormal, though not in the case of the normal, our adoption of the objective attitude is a consequence of our viewing the agent as *incapacitated* in some or all respects for ordinary interpersonal relationships. But, there is something else which, *because* this is true, is equally certainly *not* true. And that is that there is a sense of "determined" such that (1) if determinism is true, all behavior is determined in this sense, and (2) determinism might be true, i.e., it is not inconsistent with the facts as we know them to suppose all behavior is determined in this sense, and (3) our adoption of the objective attitude toward the abnormal is the result of a prior embracing of the belief that the behavior, or the relevant stretch of behavior, of the human being in question *is* determined in this sense. (119)

Strawson sees a contradiction (this is the second point at which he argues by appeal to a contradiction) in asserting all four of the following:

- P0: "In the case of the abnormal . . . our adoption of the objective attitude is a consequence of our viewing the agent as *incapacitated* in some or all respects for ordinary interpersonal relationships."
- P1: If determinism is true, then all behavior is determined* (determined in the sense that all behavior is, if determinism is true).
- P2: Determinism might be true; it is not inconsistent with the facts as we know them.
- P3: "Our adoption of the objective attitude toward the abnormal is the result of a prior embracing of the belief that the behavior, or the relevant stretch of behavior, of the human being in question" is determined*.

To rehearse the apparent contradiction once more: The "ordinary interpersonal relationships" in P0 are those that we have been engaged in all along—those we know and love, and so know to be actual. Thus, that some people are capable of these relationships is among "the facts as we know them." Thus, it could not be that, when we regard certain people as incapacitated, it is because we regard them as subject to some shortcoming to which everyone might (already) be subject, consistent with what we know. Because, again, we already know that some people enjoy the capacities required. Thus, at least "in the case of the abnormal," we never adopt the objective attitude because we believe a person's behavior is "determined" in a sense that all behavior could be, consistent with what we know. Rather, when we adopt the objective attitude, we do so because we see their behavior as forced, or caused, or stemming from forces outside their control *in a way* or *to a degree* that renders tolerably ordinary relating impossible. In light of the contradiction, Strawson rejects P3. If determinism is true, then being determined* does not render us incapacitated for ordinary interpersonal relationships. And, since we exempt because people are thus incapacitated (P0), we do not exempt them because their behavior is determined* (P3).

Objections

The opponent will be unimpressed. Most glaringly, while Strawson rejects P3, the opponent will simply reject P0 (the claim that, when we exempt an outlier, we do so because we see the person as incapacitated for ordinary relating).

Strawson seemed to suggest that P0 is implicit in the fact that we are incapable of sustained objectivity of attitude—that fact was to connect the possibility of adopting a wholly objective attitude with the impossibility of ordinary relating, and that connection was meant to support the claim that we exempt an outlier because we believe they are incapacitated for ordinary relating (P0). However, once we grant P0, Strawson argues that we need to deny P3.

In response, the opponent will raise a version of the crucial objection: Strawson once again is moving illicitly from facts about what we could or would do to conclusions about what we should

do—to conclusions about our reasons for exempting. The opponent will not allow it.

Thus, for the opponent, Po receives no support from Strawson's claim about our capacities for sustained objectivity; for the opponent, Po must rely for support only on the strength of Strawson's quick survey and taxonomy of cases of exemption. The opponent will interpret those cases differently and so will reject Po in favor of some other principle, perhaps a principle about control (we rightly exempt those who lack control) or about alternative possibilities (we rightly exempt those who lack alternatives). These other principles might be universally violated, if determinism is true.

Notice, though, that the argument has revealed something both interesting and important: The underlying disagreement turns on whether, when we adopt the objective attitudes in the case of outliers, we do so because ordinary relating has become *impossible,* as Strawson insists, or whether, instead, engaging in ordinary relationships is subject to some other kind of standard, some standard that even current, actual, working relationships might fail to meet, rendering them *illicit*.

Notice, too, that even acknowledging the distinction between conditions that render ordinary relating impossible and standards that show it illicit will present a challenge for the opponent. We have already seen that, due to the difference between excuse and exemption, the opponent must rely on cases such as mental illness (and not cases in which we were simply mistaken about the quality of will) to serve as evidence for a standard that will be universally violated if determinism is true. However, once we acknowledge this new distinction, such cases do not clearly provide evidence of the kind the opponent needs: these are cases in which ordinary relating is not merely inappropriate or illicit, but is (also) impossible. (If the opponent were instead to appeal to control, claiming that such cases such as mental illness reveal that we need to be in control and that the truth of determinism would show that we do not really or ultimately control anything, Strawson could again reply that, in the cases presented, people lack the *kind* or *degree* of control required for tolerably ordinary relating, which we evidently possess. And so on.) Much better evidence, for the opponent, would be

supplied by cases in which it would be plainly possible to continue relating in the ordinary way, but in which it is illicit to do so, because the person's behavior is determined in some sense (because they "could not have done otherwise" or "had no choice" or things were "out of their control"). Such cases are hard to find. In fact, it can seem disrespectful always to regard objectively someone capable of ordinary relating.[5]

Still, the opponent may yet insist that there are standards on the appropriateness of ordinary interpersonal relating that could be universally violated (standards of control or of alternative possibilities), one of which would be violated if determinism is true, requiring us to give up interpersonal relating. I believe Strawson rejects the possibility of such standards, and this rejection, like P0, rests on his underlying, naturalistic picture of morality. We will return to this in the next chapter.

However, before doing so, we should consider two further objections to Strawson's argument to the further point.

5. It can seem disrespectful, though sometimes we might feel the need to do so: consider someone who, we think, has the capacities required for ordinary relating, but who is, say, too much of a jackass to continue to take seriously. We use our resource to "avoid the strains of involvement." Is failing to use our resource, and maintaining the usual demands, *illicit*? I do not think so; it is just exhausting. Is using our resource disrespectful? That seems a complicated question—but, insofar as it involves distancing ourselves from someone capable of tolerably ordinary relating, it has the feel of disrespect. (I will later make a distinction between natural capacities and socially developed capacities. The jackass may, after all, lack the capacity for tolerably ordinary relating due to a failure of social development. But, as we will see in the final chapter, though that may render him incapable of tolerably ordinary relating, it may not exempt. See chapter 5, footnotes 3 and 20.)

Cheating in a game would provide the kind of example the opponent needs: If you cheat by, e.g., using performance-enhancing drugs, ordinary play is still possible, but it is also illicit. It seems to be in principle possible to learn that everyone has always been cheating (in a way that it is not possible to learn that no one has ever been capable of ordinary play). I take up this kind of example in the final chapter. See chapter 5, footnote 46.

First, the argument reaches its conclusion by claiming that P0–P3 are inconsistent. But, strictly speaking, they are not. P0 claims that we exempt because we *believe* that the person is incapacitated for ordinary interpersonal relationships. P3 claims that, when we exempt, we do so because we *believe* they are determined.* Adding, to these, P1 and P2 (adding that everyone might already be determined*) would show that we believe a collection of things that cannot all be true. But we sometimes do fall into contradiction. And it seems our actual moral and interpersonal practices might have, embedded in them, *this* contradiction—it is, after all, the ambition of the generalization strategy to show our practices incompatible with the truth of determinism. So, given our fallibility, perhaps we do sometimes exempt because a person's behavior was determined*, despite thinking both that we exempt outliers and that determinism might be true of everyone. It seems that Strawson needs another premise: that our reasons for exempting will not drive us into contradiction with what we know. We will return to this issue in the final chapter.[6]

A second problem with the argument remains: Even if we grant Strawson both P0 and the extra premise, the argument would not establish the further point as Strawson has stated it—as he is well aware. The argument from P0–P3 only reaches the conclusion that "*in the case of the abnormal*" we do not adopt the objective attitude because we believe the behavior in question is determined*. Nonetheless, having reached this partial conclusion, Strawson immediately asserts, "Neither in the case of the normal, then, nor in the case of the abnormal is it true that, when we adopt an objective attitude, we do so *because* we hold such a belief" (119). This is surprising. How can he simply draw the same conclusion, in the case of normal?

6. See chapter 5, footnote 40. Also: Why not present the argument, not as about what we *believe* to be reasons, but rather as about what reasons there *are*, so to speak? Once we grant (what we might call P0*) that the reason to exempt the abnormal is *the fact that* they are incapacitated for ordinary interpersonal relationships, then it follows that determinism, if true, would not provide this reason for exempting everyone—because, as Strawson put it, "it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition." Note, though, Strawson *did* present this argument: it is his original, seemingly facile one.

Strawson's silence would be bizarre if the purpose of making explicit the further point were to address the question of whether, given the truth of determinism, we should use our resource to exempt even the normal. We should therefore conclude that this is not Strawson's purpose. He is not, here, addressing the crucial objection (as we have seen, the argument he gives re-raises the crucial objection in a new guise). I have suggested that his aim is instead to show that, once we grant Po, we have undercut the generalization strategy. That strategy is not well motivated by typical cases in which we use our resource: we do not use our resource for reasons that anyone would be tempted to think are similar to the thesis of determinism. Rather, as Strawson reminded us early in the current passage, we use our resource for "reasons of policy or self-protection," for reasons of therapy or curiosity or exhaustion, or avoiding "the strains of involvement." Perhaps we do so on a whim. We do not use our resource because we think the person "could not help himself," or because his behavior was "out of his control," or "forced upon him" or because he "had no choice."[7] And thus, having dealt with the abnormal case, Strawson feels he can simply draw the conclusion that neither in that case nor in the case of the normal do we shift to a more objective attitude because we believe a person's behavior is determined*.[8]

7. It is true that we might bring ourselves to adopt the objective attitude, in the case of the normal, by focusing on, as Strawson puts it, "how he works" (119). But, though this is our method, our reasons for undertaking this method are reasons of policy or self-protection.

8. The opponent may raise a third objection: When Strawson first admitted that his initial argument was facile, he seemed to allow the opponent to group any overlooked reasons for repudiating the reactive attitudes (any reason that could be apply to the statistically normal) as a "use of the resource." (He said, "whatever is too quickly dismissed in this dismissal is allowed for in the only possible form of affirmative answer that remains" (118).) But, at this point in the argument, Strawson is being more strict about what might count as use of the resource. He thus seems to be equivocating in a way that rules out principles that could generalize (about control or alternative possibilities).

I believe Strawson is, in fact, doing exactly this. It would be a problem, dialectically, if the argument to the further point were meant to rule out these alternative

The opponent will remain unsatisfied. They will remain convinced that Strawson has overlooked some further condition on the appropriateness of moral responsibility, one that is neither captured by P0 nor well understood as a use of the resource (though, as noted above, Strawson's distinctions will make it difficult for the opponent to support such a condition by appeal to cases). And thus the opponent will remain convinced that Strawson has not yet adequately answered the question he set for himself: he has not shown that acceptance of the truth of determinism *should not* lead to the decay or repudiation of the reactive attitudes. Again, the opponent might grant that the reactive attitudes will not decay—that, given our nature, abandoning them will be impossible—but, the opponent will continue to think, we might, at least, *repudiate* them—and, certainly, we *should* do so, if determinism is true. We are, at last, ready to consider the crucial objection. In doing so, we will better understand why Strawson thinks there are no further conditions—why there are no standards on the appropriateness of ordinary interpersonal relating, but only conditions of its possibility.

---

principles. But it is not. As we will see in the next chapter, Strawson sees those principles as already ruled out by his underlying picture of morality. The argument for the further point assumes this. The argument from P0–P3 simply makes plain that, if determinism is true, then being determined* does not render us incapacitated for ordinary interpersonal relationships. Strawson can then conclude that, *on his picture*, there is no reason that will generalize.

4

Addressing the Crucial Objection

A QUICK REVIEW: In the first chapter we considered Strawson's initial argument to the conclusion that the acceptance of a general thesis neither would, could, nor should lead us to exempt everyone. We do and should exempt in the outlier cases, but it could not be that every case is an outlier. In the second chapter we considered Strawson's attempt to address what he called the "the only possibility that remains": we do not exempt *only* in the outlier cases; we sometimes also use our "resource." Perhaps, then, we could, would, or should use our resource at all times? Strawson answered by claiming that our commitment to interpersonal relationships renders using the resource at all times "practically inconceivable": though, strictly speaking, we *could* do it, we would not. Strawson's answer seems not to fully address his question: it seems to leave untouched whether we *should* use our resource at all times. I called this "the crucial objection."

Instead of addressing the crucial objection, Strawson turned to make explicit a further point: he moved from the claim that we are not capable of suspending the reactive attitudes at all times to the claim that we do so, in the case of the outliers, because ordinary relating is impossible. However, since ordinary relating is actual, and since determinism, if true, is already true, we know that, if determinism is true, then being determined in whatever sense it specifies (being "determined*") does not render ordinary relating impossible. Thus, Strawson concludes, it is not among the reasons for which we exempt outliers. Strawson took it to be obvious that we do not use our

resource for this reason either, and so he concluded that we never adopt an objective attitude because a person is determined*.

In making this point explicit, it will seem to the opponent that Strawson has once again overlooked the possibility of other kinds of conditions on the appropriateness of ordinary relating (e.g., conditions about control or about alternative possibilities or fair opportunity). The opponent will be frustrated.[1]

Continuing, then, in Strawson's text: Immediately after making explicit the further point, Strawson seems to think he has completed his argument. He provides a summary:

> So my answer has two parts. The first is that we cannot, as we are, seriously envisage ourselves adopting a thoroughgoing objectivity of attitude to others as a result of theoretical conviction of the truth of determinism; and the second is that when we do in fact adopt such an attitude, in a particular case, our doing so is not a consequence of a theoretical conviction which might be expressed as "Determinism in this case," but is a consequence of our abandoning, for different reasons in different cases, the ordinary interpersonal attitudes. (119–20)[2]

It is difficult to understand why Strawson thinks he is finished. Again, the first part of the two-part answer shows that we *would not* give up

1. We also saw that Strawson's distinctions will make it difficult for the opponent to find evidence to support an alternative condition: cases of exemption (rather than of excuse or use of the resource) do, in fact, seem to be ones in which tolerably ordinary relating has become not just inappropriate, but unworkable. It would have been useful for Strawson to make *this* point more explicit.

2. The "different reasons in different cases" refers to a list he provided, a few sentences earlier, of reasons for which one might be incapacitated for ordinary interpersonal relationships: "He is thus incapacitated, perhaps, by the fact that his picture of the world is a pure fantasy, that he does not, in a sense, live in the real world at all; or by the fact that his behavior is, in part, an unrealistic acting out of unconscious purposes; or by the fact that he is an idiot, or a moral idiot" (119). I have been focusing on the case of an impulse control disorder, as I think it a better case for the opponent. Strawson's text supports the reading I have given: these different reasons exempt insofar as they render the person incapacitated for ordinary interpersonal relationships.

ordinary relating, but seems not to address the question of whether we *should*. The second part of the answer (the further point) claims we never, in fact, exempt people because their behavior was determined in the way in which all behavior would be, if determinism is true. If we were to grant the further point, we would thereby grant that there is not, *in our present practices*, a reason for which we exempt that would, if determinism is true, generalize to exempt everyone. But, one might think, merely showing that something is not *now used* as a reason does not show that it *is not* a reason. For that, one might think, you need an argument. Thus, one might conclude that, even granting the further point, Strawson has not yet shown that the truth of determinism *should not* lead to the decay or repudiation of the reactive attitudes

So the opponent is again baffled, impatient to raise the crucial objection: How can facts about what we now, *in fact*, do, or about what we *would*, in fact, do, lead us to conclusions about what we *ought* to do—what we *should* do, if we learn that determinism is true.

Immediately after summarizing his two-part argument, Strawson explicitly considers this objection:

> It might be said that all this leaves the real question unanswered, and that we cannot hope to answer it without knowing exactly what the thesis of determinism is. For the real question is not a question about what we actually do, or why we do it. It is not even a question about what we would *in fact* do if a certain theoretical conviction gained general acceptance. It is a question about what it would be *rational* to do if determinism were true, a question about the rational justification of ordinary interpersonal attitudes in general. (120)

Strawson then provides a remarkably unhelpful reply, in two parts:

> To this I shall reply, first, that such a question could seem real only to one who had utterly failed to grasp the purport of the preceding answer, the fact of our natural human commitment to ordinary interpersonal attitudes. This commitment is part of the general framework of human life, not something that can come up for review as particular cases can come up for review within this general

> framework. And I shall reply, second, that if we could imagine what we cannot have, viz., a choice in the matter, then we could choose rationally only in the light of an assessment of the gains and losses to human life, its enrichment or impoverishment; and the truth or falsity of the general thesis of determinism would not bear on the rationality of *this* choice. (120)

Apparently, we need to examine even more closely Strawson's thought about "the fact of our natural human commitment to ordinary interpersonal attitudes," whose full purport we have yet failed to grasp. In particular, we need to understand why he would think our commitment cannot "come up for review" in the way that particular cases can come up for review "within this general framework." We also need to understand how this claim about what can and cannot come up for review meets the crucial objection—an objection that was not put forward as questioning a "framework," nor as seeking an "external" justification, but which simply suggested that there might yet be a reason to exempt even the normal. And, finally, we need to understand why choosing our commitments rationally (if we could) would require choosing "only in the light of an assessment of the gains and losses to human life."

Unearthing Strawson's Naturalism

To begin, note the use of the term "natural." That word carries a significance in Strawson's mouth one might not expect. We find a clue in a footnote late in the article, when he is again discussing the framework that cannot come up for review. Here is the text to which the footnote is attached:

> Inside the general structure or web of human attitudes and feelings of which I have been speaking, there is endless room for modification, redirection, criticism, and justification. But questions of justification are internal to the structure or relate to modifications internal to it. The existence of the general framework of attitudes is itself something we are given with the fact of human society. As a whole, it neither calls for nor permits an external "rational"

> justification. Pessimist and optimist alike show themselves . . . unable to accept this. (131)

And here is the footnote:

> Compare the question to the justification of induction. The human commitment to inductive belief-formation is original, natural, non-rational (not *ir*rational), in no way something we could choose or give up. Yet rational criticism and reflection can refine standards and their application, supply "rules for judging cause and effect." Ever since these facts were made clear by Hume, people have been resisting acceptance of them. (131)

Strawson seems to be thinking that, because it is "original, natural, non-rational . . . in no way something we could choose or give up," our commitment to inductive belief-formation is not something we could justify, but also not something that needs to be justified. This is so, even though the formation of *particular* beliefs is subject to justification (they can come up for review, within the general framework), and, indeed, even though the practice can be improved through rational reflection—supplying "rules for judging cause and effect."

Likewise, for Strawson, our commitment to participation in interpersonal relationships, and so to *some or another* framework of demands-and-attitudes, is also original, natural, nonrational (not *ir*rational), not something we could either choose or give up—that is, he seems to think, not something for which there are reasons, not something we could justify, but also not something that needs to be justified. Further, our engagement in such relationships provides a framework within which there are reasons for particular actions or attitudes on particular occasions—reasons to resent or to admire, to excuse or to exempt. Moreover, and importantly, rational criticism and reflection can refine the standards and improve the practice—supplying rules of justice, say. It may be that we should all be kinder and gentler. It may even be that the particulars of our practices should change dramatically. But, Strawson thinks, the framework, itself—that is, the fact that we engage in *some such* system—neither calls for nor permits justification.

Why does Strawson think that such frameworks need no justification? He presents a case for inductive belief formation in the ninth chapter of an early work, *Introduction to Logical Theory*.[3] I recommend it to the reader. Here is one long passage:

> So to ask whether it is reasonable to place reliance on inductive procedures is like asking whether it is reasonable to proportion the degree of one's convictions to the strength of evidence. Doing this is what "being reasonable" *means* in such a context.
>
> . . . It is generally proper to inquire *of a particular belief*, whether its adoption is justified . . . we are [then] appealing to, and applying, inductive standards. But to what standards are we appealing, when we ask whether the application of inductive standards is justified or well-grounded? If we cannot answer, then no sense has been given to the question. Compare it with the question: Is the law legal? It makes perfectly good sense to inquire of a particular action, of an administrative regulation, or even, in the case of some states, of a particular enactment of the legislature, whether or not it is legal. The question is answered by an appeal to a legal system, by the application of a set of legal (or constitutional) rules or standards. But it makes no sense to inquire in general whether the law of the land, the legal system as a whole, is or is not legal. For to what legal standards are we appealing?
>
> The only way in which a sense might be given to the question whether induction is in general a justified or justifiable procedure, is a trivial one which we have already noticed. We might interpret it to mean, "Are all conclusions, arrived at inductively, justified?" i.e., "Do people always have adequate evidence for the conclusions they draw?" The answer to this question is easy, but uninteresting: it is that sometimes people have adequate evidence, sometimes they do not. (257)[4]

3. Peter F. Strawson, *Introduction to Logical Theory* (London: Methuen, 1952), chap. 9. I am grateful to Roger White for bringing it to my attention.

4. (We would all be excused under the reign of universal good will, and inductive belief formation would be justified, generally, under the reign of universal good reasoning.)

This passage contains what I called, in the introduction, the broadly Wittgensteinian thought: it makes no sense to question or to criticize a practice, as a whole, in terms that owe their meaning to that practice.[5]

Note, though, that even if we grant this thought, we do not yet arrive at Strawson's stronger claim in "Freedom and Resentment": that certain of our commitments are natural and nonrational in a way that precludes *any* request for justification. We can, after all, seek an *extralegal* justification of the legal system: we can ask whether we would simply be better off without it—in the language of the crucial objection, we can ask about "the rational justification of the [the legal system] in general," and we can answer by appeal to the gains and losses to human life. Thus, while the broadly Wittgensteinian thought might address an opponent who asks whether it would be *moral* (or just, or fair) to continue in our moral practices, that is not the opponent Strawson gives to himself. Rather, the crucial objection, as Strawson poses it, asks what it would be *rational* to do, if determinism is true. This *extramoral* question can be asked even if one grants the broadly Wittgensteinian thought—and it could be answered, as the optimist answers it, by appeal to the gains and losses to human life.[6] Strawson clearly thinks such a question is not "real," something we "cannot have." It is not yet clear why he thinks this.

At this point it will be helpful to consider a much later work, Strawson's Woodbridge Lectures from 1983, published in 1985 as

5. There is much that could be said about this thought. I leave it for another occasion.

6. It would be what Rudolf Carnap called an "external" question about the practice, and the broadly Wittgensteinian thought, as presented, seems not to rule out such questions. On "external" questions, see Carnap, "Empiricism, Semantics, and Ontology," *Revue Internationale de Philosophie* 4 (1950): 20–40. See also Strawson's discussion of Carnap in Strawson, *Skepticism and Naturalism*, 6ff., and "Carnap's Views on Constructed Systems versus Natural Languages in Analytic Philosophy," in *The Philosophy of Rudolf Carnap*, Library of Living Philosophers, ed. Paul A. Schilpp (LaSalle, IL: Open Court, 1963).

Skepticism and Naturalism.[7] In the first chapter he considers a response to skepticism, which he identifies with both Hume and Wittgenstein, that he initially calls "Naturalism," and later calls "social naturalism."

Strawson starts by characterizing two "Humes," a skeptic and a naturalist. Hume the Skeptic raises the problem with induction: we have no independent grounds on which to justify it. According to Strawson, Hume the Naturalist responds to Hume the Skeptic by, so to speak, simply walking out of the study: "Skeptical doubts are not to be met by argument. They are simply to be neglected . . . because they are *idle*; powerless against the force of nature, of our naturally implanted disposition to belief" (13).

By giving, as the *reason* for the Naturalist's neglect, the fact that the doubts are "powerless against . . . our naturally implanted disposition," Strawson here seems to be giving what I called in the Introduction the simple Humean response: Our psychologies ensure we will not change, and so we should just stop talking about the skeptical doubt. The question is "unreal" and something "we cannot have," simply because of our psychological limitations.[8]

7. *Skepticism and Naturalism.* Strawson's views were developed over a long time. Many think they changed very significantly from *Individuals* to *Skepticism and Naturalism*, especially in response to pressure from Barry Stroud's "Transcendental Arguments," *Journal of Philosophy* 65, no. 9 (1968): 241–56. I doubt they changed so much. For discussion, see John J. Callanan, "Making Sense of Doubt: Strawson's Anti-Scepticism," *Theoria* 77, no. 3 (September 2011): 261–78.

8. Strawson sounds most like the simple Humean in his second statement of the argument in "Freedom and Resentment": "Finally, to the further question whether it would not be *rational*, given a general theoretical conviction of the truth of determinism, so to change our world that in it all these attitudes were wholly suspended, I must answer, as before, that one who presses this question has wholly failed to grasp the import of the preceding answer, the nature of the human commitment that is here involved: it is *useless* to ask whether it would not be rational for us to do what it is not in our nature to (be able to) do" (126). He seems to deny a simple Humean approach in his "Reply to Ernest Sosa": "It is not merely a matter of dismissing the demand for justification of one's belief . . . on the ground that one can't help believing it. That would be very weak indeed" (370). In *The Philosophy of P. F.*

But the simple Humean thought is unsatisfying, for both textual and philosophical reasons.

We noted the first textual problem in the Introduction: If we interpret Strawson as making the simple Humean point, we should not expect him to charge his opponent with *incoherence*. Yet he does, twice.

A second, related textual problem is this: The simple Humean interpretation renders most of Strawson's article unnecessary. If his response is the simple Humean one, why did he bother to lay out the distinction between reactive and objective attitudes or the distinction between excuse and exemption? Why give the seemingly facile argument, narrow our attention to cases of using our resource, or make explicit the further point? The simple Humean response recommends a far shorter article (or, perhaps, no article at all—get out and enjoy your day).

Finally, Strawson claims that, by raising the crucial objection, his opponent "utterly failed to understand the purport of the preceding answer." But, if the preceding answer were the simple Humean one, then, in raising the crucial objection, the objector would not be *misunderstanding*; the objector would be straightforwardly objecting (what we in fact do, or what we could do, the objector says, is not to the point).

The simple Humean response also seems philosophically unsatisfying. It seems an invocation of something like the following general principle: if someone lacks the ability to do or change something, then their continuing in the status quo neither calls for nor permits justification. But that principle seems, at the very best, highly contentious.

Strawson is not invoking such a general principle. Instead, he sees Hume pointing to a small number of "framework" commitments, and it is only these to which questions of justification do not apply. Once "inside" a framework, questions of justification have free rein—perhaps even in cases in which we lack the ability to change.

Strawson, Library of Living Philosophers, ed. Lewis Edwin Hahn (LaSalle, IL: Court Press, 1998).

Whether a lack of ability to change forecloses questions of justification will usually be a question to settle within the framework (and the answer, presumably, will be, sometimes yes, sometimes no). Strawson makes the point in a passage that recalls our footnote:

> This does not mean that Reason has no part to play in relation to our beliefs concerning matters of fact and existence. It has a part to play, though a subordinate one. . . . (Here we may recall and adapt that famous remark about Reason and the passions). Our inescapable natural commitment is to a general frame of belief and to a general style (the inductive style) of belief-formation. But within that frame and style, the requirement of Reason, that our beliefs should form a consistent and coherent system, may be given full play. Thus . . . Hume could quite consistently proceed to frame "rules for judging cause and effect." Though it is Nature which commits us to inductive belief-formation in general, it is Reason which leads us to refine and elaborate our inductive canons and procedures and, in their light, to criticize, and sometimes to reject, what in detail we find ourselves naturally inclined to believe. (13–14)

Reflection "inside" the framework can lead us to reject our natural inclinations—and, presumably, to minimize, contain, and work around those we cannot change. The simple Humean response is too simple.

If, following Strawson, we grant Hume the Naturalist's claim that (only) certain "framework" commitments are immune to questions of justification, we can next wonder *which* commitments belong to the framework and *why* those, and only those, need no justification.[9]

Taking the first question first: For Hume the Naturalist, only two commitments belonged: "the existence of body and . . . the general reliability of inductive belief formation" (18). But on this question

9. Note that to ask *why* a commitment is counted as part of a framework, as well as why those, and only those, require no justification, may be to start to look for a kind of justification.

Strawson prefers Wittgenstein (primarily, he says, the Wittgenstein of *On Certainty*), who has a more fluid picture. For Wittgenstein, Strawson says, the framework is "dynamically conceived: what was at one time part of the framework may change its status, may assume the character of a hypothesis to be questioned and perhaps falsified" (18).[10]

Moreover, for Wittgenstein, society, rather than nature, establishes the framework: "Wittgenstein does not speak, as Hume does, of one exclusive source, viz., Nature, for these *préjugés*. Rather, he speaks of our learning, from childhood up, an activity, a practice, a social practice" (19). Moreover, the beliefs that form Wittgenstein's framework "are not judgements we actually make or, in general, things we explicitly learn . . . but rather reflect the general character of the practice itself, form a frame within which the judgments we actually make hang together in a more or less coherent way" (19).

By considering the Wittgenstein of *On Certainty* and Hume the Naturalist, Strawson arrives at a position he calls "social naturalism." He summarizes it thus:

> [Hume and Wittgenstein] have in common the view that our "beliefs" in the . . . general reliability of induction are not grounded beliefs, and at the same time are not open to serious doubt. They are, one might say, outside our critical and rational competence in the sense that they define, or help to define, the area in which that competence is exercised. To attempt to confront the professional skeptical doubt with arguments in support of these beliefs, with rational justification, is simply to show a total misunderstanding of the role they actually play in our belief-systems. The correct way with the professional skeptical doubt is not to attempt to rebut it with argument, but to point

10. So whereas for Hume the Naturalist specific claims are beyond question, for Strawson claims *that play certain roles* are beyond question. This brings the issue of induction closer to the "framework" we have been considering: *some or another* system of demands and expectations playing a certain role (namely, making possible human society).

> out that it is idle, unreal, a pretense; and then rebutting arguments will appear as equally idle; the reasons produced in those arguments to justify induction or belief in the existence of body are not, and do not become, *our* reasons for those beliefs; there is no such thing as *the reasons for which we hold* these beliefs. We simply cannot help accepting them as defining the areas within which the questions come up of what beliefs we should rationally hold on such-and-such a matter. (19–20)

Note that, in this passage, the reason given for neglecting the doubt, the reason these commitments need no justification, is not (simply) our natural psychological limitations, but (also) the role the commitments play, a role that Strawson seems to think must be, and is appropriately, ungrounded.[11]

While one may have reservations about Strawson's social naturalism, it is, I think, at least better motivated and less unsatisfying than the simple Humean thought.

Social Naturalism and the Central Argument

We can now add to our earlier picture. We can now see that Strawson's claim about the seriousness of our commitment to interpersonal relating is an expression of his underlying social naturalism. Moreover, this social naturalism has the implications we saw Strawson take for granted when making explicit his further point: insofar as it remains possible to engage in tolerably ordinary interpersonal

11. One can wonder whether Strawson unnecessarily weakens his position by appealing to uselessness or inability to change instead of appealing simply to the claim that certain commitments serve a certain function or role, and that, in that function or role, they are rightly ungrounded—they are not held for reasons. And, *therefore*, one might want to say, questioning them is idle: it is idle, not simply because of a psychological inability, but rather because the role they play is not, need not be, and, perhaps, could not be grounded in reasons. And if one were to make this argument (as Strawson almost does), one can then say (as Strawson does) that questioning their justification shows "misunderstanding."

relationships, we will do so, and, crucially, our doing so will not be illicit.[12] There are no general conditions on the appropriateness of ordinary interpersonal relating that are not also conditions on its possibility. His social naturalism thus puts to rest, for Strawson, not only the question of whether acceptance of the truth of determinism *would* lead to the decay or repudiation of the reactive attitudes, but also the question of whether it *should*. Given our natural commitment to interpersonal relating, engaging in such relationships, where possible, does not require justification; participation in ordinary interpersonal relations is not the sort of thing for which we have reasons.[13] The question of whether we *should* participate is thus idle. If we go on to ask whether, given the truth of determinism, it would nonetheless be rational to abandon human society, we would have "utterly failed to grasp the purport of the preceding answer, the fact of our natural human commitment to ordinary interpersonal attitudes" (120).

Note, too, that, given the social naturalism, Strawson's further point falls out as easily as Strawson thought it would: if, as he thinks, there are no conditions on the appropriateness of ordinary relating that are not also conditions on its possibility, then, given that ordinary relating is evidently possible (because sometimes actual), no general thesis could both be true and give a condition on the appropriateness of ordinary relating. We are thus assured that the generalization strategy will fail—if determinism is true, then we never exempt anyone because they are determined*.

We can also now see, in light of Strawson's social naturalism, why he hopes to adjudicate the dispute between the optimist and pessimist by, so to speak, convincing each to stop talking one step earlier—by convincing the optimist to stop talking about efficacy and consequences and the pessimist to stop talking about determinism. The optimist thinks our practices can be shown justified by appeal to the

12. Or, more accurately, insofar as it remains possible, we will *typically* do so—occasionally we use our resource. In some of those cases (e.g., when in the role of a therapist), there may be specific reasons (e.g., therapeutic goals) that would render ordinary relating in those cases illicit.

13. Not even Carnap-style external reasons.

good consequences they produce. The pessimist (correctly) thinks that this justification distorts, beyond recognition, the notions of desert, blameworthiness, and justice—the optimist would replace these with merely objective relating. But the pessimist, in turn, thinks that ordinary, nonobjective relating will be justified only if determinism is false. From Strawson's point of view, each disputant commits the same of error: they each think that ordinary relating requires some justification. It does not. Thus they each, in their own way, answer an unreal question and thereby "overintellectualize the facts" (131).

Thus, Strawson's social naturalism underlies the first part of his two-part answer, and that first part, itself, is meant to address the question of whether acceptance of the truth of determinism either would or should lead to the decay or the repudiation of the reactive attitudes—either in the normal or in the outlier case.

However, once we understand this, we are left with another puzzle (one also raised for the simple Humean interpretation): If Strawson is, in the end, going to rely on his underlying social naturalism, why does he bother to do all the other work of the article? Why did he draw our attention to the distinction between reactive and objective attitudes, mark the difference between (so called) excuse and exemption, consider whether each could be universalized, and give the seeming facile argument? Why does he go on to make explicit the further point? Why not just appeal directly to the social naturalism? Much of the article might now seem unmotivated. It may now seem that Strawson, himself, is in danger of overintellectualizing.

I believe there are two replies to this objection. First, Strawson characterized the initial dispute between the optimist and the pessimist as centered on "something vital" that the optimist had wrongly left out and that the pessimist was rightly anxious to get back. The pessimist "in the grip of his anxiety . . . feels the vital thing can be secured only if, beyond the facts as we know them, there is the further fact that determinism is false" (108). Strawson wants to restore the vital thing, without relying on the falsity of determinism.

We can make the same point using a metaphor from *Skepticism and Naturalism*. There, Strawson provides a quotation from Wittgenstein that, Strawson says, "neatly sums things up [the social naturalist] point of view: 'It is so difficult to find the beginning. Or, better: it is difficult to begin at the beginning. And not to try to go further back' (471)" (24). From Strawson's point of view, both the pessimist and the optimist start, so to speak, too far back. To restore the vital thing is to start at the beginning.

The vital thing, we learn at the end of Strawson's article, is the acknowledgment of the character and role of the network of reactive attitudes.[14] The optimist's position suggested, to the pessimist and to Strawson, that the optimist advocates relating to others in a purely objective way. The pessimist rightly recoils, but wrongly thinks that only the falsity of determinism would prevent that disaster. Strawson hopes that careful attention to the network of reactive attitudes, to "what it is actually like to be involved in ordinary interpersonal relationships" (113), will convince both parties that they need not appeal to further facts for justification; they can, instead, with the social naturalist, start at the beginning, with the fact of these attitudes.

Thus, the first answer to the current objection (why did Strawson do all the work of the article?) is Strawson's (perhaps naïve) hope that, by drawing our attention to what it is actually like to be involved in ordinary interpersonal relationships, going into some detail about the reactive attitudes, contrasting them with the objective attitude, marking the difference between excuse and exemption, and considering whether we would or should use our resource all the time, he might convince the disputants to accept the upshot of his social naturalism: that this network of attitudes does not require further justification; we can start here, at the beginning.

14. The second, difficult sentence of the following quotation is worth parsing: "It is a pity that talk of the moral sentiments has fallen out of favour. The phrase would be quite a good name for that network of attitudes in acknowledging the character and place of which we find, I suggest, the only possibility of reconciling these disputants to each other and to the facts" (132).

The optimist, "being the sort of man he is," might accept Strawson's position—he may recognize that his consequentialist form of justification threatens to cast all our relating as objective, and so may accept Strawson's social naturalism (and so adjust his talk of consequences).[15] But the pessimist is unlikely to be convinced—and Strawson knows this.

The pessimist's recalcitrance provides the second answer to the current objection: Strawson not only wants to restore the "vital thing" but also hopes to address—or, perhaps better, to destabilize—the pessimist's worries. With his distinctions on hand, he can hope to do so with his second, more concessive reply to the crucial objection. Let us now turn to this.

Notice, first, that the pessimist's concern is a *moral* concern: he is asking whether, if determinism is true, it is just or fair, *morally* appropriate, licit, or correct to hold people responsible. The pessimist thinks there is some moral standard on aptly holding others responsible (perhaps about control or about opportunities or possibilities) that will be violated if determinism is true.

We have seen that Strawson rejects such standards, due to his social naturalism. The pessimist does not accept the social naturalism. However, we also saw that Strawson's distinction between (what we called) excuse and exception, together with his interpretation of the cases, will make it difficult for the pessimist to find evidence for the sort of standard he needs, in the cases. We will return to the skeptic's objections to social naturalism in the final chapter.

For now, suppose the pessimist grants the further point—as things currently stand, we never exempt because we believe people are determined*. If the pessimist grants this but continues to ask whether, in light of the truth of determinism, it would be *morally legitimate* to continue to relate to others in the ordinary way, then Strawson can then ask, as he did with the legal system, "to what [moral] standards

15. Strawson concedes, late in the article, that "when certain of our beliefs about the efficacy of some of these practices turns out to be false, then we may have good reason for dropping or modifying those practices" (133). This much of the optimist's position is consistent with Strawson's own picture.

are we appealing?" Once the pessimist grants that being determined* is not a reason we now use to exempt, he cannot deploy the generalization strategy. In fact, once the pessimist grants that the fact that someone is determined* is not among our reasons for exempting—that it does not, so to speak, appear in the moral rulebook—it is difficult to see how the pessimist can appeal to its truth as morally relevant, at all.[16] It may now seem that asking whether, given the truth of determinism, engaging in our moral practices would be morally legitimate is like asking whether, given the truth of determinism, the legal system would be legal. But, again, Strawson thinks it makes no sense to question or to criticize a practice, as a whole, in terms that owe their meaning to that practice. Doing so is, he thinks, the essence of certain skeptical positions, which are to be rejected as absurd. So, if the pessimist grants the further point, then the question he would like to raise cannot be coherently asked. He would be vulnerable to the broadly Wittgensteinian argument (as hinted at in the footnote about induction).[17]

16. He may try to mount an argument that starts not in cases, but in an account of the nature of morality. We will consider this in the next chapter.

17. I do not see Strawson making this broadly Wittgensteinian argument, explicitly, in "Freedom and Resentment," but I do believe he expects his reader to fill it in. It is a standard sort of argument for Strawson to make, and one standard interpretation of "Freedom and Resentment" understands him to be relying on it.

Compare his treatment of skepticism in *Individuals: An Essay in Descriptive Metaphysics*: "This gives us a more profound characterization of the sceptic's position. He pretends to accept a conceptual scheme, but at the same time quietly rejects one of the conditions of its employment. Thus his doubts are unreal, not simply because they are logically irresoluble doubts, but because they amount to the rejection of the whole conceptual scheme within which alone such doubts make sense. So, naturally enough, the alternative to doubt which he offers us is the suggestion that we do not really, or should not really, have the conceptual scheme that we do have; that we do not really, or should not really, mean what we think we mean, what we do mean. But this alternative is absurd. For the whole process of reasoning only starts because the scheme is as it is; and we cannot change it even if we would" (35–36; see also 106, 109).

Compare, too, what Strawson and Grice characterize as "a typical example of a philosopher's paradox. Instead of examining the actual use that we make of the

Even so, the pessimist could raise the crucial objection as stated by Strawson: the pessimist could go on to ask whether it is *rational* to continue ordinary relating. Strawson says that such a question is unreal. But he also gives his second, more concessive, reply. He says that, if one insists on "imagining what we cannot have" and so asking this question, we *still* could not answer it in the way that the pessimist would like to, namely, by considering whether anyone "really" deserves resentment, indignation, punishment, and so on. Once one grants the further point, that possibility is ruled out by the broadly Wittgensteinian thought. Instead, we would have to answer it by appeal to some *extramoral* standard—by appeal to the "gains and losses to human life" (such as, e.g., longer life expectancy, greater wealth, therapeutic improvements, etc.). But, this is just the way the *optimist* addresses the question. And Strawson and the pessimist are in agreement that this is the wrong way to answer the question that the pessimist wants to ask.

Thus, I believe, Strawson engages in much of the work of the article not only to restore the vital thing, but also in an effort to destabilize the pessimist's worries: The further point would disarm the generalization strategy. With that strategy disarmed, the broadly Wittgensteinian thought (which remains in the background) might be used to argue that it makes no sense to ask whether the moral system is, as a whole, morally legitimate. The most one could then intelligibly ask, Strawson thinks, is whether our moral practices are, in some other way, good or bad—but that is not the question the pessimist wanted to ask.[18] Moreover, it is not a question Strawson thinks real. We should, instead, rest content in the fact that the existence of some such framework of interpersonal demands and reactions is a natural part of human life—a part that can be improved

---

notion . . . the philosopher measures it by some perhaps inappropriate standard . . . and because it falls short of this standard, or seems to do so, denies its reality, declares it illusory." H. P. Grice and Peter F. Strawson, "In Defense of a Dogma," *Philosophical Review* 65, no. 2 (April 1956): 147.

18. One might think that, in asking whether they are rational, one was asking not whether they are good or bad, but simply whether they are *coherent*. We will consider this in the final chapter.

and refined, to be sure, and can even be improved and refined for moral reasons, in light of our increased understanding of our place in nature—but not part of life that could or even should be abandoned.

Of course, the pessimist will not be satisfied. In the next and final chapter we will consider the objections that remain. In so doing, we will bring to light more of Strawson's underlying metaethical picture.

5

The Remaining Objections

WE HAVE, I HOPE, come to understand the central argument of "Freedom and Resentment," and, in particular, how it relies on Strawson's underlying metaphysics of morals—a metaphysics of morals that is, in turn, an expression of Strawson's broader social naturalism. I believe that, so understood, the argument is powerful. But, of course, objections can be raised. The pessimist, in particular, will remain dissatisfied. In this final chapter we will consider what I imagine to be the pessimist's remaining dissatisfactions and provide, on Strawson's behalf, some possible replies. Though the objections are serious, I believe robust replies can be supplied.

Intermediate Principles and Cases

The pessimists' continued dissatisfactions spring from his firm conviction that there are conditions not just on the possibility or workability, but on the *moral legitimacy* of relating to others in the ordinary way, in ordinary circumstances. The pessimist believes that it is unfair (or unjust, or morally illicit) to hold someone responsible when their action was out of their control, or when they lacked alternative possibilities, or when their behavior was forced or fated or not up to them. Moreover, the pessimist believes, if determinism is true, then everyone is always in this condition: all of us always lack control, possibilities, or some such. And thus, if determinism is true, ordinary relating is always morally illegitimate. We called this argument "the generalization strategy." It depends on the claim that there are moral standards that all ordinary interpersonal

relating could fail to meet, requiring us to adopt the objective attitude universally.

We have seen that, as a consequence of his social naturalism, Strawson rejects such standards. He thinks there are no standards on the legitimacy of engaging with others in the ordinary way, only conditions of its possibility. (And, since such ordinary engagement is evidently sometimes possible, no thesis that is both general and true will show it inappropriate.) Moreover, we have seen that Strawson's arguments in "Freedom and Resentment" presuppose the social naturalism that supports this rejection. The unconvinced pessimist will not accept social naturalism.

Absent social naturalism, Strawson's rejection of the pessimist's standards would rest simply on his examination of our practices—that is, it would rest on his interpretation of cases in which we suspend or modify the reactive attitudes—together with an assumption that the disagreement must be resolved by appeal to cases.[1] But the pessimist will no more accept Strawson's interpretation of cases than the social naturalism.

How to adjudicate this disagreement? In one way, Strawson seems to have the advantage: he can, at least, support his interpretation of the cases by situating it in his larger theoretical picture (the social naturalism). The pessimist could respond with his own metaphysics of morals—that is, the unconvinced pessimist could support his interpretation of the cases by providing an argument that starts from claims about the nature of moral requirement or moral demand and reaches the conclusion that moral demands require a form of control, possibility, originality, spontaneity, or some such, that is ruled out by the truth of determinism. We would then be left to compare the pessimist's metaphysics of morals with Strawson's.

Strikingly, though, those who argue the pessimist's position rarely delve into the nature of moral requirement or moral demand. They typically, instead, trade in intermediate moral principles (their standards about desert, control, alternatives, possibility, originality, and

1. The article is often interpreted in this way. Strawson's interpretation of cases gives *some* reason to reject the standards.

so on), which they usually simply present as intuitively compelling, illustrating with cases. However, once such an intermediate principle is questioned (because, for example, it is incompatible with the truth of determinism) and an alternative interpretation of the cases is presented, something more than simple appeal to intuition is called for.

One might think the pessimist should return to the cases, now appealing to them as *evidence for*, rather than merely *illustrations of*, his principles. However, as we saw in chapter 3, even absent the social naturalism, the distinctions Strawson has drawn make it difficult for the pessimist to support his intermediate principles in this way.

To review: The pessimist's trouble comes in two stages. First, Strawson's distinction between (what we called) excuse and exemption requires the pessimist to choose his cases carefully. Often, when we learn that someone was caused to do something (or, that they had no alternative, or what they did was out of their control, or not up to them, or forced, or fated), we thereby learn that we were *mistaken* about the quality of their will—they really meant no harm, or it could not be avoided, or they slipped, and so on. These are cases of "excuse"—in our rough slogan, these are cases in which the will was not ill.[2] However, the truth of determinism will not show that we have always been mistaken about the quality of one another's wills—that no will has ever been ill. That, Strawson pointed out, would be a consequence of the reign of universal good will, not universal determinism. Thus, if determinism is to provide a reason to suspend or modify all reactive attitudes, it would have to be a reason to "exempt", rather than to "excuse": it would have to be a reason that shows, in our rough slogan, that the (sometimes poor) quality of other's wills never matters in the way we have thought it does. And so, in light of this distinction between excuse and exemption, the pessimist must choose his cases carefully: he must find a case in which the quality of a person's will does not matter because that person's will is somehow out of their control (or forced or fated or not up to them). There are such cases; we

2. See, again, chapter 1, footnotes 5 and 6.

considered the case of an impulse control disorder. The pessimist might then suggest that, if determinism is true, every case is like this one.

However, Strawson poses a second challenge by providing an alternative interpretation of these cases: According to Strawson, when we exempt someone with an impulse control disorder, we do not do so *simply* because their will is somehow out of their control (or forced or fated or not up to them). Rather, we exempt them because their will is out of their control *in such a way* as to render them incapacitated for ordinary interpersonal relating. Once this alternative way of interpreting the relevant cases is on the table, the pessimist may have difficulty finding evidence, in cases, to support his interpretation over Strawson's. It is difficult to locate cases in which characteristically engaged relating is tolerably possible but is nonetheless illicit.[3]

A Pessimistic Metaphysics of Morals?

Perhaps, then, the unconvinced pessimist should attempt his own metaphysics of morals, after all. What kind of overarching account of the nature of moral demands and requirements would support a pessimistic position?

The optimist's consequentialism certainly does not—it renders the possible truth of determinism irrelevant. A contractualist picture would do the same, though for somewhat different reasons.[4] An ethics grounded in human excellence or human nature would also, it seems, accommodate itself to the facts on the ground—to whatever turns out to be true of excellent humans choosing well.

3. Engaged relating is possible but, perhaps, ill-advised or unwise, in the case of the jackass: better to use your resource. However, merely being unwise or ill-advised is not enough for the pessimist's argument. It must be illicit, unfair, undeserved. See chapter 3, footnote 5 and footnote 20 in this chapter.

4. If the moral standards are the ones we would all agree to under certain circumstances, we would, presumably, take into account, in our agreement, whatever limitations are true of us. See T. M. Scanlon, *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998), chaps. 6 and 8.

If one believes the moral standards are established by divine command, grounded in divine nature, or (especially) enforced by a deity who is sovereign over the physical world, one could then find a way to argue that moral standards require, for their legitimate application, a kind of freedom that is incompatible with determinism. However, it seems unlikely that one would, in this case, be pessimistic—it seems likely that one would also find, in one's theology, reasons to expect determinism to be false. Likewise, a Kantian could perhaps find resources for arguing (rather than asserting) that the moral law requires a kind of freedom that is incompatible with determinism, but the Kantian may also argue that, although our independence from the physical world does not appear when we consider ourselves as merely empirical subjects, it is nonetheless true of us as we are in ourselves. Neither of these metaphysics is pessimistic.

Since I find a pessimistic metaphysics of morals difficult to locate, I will not, here, consider one. I will, instead, consider the strategy I believe the pessimist is most likely to pursue.

Against Social Naturalism

Rather than present his own metaphysics of morals, the pessimist is likely to go on the offensive; he is likely to argue that Strawson's social naturalism cannot provide a satisfying account of moral standards. It is a good strategy. Serious challenges can be raised to Strawson's underlying picture. In fact, it will seem to many, even non-pessimistic, ethicists simply a nonstarter. (I suspect this is why Strawson's argument has typically been misunderstood.)

On Strawson's socially naturalistic picture, moral standards seem to appear with, and to be sustained by, human social activity. They are constituted, at least in part, by actual moral practice.[5] Thus it seems

5. An alternative might see moral standards as constituted not so much by practice as by a practical *problem*: the problem of sharing a world with others. This is how I understand contractualism. With more time, I would compare it to the position in the text. Both ground moral standards in something actual. Both can allow them to reach beyond that starting point. My initial sympathies are with the contractualist, but I find Strawson's position intriguing.

that the practice is somehow prior to the standards—or, better, not posterior to them.[6] Strawson thus concludes that the standards are attuned to our actual capacities—capacities we already know we currently enjoy, whether or not determinism is true.

The pessimist, as well as many who are not pessimistic, will raise a serious objection: The purpose and role of moral standards is to govern our social practice. The standards are not meant to *reflect* the practice, but rather to *regulate* it. When there is divergence between the standards and our practice, it is the practice, not the standards, that must change. Hence, it might seem, the standards must be established independently of, must find a ground somewhere other than, our social practice—otherwise they will not have the right kind of "authority" over the practices.[7] They will lack critical purchase. Thus, one might think, social naturalism is inadequate as an account of the nature of moral expectations and demands, precisely because it allows moral standards to adjust to what is ordinary: the same feature

6. This is not a common thought in ethics, but similar thoughts are more common elsewhere. Wittgenstein is associated with the idea that use establishes meaning. Strawson is ambivalent about this thought. (See the final chapter of Strawson, *Skepticism and Naturalism*.) Compare the reaction of Strawson and Grice to Quine's thought that conventional rules might establish synonymy: "The notion of synonymy by explicit convention would be unintelligible if the notion of synonymy by usage were not presupposed. There cannot be law where there is no custom, or rules where there are not practices (though perhaps we can better understand what a practice is by looking at the rule)." Grice and Strawson, "In Defense of a Dogma," 153.

As applied to ethics, it is an intriguing picture and (assuming it can withstand the objections raised below) one we might do well to consider, especially in light of the challenge from evolution advanced by Sharon Street in, e.g., "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies* 127, no. 1 (2006): 109–66. (Strawson identifies himself as a realist, though Street would not so identify him. See Strawson, *Skepticism and Naturalism*, chap. 4.) Consider, too, T. M. Scanlon's evocation of Carnap in his discussion of the metaphysics of reasons, in *Being Realistic about Reasons* (Oxford: Oxford University Press, 2014). Time may be ripe for a midcentury revival.

7. Always question "authority."

that allows Strawson to make his argument against the relevance of any general thesis also renders his picture unacceptable.

This concern about authority, or about critical purchase, can make Strawson's underlying picture seem simply a nonstarter. After all, the actually existing framework of expectations and demands in any given human society can be, and often is, ghastly, as any brief glance through history or across the globe will attest—or a more searching gaze here and now. Strawson has claimed that the contents of the demands in a society are determined, at least in part, by what is *statistically ordinary*—but, atrocities are ordinary. In certain societies, it is ordinary to punish women for being raped, or to beat those of lower castes for failures to show deference. In our own society, it is ordinary simply to ignore the desperate but satisfiable needs of those distanced from us, either by time and space or by social class or category. Genocide is, in some sense, ordinary. Xenophobia is certainly ordinary—and likely part of our inheritance in human nature. But ordinariness does not render these permissible. These are flagrant violations of reasonable demands for respect and goodwill.

One might try to reply, on Strawson's behalf, by distinguishing between ordinary *practices*, which can be ghastly, and ordinary *capacities*, to which any practice must be sensitive: The pessimist claims that, if determinism is true, then we all lack a capacity presumed by moral demands. One might try to insist that moral demands will presume no more than the typical capacities while allowing moral demands to remain insensitive to, and thereby potentially critical of, the typical practices.

This will not work, as a reply for Strawson. First, it cuts against the social nature of his social naturalism, which relies on practices. Second, capacities for respecting others (e.g., for overcoming xenophobia, or delaying one's own gratification for the sake of another, or resisting envy) can vary just as capacities for memory or inhibitory control can.[8] We saw, earlier, that demands that exceed the

8. Or, at least, Strawson should think so. Why deny that capacities for respecting others can vary just as any other capacity? Presumably because one affirms a strong

ordinary capacity for memory, attention, or inhibitory control will not be part of the adjustable system. However, if capacities for respecting others can vary just as capacities for memory, attention, or inhibitory control can, then it might seem that Strawson cannot avoid the conclusion that demands that exceed the current capacity for respect for or sensitivity to others will also not be part of the adjustable system. If so, then it will turn out that, as long as what *we*

---

version of "ought implies can," that is, presumably because one thinks that each person facing a legitimate moral demand must have the capacity to satisfy that demand (or to bring it about that they do), regardless of their contingent history or circumstance. Call this capacity "moral freedom." Presuming moral freedom opens one to the pessimist's worries: a capacity to satisfy moral demands, regardless of one's history or circumstances, seems to require an independence from history and circumstances that the truth of determinism might undermine. Strawson should not presume moral freedom. (Though see footnote 25.)

The presumption that ought implies can is, I think, standard in many contemporary moral intuitions, and it motivates the pessimist's position. Note, though, that there are two ways of accommodating it: assert ability or deny obligation. In the second critique, Kant argued from the fact of moral obligation to the (noumenal) reality of moral freedom. See Immanuel Kant, *Critique of Practical Reason*, Cambridge Texts in the History of Philosophy, trans. Mary Gregor (Cambridge: Cambridge University Press, 1997), 5:30. Others argue in the contrapositive, from the fact of human limitation to the limitation of legitimate moral obligation. The pessimist is the limiting case of the latter strategy, arguing from the elimination of moral freedom to the elimination of morality. Later in this chapter I will consider the implications, for Strawson, of the fact that moral intuition includes some commitment to moral freedom.

For now, note a contrary thought, also standard in the moral intuition of many: Sometimes people are simply too weak of will, or too vicious of character, too selfish or petty, to rise to the occasion and do what is nonetheless required of them. The process of moral development and education leaves some of us inadequate to the moral task, subject to what we might call "original sin." This second thought seems to me clearly correct: it is simply not the case that each of us always, regardless of our history and circumstances, has the capacity to do what is nonetheless rightly required of us. Strawson's naturalism should leave him with the same opinion.

Thanks to Sarah Buss and to an anonymous reviewer for pressing for clarification.

would call disrespect is sufficiently entrenched in the culture, it will not be disrespect, after all.[9]

In sum, the objector claims that Strawson's social naturalism is inadequate as an account of the nature of moral and interpersonal standards because such standards must have the ability to criticize the status quo, and do so radically, not merely at the edges or incrementally. By relying, as Strawson does, on actually existing capacities and practices, he has deprived moral standards of the independence they require to do the work they are meant to do. Again, the feature that allows Strawson to make his argument also renders his picture unacceptable.

A Defense of Social Naturalism

In reply, Strawson should first point out that he has not advocated a merely descriptivist ethics. He has instead claimed that any human society will include a certain general, formal feature: a framework of expectations and demands constituted by certain characteristic reactions to the violation or supersession of those expectations or demands. Participation in some or another such framework, he claims, is "original, non-rational . . . in no way something we could choose or give up" (131n7), and so, he thinks, beyond questions of justification. We have seen that he also believes the existence of a working human society guarantees that certain *minimal* demands and expectations are typically met. We have surmised that those minimal demands must be attuned to the statistically ordinary capacities. Thus, Strawson can conclude that most people enjoy the capacities required for minimal participation in the framework. But, within the framework, Strawson has claimed, there is "endless room for modification, redirection, criticism and

9. The point must be put carefully. If "disrespect" is simply whatever violates the standards currently at work, disrespect will always be wrong. However, this observation does not address the concern. The objector worries that *what we now regard as* serious violations of human dignity (or equality or freedom or self-governance) may, in certain circumstances, no longer count as disrespect.

justification" (131).[10] Thus he makes room for critical purchase, within the practice.

One might wonder, what provides that critical purchase? On what basis do we modify, redirect, criticize, and so on? Strawson mentions rational consistency.[11] However, given the flexibility of his social naturalism, it seems that Strawson can and should also allow that certain *ideals* can and will be incorporated into the moral and interpersonal framework—ideals that set expectations above the minimum, which expectations may not typically or ordinarily be met. We could, for example, incorporate into the framework a concern for equality, or self-governance, or the freedom of each consistent with the freedom of all, or symmetric standing in determining the terms of our respectful relating, or a concern for producing certain goods, or exemplifying human excellence, or avoiding pain in sentient creatures, or caring for the natural world.[12]

To see how this is possible, note again the difference between what we might now call *standards of regard*, that is, expectations and demands "within the framework," to which we respond with the reactive attitudes, and what we might call *exempting conditions*, that is, the conditions that preclude being a participant in the framework, "a term in moral relationships" (124), someone to whom we apply the

10. "Endless" is overstatement—the minimal requirements must be kept, and Strawson insists on reciprocity.

11. I do not know how far he thinks consistency will take us—it is a Kantian hope that it will take us very far. See footnote 25.

12. Perhaps, just as Strawson was unwilling to call the framework "moral" unless it included reciprocity, we should be unwilling to do so unless it includes such ideals. But, like Strawson, I would also wish to keep our sense of the "moral" spare enough to allow for a variety of ideals to thrive in a liberal society.

What I am calling "symmetric standing in determining the terms of our respectful relating" appears, I believe, in Scanlon's contractualism. As an ideal, it supports what I call below the "counter-pressure." See Scanlon, *What We Owe to Each Other* and Pamela Hieronymi, "Of Metaethics and Motivation: The Appeal of Contractualism," in *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, ed. Rahul Kumar, Samuel Scheffler, and R. Jay Wallace (Oxford: Oxford University Press, 2011), 101–28.

standards of regard. The first are the standards by which we determine (in our rough slogan) whether your will was ill. The second are the conditions that determine (in our rough slogan) whether your (possibly ill) will matters to us in the usual way.[13]

Strawson believes that our "natural commitment to ordinary interpersonal relating" ensures that exemptions will be rare. However, one could maintain this without insisting that the standards of regard will typically be satisfied. Although Strawson did claim that the existence of a society guarantees that a certain *minimum* set of demands and expectations will typically be satisfied, maintaining this is consistent with allowing other, more robust demands and expectations that are not typically satisfied—perhaps even only rarely satisfied (in just the way that certain standards of evidence are typically not satisfied, due to, for example, confirmation bias).

Thus, one might think that Strawson can avoid the serious objection simply by allowing certain ideals to be incorporated into the framework of moral and interpersonal demands and expectations, where these ideals will set standards of regard—and perhaps set them much higher than typically satisfied by the usual practices—without affecting the exempting conditions.

However, Strawson is not out of the woods. More must be said, because we have seen that, on his picture, the fact that exemptions

13. I say "exempting *conditions*" rather than "*standards* of exemption" attempting to mark the fact that, for Strawson, nonexemption is simply the default, not something that needs or requires justification. I mean simply the conditions (such as those Strawson has listed) under which we exempt.

It seems to me that the difference between standards of regard and exempting conditions is the difference at which interpreters sometimes gesture with the slippery spatial metaphors of "internal" and "external." I very much distrust these metaphors, and I have tried to minimize their use. (I likewise dislike the metaphors of "constituted by," "prior," and "posterior," on which I earlier relied. I welcome clarification.)

In the next page of the main text I note that, for Strawson, the need to limit the exempting conditions puts pressure on the standards of regard—the two interact. I believe this is, in part, why the "internal/external" metaphor is so unsatisfying.

will be rare can put significant pressure on, and in fact will help to determine, the standards of regard.

We saw the effects of this pressure earlier, when we considered the community of people who are naturally equipped with only the degree of inhibitory control, memory, and attention that we typically enjoy when fairly inebriated. As things stand, we typically regard certain sorts of outbursts or revelations, or certain moments of distraction, inattention, or forgetfulness, as disrespectful—violations of the standards of regard. Though, if they appear in isolated bouts of drunkenness, we sometimes exempt (saying, e.g., "he wasn't himself").[14] However, in a community of people whose natural capacities are diminished, finding the same outbursts (or, distraction, inattention, and so on) disrespectful would be both unreasonable and emotionally unsustainable. Thus, it seems, people in that community would not be prone to respond to those outbursts with negative reactive attitudes. Importantly, though, in such a community, it will not be the case that everyone is constantly *exempting* one another from some unchanging, universal standards of regard (standards that *we* typically meet but *they* cannot). Rather, in such a community, the standards of regard adjust to suit the actual capacities.[15]

Why think that the standards of regard adjust? Because those standards *just are* the expectations and demands constituted by the reactive attitudes. As Strawson put it, "the making of the demand *is* the proneness to the attitude." If failing to react to certain outbursts is the norm, then that failure to react is not a case of *exempting* from the usual expectations or demands (nor a use of the resource). Rather, no expectation has been violated; no demand is in place. To be sure, one may maintain certain *beliefs* about what would be best, or even about what would be morally ideal, without any proneness to the

14. For details on drunkenness, see chapter 2, footnote 14.

15. I believe this sort of thing happens in many actual cases—perhaps most vividly in nursing homes. Those in such communities need not be simply exempting one another—as Strawson would allow. See chapter 1, footnote 7; chapter 2, footnote 15; and this chapter, footnote 17 and 35.

reactive attitudes. One might well think it would be better, morally speaking, if people refrained from those outbursts. However, if those ideals are not (in a society or for an individual) among the standards whose violations tend to elicit reactive attitudes, then the violation of those standards does not, in fact, *matter* (in that society or to that individual) in the way that violations of the standards of regard matter—those ideals are not ones to which people are held responsible in the usual way. Thus, they do not qualify as among the standards of regard.[16] The standards of regard shift with the reactive attitudes.[17]

In the community with diminished natural capacities for memory, attention, or inhibitory control, an adjustment of the standards of regard seems welcome. We can note, too, that the same dynamic will allow the standards of regard to adjust upward—if people develop greater capacities for memory or attention, they might ask more of another, and, insofar as it is workable, it may become expected. That seems a happy outcome.

However, while the flexibility seems welcome in those two cases, it invites the serious objection: what is to prevent those standards from adjusting downward, to accommodate the circumstances, until what we would regard as serious disrespect no longer matters?

Again, the worry is not idle: humans are not born possessing the emotional capacities required to face the ethical and interpersonal challenges of adult life; we require moral and personal education

16. The fact that certain ideals that might plausibly be called "moral" might not be incorporated into the standards of regard might clarify a variety of puzzling features of "morality."

17. Earlier I claimed the question of whether the use of the resource was a case of exemption was delicate: there are grounds for saying that, when you use your resource, you simply ignore violations of standards that you nonetheless maintain. However, the *universal* use of the resource would eliminate the demands: there would be no basis for claiming that the expectations or demands remain in place. Thus it turns out that society-wide "exemption" is not possible—if nonreaction becomes the norm, then the operating standards of due regard have changed. (Again, the introduction of the label "exemption" can be misleading.) (See, again, chapter 1, footnote 7; chapter 2, footnote 15; and this chapter, footnote 15.)

and development. Tragically, those processes can go awry. As a result, people often enough arrive at adulthood simply too weak of will, or too vicious of character, or too selfish or petty, or too full of self-doubt or self-criticism, or too stubbornly blind to the needs or experiences of others, to rise to the occasion and do what, one might think, is nonetheless required of them. Often enough, people arrive at adulthood without the capacities required to treat others (or, perhaps certain classes of others) with the deference or accommodation they would expect for themselves (and, often enough, also without the wherewithal to recognize this fact about themselves).[18] Worse, a given culture might positively reinforce and regularly reproduce these failures.[19] Why, then, would the standards of regard not adjust to *those* incapacities, in those circumstances, just as it did for those incapable of greater memory, attention, or inhibitory control?

Strawson should respond first by granting that there will, indeed, be pressure for the standards of regard to adjust to the status quo, to accommodate the tendencies and incapacities of the majority (or the culturally dominant). He should then point out that there will *also* be counter-pressure, coming both from the needs and interests of those disadvantaged and from demands of consistency, to resist this pressure—to maintain in, or to incorporate into, the standards of regard a concern for needs and interests of each and to apply those enriched standards consistently across the population. Insofar as this dynamic of pressure and counter-pressure accurately describes our social and moral life while allowing for critical purchase, Strawson will, I think, have a robust defense against the serious objection.

To illustrate and elaborate upon this reply: Let us start by reconsidering the pressure to accommodate the majority, but now examining, not limitations of memory, attention, or the like, but rather widespread and serious failures of interpersonal respect. Suppose you somehow come to live among people the vast majority of whom are,

18. See footnote 8.

19. For only one current example, consider the current concern about "rape culture."

as you see it, regularly violating basic demands of interpersonal respect through, for example, a strict system of caste, or racism, or sexism, or other social stratification. Given your conviction that each person deserves equal respect and equal treatment, it seems you should be regularly, frequently indignant, even outraged, and, if you fall into one of the disfavored categories, resentful. But those emotions are costly, both personally and interpersonally. It will therefore be difficult to avoid either regularly using your resource, to cope with "the strains of involvement," or else coming to see those you live with as somehow morally immature or diseased.[20] However, if you do either, you will have, in effect, exempted those around you from the standards you believe they are violating. By doing so consistently, you will shift your standards of regard. Although you may remain convinced, intellectually, that those you live around are being seriously disrespectful, neither your actions nor your reactions will embody that conviction. It may become difficult to maintain, publicly, your claim that the violations are serious ones—that they really matter. Although you may continue to believe that a serious moral requirement is being violated, you no longer include that requirement among the interpersonal expectations or demands to which you hold others responsible. Your moral convictions may seem to be so only in name.[21]

20. One might note that, on Strawson's picture, the line between using your resource to avoid the strains of involvement and exempting a person due to incapacities will not be a sharp one. Thus, the line between vice and disease is also not sharp. This seems to be so, in practice: we respond to the extreme jackass and the psychopath in much the same way. The blurring of this line seems to me salutary: our main interest in drawing the line is to determine who is to be held responsible. But, for Strawson, this is just the question of whether ordinary relating is tolerably possible, and in neither case is it. We can draw a line by saying that, in one case, the impossibility is due to pathology, while, in the other, it is due to inadequate moral development. But, if that difference makes no significant difference in practice, we have an explanation for why the line can seem difficult. See chapter 3, footnote 5 and this chapter, footnote 3.

21. If the example seems unconvincing, one might instead consider the current plight of the global poor, or the treatment of animals in factory farming, and how

Alternatively, you might, instead, continue to suffer the strains of involvement, you might continue to respond to others as responsible adults violating important standards of regard—that is, you may continue to relate to others with the resentment or indignation called for by your convictions.[22] If so, you will be holding others to standards that they, with the crowd, neither recognize nor regard as reasonable. *They* will then resent *you*. As time goes on, it is likely they will either turn against you (the list of martyred moral reformers is long) or else begin to use their resource to respond to you more objectively: you will become a problem, an issue, or perhaps a kind of curiosity or museum piece. You will then be left outside the scope of ordinary interpersonal relationships.[23] If you are one and they are many, this will not be so difficult for them.

---

those facts do or do not manifest in one's own interpersonal and intrapersonal reactions to everyday decision making.

22. It is crucial for the plausibility of the position I offer to Strawson that the reactive attitudes are, as I would put it, neither voluntary nor involuntary. Though we can sometimes voluntarily use our resource to step away from them, so to speak, these attitudes are not *themselves* voluntary—they cannot be simply adopted or abandoned for any reason one would like (as the position of the optimist seems to suggest). Nor are they involuntary—they are not merely passive occurrences to which we are subject, like a sensation or a perceptual experience. I have elaborated on the form of agency embodied in these attitudes elsewhere, calling them "non-voluntary" activities. See, e.g., "Controlling Attitudes," *Pacific Philosophical Quarterly* 87, no. 1 (March 2006): 45–74; "The Wrong Kind of Reason," *Journal of Philosophy* 102, no. 9 (September 2005): 1–21; "Believing at Will," *Canadian Journal of Philosophy* 35, suppl. (2009): 149–87; "I'll Bet You Think This Blame Is about You," in *Oxford Studies in Agency and Responsibility: Essays on Themes from the Work of Gary Watson*, ed. Justin Coates and Neal Tognazzini (Oxford: Oxford University Press, 2019), 60–87.

Strawson seems to want to point to this aspect of the reactive attitudes in the final paragraph of "Freedom and Resentment": "These practices, their reception, the reactions to them, really *are* expressions of our moral attitudes and not merely devices we calculatingly employ for regulative purposes. Our practices do not merely exploit our natures, they express them. Indeed the very understanding of the kind of efficacy these expressions of our attitudes have turns on our remembering this" (133).

23. The dilemma (curtail your standards or be criticized or ostracized) seems to be faced by many, both progressive and conservative.

It seems, then, that the natural human commitment to ordinary, engaged relating will, indeed, put pressure on the standards of regard—pressure toward conformity or toward what is typical—even if what is typical is seriously disrespectful. The pressure will make it difficult to respond to widespread disrespect as disrespectful, just as the objector surmised. The standards are in fact in danger of slipping. Again, Strawson should grant this.

Having granted this, Strawson should also point out that *it is* difficult to respond to widespread disrespect as disrespectful, in just the way described. One might see it as a strength of Strawson's picture that it captures some of this difficulty.

Turning, next, to the counter-pressure: It may seem odd to think that there could be any counter-pressure for change or reform of the poor practices of the majority, given both the fact that the standards of regard must be emotionally and interpersonally sustainable and the fact that people may be simply incapable of showing respect for others (or even of recognizing their own incapacity to do so). However, I believe that maintaining atypical standards of regard—continuing in emotionally engaged expectation and demand, even when those expectations or demands are widely unrecognized—can be emotionally and interpersonally sustainable, in the right circumstances. Though it will surely be difficult.[24]

Let us start by making a distinction. We have noted that many ordinary adult capacities are gained through successful personal, interpersonal, and moral education and development. We might then distinguish between what we could call *purely natural* capacities and limitations—such as our ability to remember many things but not everything—and what we could call *socially developed* capacities and incapacities—capacities that are the result of socialization, education, and inculturation, such as the ability to lose gracefully; or to relate, without fear, to those very different from you; or to find disagreement interesting rather than threatening; or to recognize the needs of others and accommodate them at some cost to yourself.

24. Again, engaging is not voluntary, not something that can be done for strategic reasons. See footnote 22.

Demands or expectations that exceed the typical *natural* capacities will be consistently met with frustration, and, it seems, such demands will be both unreasonable and unsustainable—as we saw in our imagined community with diminished capacity for attention, memory, and inhibitory control. The standards of due regard will attune themselves to the typical natural limitations.

However, the same need not be said of demands that exceed the typical *socially developed* capacities: Such demands may also be regularly met with frustration, but it may not follow that they are either unreasonable or unsustainable. Whether they are so will depend on the case. They may be both reasonable and sustainable if enough is at stake for those making the demand. The fact that the demand could be accommodated with the right social or cultural changes, together with the fact that, given the needs and interests at stake, the cost of making those changes is worth paying, makes the demand reasonable.[25]

25. Of course, those ignoring the demand will resist seeing it as reasonable. Seeing it as reasonable—seeing the costs of accommodation as worth paying—will require some appeal to consistency (e.g., "You would pay the cost, if the needs and interests at stake were your own"). The appeal to consistency will typically be responded to, by the dominant, with various spurious distinctions (e.g., "If the needs and interests were mine, then the costs would indeed be worth paying. But there is this or that difference between us, and we are consistently applying a standard that is sensitive to that difference," or "separate but equal" or . . .). Thus, I believe appeals to consistency, alone, will do little work. But paired with needs and interests, and (crucially) in the presence of empathy and ego strength, they are extremely powerful.

Note, too, that the reasonableness of the demand does not depend on the capacities of the *individual* subject to it on a given occasion. A given person may be too entrenched in, say, their chauvinism to rise to the occasion. That does not render resentment or indignation unreasonable.

An inability to rise to the occasion might render imposing sanctions or making requests unreasonable, but the reactive attitudes are neither sanctions nor requests. Such attitudes would also be unreasonable if they included a belief that the person could have done otherwise—but, I would argue, they do not. See my "The Force and Fairness of Blame," *Philosophical Perspectives* 18, no. 1 (2004): 115–48; "Reflection and Responsibility"; "I'll Bet You Think This Blame Is about You"; "Fairness, Sanction, and Condemnation," in *Oxford Studies in Agency and Responsibility*, ed. David Shoemaker (Oxford: Oxford University Press, 2020). And, see, again, footnote 22.

And, the needs and interests at stake can make the demand sustainable (though no doubt difficult), for those who make it. (It will help if they form a supportive community.)[26] Thus it is possible for demands that exceed the typical *socially developed* capacities to be both reasonable and sustainable (though, costly), if enough is at stake for those making the demand.[27]

Note, though, that reformers will introduce ambiguity into the standards of regard at work in their corner of the world—there will be, at work in that corner of the world, both a higher and a lower standard, so to speak. There may be arguments in favor of the higher standard, perhaps even arguments that show it correct. However, because "the making of the demand *is* the proneness to such attitudes" (129), resolving the ambiguity in favor of the higher standard (while maintaining the same society) will require more than argument or even truth: some people on the receiving end of the higher expectation or demand must come to respond, not with counter-resentment or by ostracizing (or worse), but instead with enough empathy and ego strength to enable

---

In fact, it could be that *no one* in the society has the capacity to meet the demand, here and now. It does not follow that the demand is unreasonable. For a demand to be reasonable, it must only be the case that (a) with the right social and personal development and enculturation, most people could be brought to be able to meet the demand and (b) the interests at stake, for those disadvantaged by the status quo, are sufficient to require, as a matter of fairness, that society pay whatever costs are required to change in such a way as to enable most people to be able to meet the expectation and demand. (Thanks to Sarah Buss for pressing for clarification on this point.)

One might note that "ought implies can" turns out to be true, at the societal level—or, at least, "ought implies could-with-enough-time-and-effort." (One might be tempted to say that society, itself, enjoys moral freedom—if given time.)

26. I hope one can see, here, the role of consciousness-raising sessions.

27. One might doubt that there is a meaningful distinction between natural and socially developed capacities. Xenophobia, I said earlier, is plausibly part of our natural endowment. And yet, it seems, any particular manifestation of xenophobia could be overcome with better enculturation. One might suspect that even natural capacities such as memory and attention could be improved with training. We might then consider, instead, what we might call *socially malleable* capacities. Perhaps all capacities are socially malleable. That will not undermine the argument, as the question will then simply be which, among the malleable ones, are worth trying to change.

them to see that the demands are, in fact, reasonable ones, and so begin to react accordingly. This sometimes happens.

Thus, while Strawson should allow that there will be emotional and interpersonal pressure to conform to the majority or the dominant, he can also insist that there can be counter-pressure to incorporate, into the standards of regard, ideals that are not widely satisfied. While the fact of our natural human commitment to characteristically interpersonal relating pushes us to adjust our standards toward the majority or the dominant,[28] that same commitment, in conjunction with the needs and interests of individuals and standards of consistency, can generate counter-pressure to maintain and advance certain ideals—and so either to maintain the standards or to adjust them upward or outward. By appeal to these opposing dynamics, Strawson can deny that disrespect will inevitably cease to be disrespectful simply because it goes widely unrecognized. Certain ideals can, with effort, be incorporated into the actually existing framework and can, with effort, be upheld. They may be contested—but, of course, that is an accurate description of human social life.[29]

One may yet be unsatisfied. One obvious source of dissatisfaction is the clear injustice of the burden these dynamics of pressure and counter-pressure place on those already disadvantaged. This is a serious problem. I am sorry to say, it seems to me that this injustice is both tragic and unavoidable. It is a consequence of a general failure of humanity, of the fact that we are not better than we are—that, collectively, we fall so short.

One might instead be dissatisfied because one discerns that there remains, in this picture, the possibility that certain standards may be simply lost in history, so to speak—just as certain cultures, languages, and species have been. Even if the *ideas* are maintained—in the books, so to speak—the reality may not be.

28. Recall Thrasymachus's claim that justice is the advantage of the stronger in Plato, *Republic*, trans. G. M. A. Grube, ed. C. D. C. Reeve (Indianapolis: Hackett, 1992).

29. These opposing dynamics may explain why it seems right to say that the "central" portion of morality concerns the interests of other people rather than, say, nature or beauty, or even other sentient creatures. That "center" is where the stakes are most pressing.

However, one might think that this, too, is an accurate reflection of how things are. If so, then, in addition to capturing the emotional and interpersonal *difficulty* of pressing for social reform, a picture like Strawson's will also capture some of the *urgency* of incorporating certain ideals into the standards of regard. The reformer is subject to a kind of double-vision.

To illustrate: Though it is edifying to say, for example, that we are all moral equals, that each is endowed with no more but also no less moral standing or worth than any other, in many circumstances there is a palpable sense in which this is not yet so. Some are more equal than others.[30] The interests of some, and the voices of some, dominate the standards of regard, the actually existing moral and interpersonal framework of a society. And so the edifying claim is, in one way, not true. It might be more accurate to say that each person *ought* to have equal moral status—though, of course, if this is so (if each ought to have equal moral status) then, in another way, each one *does*, already, have a kind of equal moral status, insofar as one appeals to the ideal. (Thus the double vision.) The protester of injustice asks for this more rarified, ideal status to be made actual. The urgency of protesting injustice comes, in part, from the fact that something like the here-and-now reality of one's moral standing is on the line and, perhaps, in part from a concern that one's vision of justice might be lost to history.[31] A picture like Strawson's seems able to capture this urgency.[32]

Strawson thus has resources available to him to address the claim that his picture provides too little critical purchase. His view is not as retrograde as one might at first imagine; he can allow for calls for reform and charges of widespread injustice.

30. The phrase belongs, of course, to George Orwell, *Animal Farm* (London: Secker and Warburg, 1945).

31. I once characterized resentment as a protest against the ongoing threat posed by past wrong. It is a threat to what I am calling the here-and-now reality of one's moral standing. See "Articulating an Uncompromising Forgiveness," *Philosophy and Phenomenological Research* 62, no. 3 (2001): 529–55.

32. The ideas of this paragraph—in fact, of this entire subsection—are obviously undertheorized, but I hope they are suggestive.

The opponent may, of course, reject the view in favor of something with greater independence from the actual. The opponent may think it important to ground either the standards of regard or the ideals they might incorporate somewhere safe from the contingencies of history—somewhere they cannot be lost, so to speak. Perhaps they might find grounding in human nature, or rational nature, or divine nature, or in what people call "Plato's heaven," or in the undeniable goods and evils of pleasure and pain, or even in demands for intrapersonal consistency.

The opponent is free to seek such grounds, though I think it remarkably unclear what, exactly, we gain from that kind of grounding—how or why it serves us better, either in our interactions with others or in our own convictions, than does a robust description of, and appeal to, an ideal: the good of equality, or of the freedom of each consistent with the freedom of all, or of the absence of suffering, or the beauty of nature. Perhaps additional grounding will satisfy reason's search for the unconditioned, or perhaps it will help, politically, in the gathering of adherents.[33] It will not, though, provide any better motivation than an appeal to the ideal, itself (or, if it does, we have identified a different ideal). Again, one might say that we should start at the beginning, and not further back.

In any case, I think it safe to conclude that Strawson need not be embarrassed by his own position. He may respond to the opponent's alternative as he does to the skeptic in *Individuals*:

> Finally, we may, if we choose, see the sceptic as offering for contemplation the sketch of an alternative scheme; and this is to see

33. It is clear to me that most people believe we *do* gain—a metaphysics of morals that bottoms out in God or nature or reason or logic (rather than simply stopping at ideals or values) seems to most people to be worth fighting for, both intellectually and otherwise. But I am not clear *what* we gain. The existentialist sees such appeals as appealing because they allow one to shirk one's responsibility, or one's freedom, by allowing the equivalent of "I was just following orders." That dark thought has plausibility. (And is at least less dark than the Nietzschean diagnosis—that the metaphysics or religion allows the high priests to control or to exact revenge. See Nietzsche, *On the Genealogy of Morality*, esp. Treatises I and III.)

him as a revisionary metaphysician with whom we do not wish to quarrel, but whom we also do not need to follow. (35–36)

An Opening for the Generalization Strategy?

The serious objection claimed that the same feature that allowed Strawson to make his argument for the irrelevance of determinism also showed his view unacceptable. Having shown how to find the view acceptable, one might worry we have thereby undermined the argument for the irrelevance of determinism. That is, one might worry that, once we allow for critical purchase, we also provide a foothold for the generalization strategy. We need to consider this.

We have argued that the view can be made acceptable by allowing certain ideals to be incorporated into the standards of regard, even if those ideals are not widely, or perhaps never fully, satisfied. However, if we allow an ideal of, say, equality, to be incorporated into the standards of regard, why not also allow an ideal of, say, fairness, to be incorporated into the conditions under which we hold others responsible? That is, why not also introduce an ideal, not only into the standards of regard, but also into the *exempting conditions*? If we were to do that, then it seems that we could deploy the generalization strategy.

To this Strawson can reply with his social naturalism—he can say that, although the standards of regard may incorporate certain ideals that are rarely if ever satisfied, the exempting conditions cannot. The exempting conditions, remember, are not standards of justification. They are simply the conditions under which ordinary, engaged relating ceases to be tolerably workable. So long as ordinary, engaged interpersonal relating is tolerably possible, we need not exempt. When it ceases to be workable, we do exempt. But there is no question, for the social naturalist, of whether doing so is either justified or unjustified, and thus no standards into which to incorporate ideals.[34]

34. The plausibility of the objection—of thinking that there are standards of fairness on interacting with others in the engaged way—comes from the tendency to think of reactive attitudes either as forms of sanction or punishment or as requests

The opponent may have another go. The opponent may point out that, on Strawson's picture, it is not *simply* that when ordinary, engaged interpersonal relating ceases to be tolerably possible, we exempt—it is also the case that, if we "exempt" uniformly and consistently, in certain sorts of cases, then we have thereby changed the standards of regard. (Or, to put the point more precisely, it is impossible to exempt uniformly, because, if we uniformly fail to react, we have eliminated the standards from which we might have "exempted.")[35] However, in addressing the serious objection, we allowed that, because uniformly failing to react amounts to changing the standards, sometimes certain people will, so to speak, hold out—they will continue to engage, even when engaged interpersonal relating becomes very difficult, because the standards of regard that would be eliminated if they did not are too important to them.[36] That is to say, we have seen that the importance of incorporating certain ideals into the *standards of regard* can put pressure on the *conditions for exemption*. In the cases considered, the pressure in question was pressure to *avoid* exempting—to continue engaging with others, even when it is difficult. However, the pessimist might think that, once we allow ideals to put pressure on the conditions for engaged interpersonal relating, we have opened up the possibility that certain ideals could instead show that we ought always to "exempt," that we ought *never* to engage with others in the usual way.

How would this argument be made? To what ideal would the opponent appeal? When considering the counter-pressure, I introduced the question of whether a given expectation or demand would be *reasonable*. I suggested that only reasonable demands would be sustained—that, if a demand was unreasonable, it would not be able

---

or commands. Sanctions and punishment are subject to questions of justification, while requests and commands are inappropriately directed at those unable to satisfy them. We will consider this below. See also footnotes 22, 25, and 39 and references there cited.

35. This is the delicacy introduced in chapter 1, footnote 7, and followed in chapter 2, footnote 15, and in this chapter's footnotes 15 and 17.

36. Again, this is not voluntary. See footnote 22.

to be incorporated into the system by reformers.[37] In doing so, one might think, I have provided a point of leverage for the generalization strategy: The generalization strategist might argue that it is *unreasonable*—in particular, unfair—to demand or expect people to do what they lack an ability or capacity to do. But, if determinism is true, they might argue, none of us has the ability or capacity to satisfy the demands or expectations. Therefore, they might conclude, none of us is reasonably subject to any of these demands. Therefore, if we are to be reasonable, if we are to treat others fairly, the standards of regard should shift to the point of elimination.

It is true that, by introducing the idea of what is reasonable, I have introduced a variable that deserves more attention than I have given it. It is also true that the standards of reasonableness that I believe must be met by the reformers concern fairness.[38] Nonetheless, the appeal to reasonableness does not provide a point of leverage for the generalization strategist. The strategist's argument fails on multiple counts.

First, the argument relies on two claims: the apparently plausible claim that it is unreasonable or unfair to demand or expect people to do what they lack the ability or capacity to do and the further claim that, if determinism is true, then none of us has the ability or capacity to satisfy demands. Both are incorrect.

37. I first introduced the idea of reasonability when considering limitations of natural capacities, saying that it would be "unreasonable and emotionally unsustainable" to continue to resent someone who has diminished capacities for, for example, memory or attention. I then added the idea that the *reformer's* demands must be reasonable, where that included some appeal to what is *fair*—the reformer's demands must be reasonable in a way that, say, expecting or demanding that everyone treat you specially, or accommodate your needs to the exclusion of others, would be unreasonable. (See footnote 25.) Upholding such an obviously biased standard would be unsustainable, unless you are able to exploit differences in power to which the reformers will not have access—others will simply refuse. (An unreasonable demand *could* be incorporated into the system by the majority or the dominant, but not, I think, by the reformers.) By appealing to what is reasonable, I am relying on some basic ideas of fairness—in much the way that both Rawls and Scanlon rely on basic ideas of fairness in Scanlon, *What We Owe to Each Other*; John Rawls, *A Theory of Justice* (Cambridge, MA: Belknap, 1971).

38. See footnotes 25 and 37.

The second is incorrect because, as Strawson points out, many of us in fact satisfy the expectations and demands actually placed upon us, so many of us evidently do enjoy the capacities required to do so—and therefore, if determinism is true, it cannot be incompatible with these capacities.

The first claim is the more plausible, but it gains its plausibility from one of two ideas: either the idea that moral and interpersonal expectations or demands are like *requests* or *commands*—which can be unreasonable when directed to someone who lacks the capacity to satisfy them—or the idea that the reactive attitudes are like *punishment* or *sanction*—either of which might require, for their fairness, an adequate opportunity to avoid them. Neither is true.

A full explanation of why neither is true is more than I can take on here.[39] To give a sense of it, compare the standards of regard with the standards of good parenting or policing. The standards of parenting or policing are, themselves, neither requests nor commands (though someone may issue a request or command based on them). They do not cease to be reasonable when a particular parent or officer lacks the capacities required to satisfy them (as a request or command might). Moreover, the negative consequences that follow from failing to satisfy those standards are not, like sanction or punishment, shown unfair by a lack of opportunity to avoid. If a given parent or officer lacks the capacities required to satisfy the standards, things will go badly: feelings will be hurt and lives will be damaged. But these consequences will not be shown unfair by the fact that the parent or officer lacked the ability to satisfy the standard: If we remove a police officer from duty because they are too impetuous to be trusted in the field, the fact that that particular officer never had an opportunity to become more circumspect does not render either their removal or our distrust unfair. In the same way, being prone to resentment or indignation is not the same as making a request or command, and the standards of regard do not cease to be reasonable because a particular person lacks the capacity to satisfy them. When

39. The issues here are large. See footnotes 22 and 25, the references cited there, and footnote 34.

the standards are violated, things will go badly, feelings will be hurt, and relationships may be damaged or broken—but the wrongdoer cannot point to the hurt feelings or the broken relationships and complain of unreasonableness or unfairness. Or so I would argue.

Finally, and perhaps most to the point, in claiming that the "demands and expectations" are unreasonable or unfair unless some condition is met, the generalization strategist is in fact making a claim, entering a moral principle, about the fairness or licitness of engaged interpersonal relating (rather than a principle about the fairness or licitness of a particular request, command, sanction, or punishment). But, if some principle would, in its application, prohibit all ordinary, engaged interpersonal relating, it will be *that principle* that is unreasonable. If the pessimist thinks otherwise, he must first introduce some grounding for this claim about what is reasonable or fair (beyond the analogies with request, command, sanction, or punishment), then defend that idea even as he shows that, with the likely truth of determinism, it will undermine ordinary interpersonal relating. This will be difficult to do.

This final reply may become more clear as we turn to consider the final set of objections.

Error, Inconsistency, and Crises

The pessimist might, at this point, shift his strategy. Rather than continuing to attack Strawson's underlying social naturalism, he might instead backtrack, a bit, and point out what seems to be a serious oversight in Strawson's argument, even if we grant the social naturalism: We have been attributing to Strawson the view that some or another framework of moral and interpersonal demands will be given with the fact of human society and that it will, at least at its minimal points, be attuned to the capacities we *actually* have—attuned to what is *in fact* ordinary. Strawson could then argue that we already know that most people in fact possess the minimally required capacities. But, one might think, this is just a mistake: the socially naturalistic framework will not be attuned to what is *in fact* ordinary. Rather, at best, the framework will be attuned to what is *believed to be* ordinary. Thus, the pessimist may argue, we should grant to Strawson not

the claim that we exempt those who are incapacitated for ordinary interpersonal relationships, but rather, at best, the claim that we exempt those *we believe* to be incapacitated for what *we believe to be* ordinary relationships. Likewise, we should not grant to Strawson the claim that, when we exempt the abnormal, we do so because we believe them incapable of a kind of relating we *already know* to be actual. Rather, at best, we do so because we believe them incapable of a kind of relating we *believe* (we know) to be actual. It could turn out that we have been mistaken—such relating is not, in fact, actual. We might thus learn that no one in fact possesses the capacities our actual practice requires. If so, it seems that the generalization strategy will succeed after all; the practice will condemn itself.[40]

To put some flesh on the challenge: The pessimist might point out that we have, for centuries, thought of ourselves as spirits inhabiting a material world, endowed with something like contra-causal freedom, able to detach from, "step back" from, our circumstances and act independently. We have assumed this to be the ordinary condition. And we have, for centuries, taken the fact that someone's behavior is caused by forces outside of their control to show that they are wrongly held responsible—we have, for ages, taken that to be a reason to exempt. So it would seem that, if we learn that we are not free spirits, that we are not able to act independently of the material world, then we *also*, thereby, will learn that everyone qualifies for the exemption we have all along been granting to certain special cases. The pessimist thinks that, if we make that discovery, the system will then undermine itself: to avoid the contradiction, we ought to stop interacting in the ordinary way, despite our natural commitment.

It seems, then, that Strawson is overlooking the possibility that our practices are, themselves, based on widespread error about our own capacities. But surely Strawson should allow the possibility of such error, and surely he should also allow that such error might infect our

40. Recall there were two outstanding objections to Strawson's argument for the further point (in chapter 3, "Objections"): (1) it relies on social naturalism and (2) it presumes our reasons for exempting will not drive us into contradiction. Having defended social naturalism, we are now considering the second objection.

practices. Once he allows both, how can he rule out the possibility that we, even now, exempt people due to a condition that might be true of everyone? That is, how can he rule out the possibility that we might be guilty of exactly the kind of contradiction he imagines, when he argues against the generalization strategy? And how can he avoid the pessimist's conclusion that, faced with that kind of contradiction, we *ought* to stop interacting in the ordinary way (whether or not we can or will)?

Recall that Strawson believes that it is the job of reflection and reasoning (and of descriptive metaphysics) to reveal the principles at work within the framework of human life, in part by eliminating inconsistencies. Thus (I think) Strawson believes that the framework, although established by our practice, must be consistent: if we discover an inconsistency, we have discovered a point at which reason has revealed that we are wrong about the principles at work in our practice.[41] Strawson and the pessimist agree that our practice would be inconsistent if we exempted only some people because we believe, of them, something we know to be true of everyone. The pessimist thinks we should resolve the inconsistency by abandoning characteristically interpersonal relating. But what the pessimist imagines we ought to do is just the thing that Strawson thinks practically inconceivable, while what the pessimist imagines is unjustified (continuing in ordinary interpersonal relations) is what Strawson thinks does not require justification. Rather than accept the pessimist's conclusion that we ought to give up characteristically human relationships, Strawson concludes that we can, by reflection and reasoning, rule out the possibility that anything true of everyone is, in fact, a reason to exempt.

41. Recall, from *Skepticism and Naturalism*, "Our inescapable natural commitment is to a general frame of belief and to a general style (the inductive style) of belief-formation. But within that frame and style, the requirement of Reason, that our beliefs should form a consistent and coherent system, may be given full play. Thus . . . Hume could quite consistently proceed to frame 'rules for judging cause and effect.' Though it is Nature which commits us to inductive belief-formation in general, it is Reason which leads us to refine and elaborate our inductive canons and procedures and, in their light, to criticize, and sometimes to reject, what in detail we find ourselves naturally inclined to believe" (13–14).

Again, to put some flesh on the reply: Suppose, as imagined, that we have for centuries believed that we enjoy something like contracausal freedom, and, further, that we have thought that the fact that someone's behavior is the result of physical and psychological forces beyond their control exempts the person of responsibility for that stretch of behavior. Then we come to learn that we are not, in fact, free in the way we had imagined, that *everyone's* behavior is ultimately the result of physical and psychological forces beyond their control. On Strawson's picture, we are not, thereby, forced to conclude that everyone is exempt, that no one deserves to be blamed, or that continuing to engage in the system of demands and reactions would be unjustified. Rather, this is an occasion in which reflection has helped us to better understand the terms of our relating—reflection reveals to us the principles actually at work. In this situation, Strawson may suggest, we could, would, and should conclude, not only that we have been in error, all along, about our own freedom and metaphysical status, but *also* (in fact, therefore) that we have been in error about our own reasons for exempting.[42] Being the result of physical and psychological forces that are ultimately outside of one's control is not, in fact, what exempts those who are severely mentally ill. Rather, the person who is severely mentally ill is exempted because their conduct is far *enough* outside of their control to render them incapable of tolerably ordinary adult relationships. With his firm grip on our natural human commitment to interpersonal relating, Strawson has done us the favor of anticipating this correction and pointing out that we exempt only in the unusual cases.[43]

42. Note, we have just supplied Strawson with the missing premise for the argument from P0–P3: our beliefs will not drive us into contradiction. See chapter 3, footnote 6.

43. Notice that this line of thought has the perhaps surprising consequence that our practice will, after all, be attuned to the actual (and not just to our beliefs about what is actual). See, too, the following quotation from Strawson, "Replies," 265: "There is a quite general ambiguity in the notion of 'our ordinary concept' of whatever it may be. Should the lineaments of such a concept be drawn from its use, from our ordinary *practice*, or should we add the reflective accretions, however confused, which, naturally or historically, gather round it? The distinction is hardly clear-cut;

I hope we have now arrived at a better understanding of the role of Strawson's social naturalism in his response to the pessimist. By, on the one hand, understanding our participation in some or another moral and interpersonal framework as a natural commitment given with the fact of human society, not open to questions of justification, and, on the other, allowing that the framework is open to correction by reflection and revision, Strawson can rest assured that the practice will not embed the inconsistency imagined by the pessimist. The further point is implicit in his underlying picture, just as he said. There are no standards on the legitimacy of ordinary relating that are universally unmet.

The pessimist may object to this most recent line of thought, saying it is simply implausible to think that, in light of the discovery that we are not free spirits, we have *also* discovered that our reasons for exempting were not as we thought. The moral principles are, after all, on a naturalist picture, of our own making—how could we have been confused about them? And, more to the point, why think the framework could not, itself, simply be inconsistent?

I believe Strawson's answer is the one above—the answer of a descriptive metaphysician. But I will close by considering a nearby, more concessive, but also more radical reply that might be made by a social naturalist of a more existentialist bent. One might point out that, even if the principles run no deeper than our practices, so to speak,[44] and even if our practices *had* incorporated into them false principles, we are not then beholden to those principles when their falsehood comes to light. When an inconsistency threatens to undermine the entire system, no particular part of the system must retain special authority. Rather, the crisis raises a problem, and we both must and may solve that problem as best we can—though we are not bound to solve it in any particular way, in advance (that is what makes it a crisis).[45]

---

but where it can be made, I prefer the first alternative." I am grateful to David Beglin for bringing my attention to this quotation.

44. Strawson resists this in *Skepticism and Naturalism*, chap. 4.

45. I have, in effect, led Strawson to Sartre's young man: the standards given to us have given out; they offer no clear guidance. And I am making Sartre's suggestion: "Choose, that is to say, invent" (Jean-Paul Sartre, "Existentialism Is a Humanism"

I will illustrate this last by an analogy with baseball—or, rather, with a game also called "baseball" and much like our baseball, played in a nearby possible world. In this nearby world, the rules of baseball explicitly stipulate that if performance-enhancing drugs are used in a game, that game is to be struck from the record: its scores and statistics are not kept; it contributes not at all to the teams' or players' standings. And, in this world, "performance-enhancing drug" is explicitly defined as any chemical that improves performance but has detrimental effects on players' overall, long-term health. Moreover, as it happens, in this world, in every game that has ever been played, many players have been using chewing tobacco on the field. It is then discovered that chewing tobacco fits the definition of a performance-enhancing drug.

I hope the analogy is clear: The use of performance-enhancing drugs is a condition that is explicitly stipulated, by the rules of the game, to undermine the game.[46] The discovery is then made that this condition has always held. Thus, the reasoning of the generalization strategy would require us to conclude that, in light of the medical discovery, no game has ever been won or lost (or, better, no game has ever *really* been won or lost): we must wipe the records and reset the stats.

I hope it is also clear that we would not, in the scenario imagined, be unquestionably bound by the explicitly established rules. Rather, we would be faced with a problem—we have been brought to a kind of crisis and we need to consider how best to resolve it. To be sure,

---

[lecture, 1945]). But (like Sartre, I would argue) I am not imagining we can resolve the crisis by "radical choice," if that is a choice made by fiat, a mere picking. Rather, we must undertake serious reflection of a familiar kind—a kind of reflection that is, one might think, the province of philosophy. Charles Taylor describes the situation eloquently in the final passages of "Responsibility for Self," in *The Identities of Persons*, ed. Amélia O. Rorty (Berkeley: University of California Press, 1976), 281–99. (I will, with some trepidation, note the resonance between the idea that practice precedes standards and the thought that existence precedes essence.)

46. This rule about performance-enhancing drugs is a condition that renders a game *illegitimate* without also rendering play *impossible*—it is an example of the sort that is hard to find in the moral and interpersonal realm. See chapter 3, footnote 5.

we might, in the end, do exactly as the generalization strategy would suggest: we might stick with the old rules and decide that, heretofore, no game has ever (really) been won or lost—wipe the records and reset the stats, ditch the dip and start over. But we might, instead, revise our definition of performance-enhancing drug and carry on as before—roughly as Strawson suggested, above. Or, we might opt for a "grueish" solution: we might ignore the use of chewing tobacco up until now, while prohibiting it in the future.[47] Each of these is open, with pros and cons. There may not be a uniquely right answer. But we will have to settle on one. To think that the rules we have heretofore been following must hold sway, even in the crisis, is to be guilty of a kind of bad faith. (Likewise to think that we can just opt for any alternative, without reason.)[48]

The opponent might feel the analogy makes vivid the weakness in this position: morality is not a game and we are not at liberty to

47. On "grue," see Nelson Goodman, *Fact, Fiction, and Forecast* (Cambridge, MA: Harvard University, 1955).

48. Note that because we cannot ditch determinism in the way we can ditch dip, neither wiping the records and starting over fresh nor the grueish option is available. Unless we reinterpret the rules, we would need to give up the game. But, Strawson says, giving it up is not only not realistically possible but also not something that could be required.

Also of interest: in the case of ethics, unlike the case of baseball, there is nobody to make the decision. Rather, faced with crises, "we" come to something like a decision, through our practice. I find an interesting comparison in the current evolution of the English third-person singular pronoun. The usual menu of "he," "she," and "it" seemed to leave us without a gender-neutral way to refer to a single person in the third person. The problem could have been addressed in a number of ways—e.g., by hearing "it" as an acceptable way to refer to persons or by introducing a new word (as, I understand, happened in some English dialects). It seems, though, that we are addressing the problem by using "they" as a singular pronoun—and, at the same time, introducing a host of new singular verb forms, which (confusingly) are identical to the plural verb forms. It is not the solution I would have chosen, but it seems to be something "we" have done—something like a choice we have made, without choosing as a collective body. (This note will prompt some to point out that "they" has been so used for centuries in works of great literature. True. Please add that information wherever needed.)

change it as we like. But, of course, both Strawson and the more existentialist social naturalist agree with both claims. The disagreement runs deeper—it seems to me to be the disagreement we first considered, the rejection of social naturalism.

What would it take for the pessimist to support his pessimism? He would first have to locate plausible standards on the legitimacy of characteristically engaged interpersonal relating that will be unmet if determinism is true, and then he would need to argue that those standards will continue to apply, even as we learn our mistake about our metaphysical status. It is hard to see what could give the standards that sort of authority. To think they have that kind of authority would be to think, so to speak, that man was made for the Sabbath, rather than the Sabbath for, and by, man.

Conclusion

LET US NOW RETRACE from the beginning. Strawson is thinking, from the start, that we have a natural, nonrational commitment to engaging in characteristically interpersonal relationships—he is thinking that the quality of others' wills toward us matters to us, that we put some (or another) set of demands on the quality of others' wills, that we will react in certain (or other) ways when those demands are violated or exceeded, and that this fact about us is given with human society, not something for which there are or need to be reasons. Thus we will, as a matter of fact, typically engage with others in the characteristically interpersonal way. Moreover, the exact demands and reactions, the details of our ordinary interpersonal relationships, are themselves a natural fact, a product of life as it actually happens. And so the detail of our system will be sensitive to typical human capacities and typical circumstances. We sometimes suspend characteristically interpersonal relating when circumstances are extreme or when someone is incapable of engaging in it. More curiously, we sometimes exercise our "resource" to opt out of such relating, just to avoid the strains of involvement, or for therapeutic purposes, or out of curiosity. But it could not be the case that everyone is in unusual or extreme circumstances, nor could it be that everyone is incapable of ordinary relating—to say either is to assert a contradiction. Thus, the only condition worth considering is whether we could or should come to exercise our resource all the time, and so give up characteristically interpersonal relating. Could we do so? While saying so involves no contradiction, it seems practically inconceivable that we *would* do so. Should we do so? Engaging

in characteristically interpersonal relating is not done for reasons, nor something that requires justification, and so the question is idle—we need not take it seriously. But there is a further point that can now be made explicit: we can know, in advance, that being determined* (the sense of "determined" at issue in the thesis of determinism) is not a reason to exempt, because (given Strawson's social naturalism) we can know that the principles that govern moral and interpersonal relating will not include the contradiction that would require: if we discover an apparent contradiction in our principles, we have discovered that we ought to revise our understanding of those principles. Thus we can rest assured that nothing true of everyone will provide a reason to exempt. If this is not enough, we can also note that, even if we *could* face a choice about whether to abandon our commitment to characteristically interpersonal relating, a choice Strawson finds unreal, our reasons for making the choice would not be the kind that motivate the pessimist—they would not be moral reasons. If we were to decide that no one should be held responsible, we would have to make that decision for reasons that concern not questions of desert or justice, but the gains and losses to human life. And, as Strawson says, the truth or falsity of determinism would have no bearing on *this* choice.

P. F. Strawson, "Freedom and Resentment"*

I

Some philosophers say they do not know what the thesis of determinism is. Others say, or imply, that they do know what it is. Of these, some—the pessimists perhaps—hold that if the thesis is true, then the concepts of moral obligation and responsibility really have no application, and the practices of punishing and blaming, of expressing moral condemnation and approval, are really unjustified. Others—the optimists perhaps—hold that these concepts and practices in no way lose their raison d'être if the thesis of determinism is true. Some hold even that the justification of these concepts and practices requires the truth of the thesis. There is another opinion which is less frequently voiced: the opinion, it might be said, of the genuine moral sceptic. This is that the notions of moral guilt, of blame, of moral responsibility are inherently confused and that we can see this to be so if we consider the consequences either of the truth of determinism or of its falsity. The holders of this opinion agree with the pessimists that these notions lack application if determinism is true, and add simply that they also lack it if determinism is false. If I am asked which of these parties I belong to, I must say it is the first of all, the party of those who do not know what the thesis of determinism is. But this does not stop me from having some sympathy with the others, and a wish to reconcile

* From *Proceedings of the British Academy*, vol xlviii (1962) pp. 1–25.

them. Should not ignorance, rationally, inhibit such sympathies? Well, of course, though darkling, one has some inkling—some notion of what sort of thing is being talked about. This lecture is intended as a move towards reconciliation; so is likely to seem wrongheaded to everyone.

But can there be any possibility of reconciliation between such clearly opposed positions as those of pessimists and optimists about determinism? Well, there might be a formal withdrawal on one side in return for a substantial concession on the other. Thus, suppose the optimist's position were put like this: (1) the facts as we know them do not show determinism to be false; (2) the facts as we know them supply an adequate basis for the concepts and practices which the pessimist feels to be imperilled by the possibility of determinism's truth. Now it might be that the optimist is right in this, but is apt to give an inadequate account of the facts as we know them, and of how they constitute an adequate basis for the problematic concepts and practices; that the reasons he gives for the adequacy of the basis are themselves inadequate and leave out something vital. It might be that the pessimist is rightly anxious to get this vital thing back and, in the grip of his anxiety, feels he has to go beyond the facts as we know them; feels that the vital thing can be secure only if, beyond the facts as we know them, there is the further fact that determinism is false. Might *he* not be brought to make a formal withdrawal in return for a vital concession?

II

Let me enlarge very briefly on this, by way of preliminary only. Some optimists about determinism point to the efficacy of the practices of punishment, and of moral condemnation and approval, in regulating behaviour in socially desirable ways.[1] In the fact of their efficacy, they suggest, is an adequate basis for these practices; and this fact certainly does not show determinism to be false. To this the pessimists reply, all in a rush, that *just* punishment and *moral*

1. Cf. P. H. Nowell-Smith, 'Freewill and Moral Responsibility', *Mind*, 1948.

condemnation imply moral guilt and guilt implies moral responsibility and moral responsibility implies freedom and freedom implies the falsity of determinism. And to this the optimists are wont to reply in turn that it is true that these practices require freedom in a sense, and the existence of freedom in this sense is one of the facts as we know them. But what 'freedom' means here is nothing but the absence of certain conditions the presence of which would make moral condemnation or punishment inappropriate. They have in mind conditions like compulsion by another, or innate incapacity, or insanity, or other less extreme forms of psychological disorder, or the existence of circumstances in which the making of any other choice would be morally inadmissible or would be too much to expect of any man. To this list they are constrained to add other factors which, without exactly being limitations of freedom, may also make moral condemnation or punishment inappropriate or mitigate their force: as some forms of ignorance, mistake, or accident. And the general reason why moral condemnation or punishment are inappropriate when these factors or conditions are present is held to be that the practices in question will be generally efficacious means of regulating behaviour in desirable ways only in cases where these factors are *not* present. Now the pessimist admits that the facts as we know them include the existence of freedom, the occurrence of cases of free action, in the negative sense which the optimist concedes; and admits, or rather insists, that the existence of freedom in this sense is compatible with the truth of determinism. Then what does the pessimist find missing? When he tries to answer this question, his language is apt to alternate between the very familiar and the very unfamiliar.[2] Thus he may say, familiarly enough, that the man who is the subject of justified punishment, blame or moral condemnation must really *deserve* it; and then add, perhaps, that, in the case at least where he is blamed for a positive act rather than an omission, the condition of his really deserving blame is something that goes beyond the negative freedoms that the optimist concedes. It is, say, a genuinely free identification of the will with

2. As Nowell-Smith pointed out in a later article, 'Determinists and Libertarians', *Mind*, 1954.

the act. And this is the condition that is incompatible with the truth of determinism.

The conventional, but conciliatory, optimist need not give up yet. He may say: Well, people often decide to do things, really intend to do what they do, know just what they're doing in doing it; the reasons they think they have for doing what they do, often really are their reasons and not their rationalizations. These facts, too, are included in the facts as we know them. If this is what you mean by freedom—by the identification of the will with the act—then freedom may again be conceded. But again the concession is compatible with the truth of the determinist thesis. For it would not follow from that thesis that nobody decides to do anything; that nobody ever does anything intentionally; that it is false that people sometimes know perfectly well what they are doing. I tried to define freedom negatively. You want to give it a more positive look. But it comes to the same thing. Nobody denies freedom in this sense, or these senses, and nobody claims that the existence of freedom in these senses shows determinism to be false.

But it is here that the lacuna in the optimistic story can be made to show. For the pessimist may be supposed to ask: But *why* does freedom in this sense justify blame, etc.? You turn towards me first the negative, and then the positive, faces of a freedom which nobody challenges. But the only reason you have given for the practices of moral condemnation and punishment in cases where this freedom is present is the efficacy of these practices in regulating behaviour in socially desirable ways. But this is not a sufficient basis, it is not even the right *sort* of basis, for these practices as we understand them.

Now my optimist, being the sort of man he is, is not likely to invoke an intuition of fittingness at this point. So he really has no more to say. And my pessimist, being the sort of man he is, has only one more thing to say; and that is that the admissibility of these practices, as we understand them, demands another kind of freedom, the kind that in turn demands the falsity of the thesis of determinism. But might we not induce the pessimist to give up saying this by giving the optimist something more to say?

III

I have mentioned punishing and moral condemnation and approval; and it is in connection with these practices or attitudes that the issue between optimists and pessimists—or, if one is a pessimist, the issue between determinists and libertarians—is felt to be particularly important. But it is not of these practices and attitudes that I propose, at first, to speak. These practices or attitudes permit, where they do not imply, a certain detachment from the actions or agents which are their objects. I want to speak, at least at first, of something else: of the non-detached attitudes and reactions of people directly involved in transactions with each other; of the attitudes and reactions of offended parties and beneficiaries; of such things as gratitude, resentment, forgiveness, love, and hurt feelings. Perhaps something like the issue between optimists and pessimists arises in this neighbouring field too; and since this field is less crowded with disputants, the issue might here be easier to settle; and if it is settled here, then it might become easier to settle it in the disputant-crowded field.

What I have to say consists largely of commonplaces. So my language, like that of commonplaces generally, will be quite unscientific and imprecise. The central commonplace that I want to insist on is the very great importance that we attach to the attitudes and intentions towards us of other human beings, and the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs about these attitudes and intentions. I can give no simple description of the field of phenomena at the centre of which stands this commonplace truth; for the field is too complex. Much imaginative literature is devoted to exploring its complexities; and we have a large vocabulary for the purpose. There are simplifying styles of handling it in a general way. Thus we may, like La Rochefoucauld, put self-love or self-esteem or vanity at the centre of the picture and point out how it may be caressed by the esteem, or wounded by the indifference or contempt, of others. We might speak, in another jargon, of the need for love, and the loss of security which results from its withdrawal; or, in another, of human self-respect and its connection with the recognition of the individual's dignity. These simplifications are of use to me only if they help to emphasize how much we actually mind, how

much it matters to us, whether the actions of other people—and particularly of *some* other people—reflect attitudes towards us of goodwill, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other. If someone treads on my hand accidentally, while trying to help me, the pain may be no less acute than if he treads on it in contemptuous disregard of my existence or with a malevolent wish to injure me. But I shall generally feel in the second case a kind and degree of resentment that I shall not feel in the first. If someone's actions help me to some benefit I desire, then I am benefited in any case; but if he intended them so to benefit me because of his general goodwill towards me, I shall reasonably feel a gratitude which I should not feel at all if the benefit was an incidental consequence, unintended or even regretted by him, of some plan of action with a different aim.

These examples are of actions which confer benefits or inflict injuries over and above any conferred or inflicted by the mere manifestation of attitude and intention themselves. We should consider also in how much of our behaviour the benefit or injury resides mainly or entirely in the manifestation of attitude itself. So it is with good manners, and much of what we call kindness, on the one hand; with deliberate rudeness, studied indifference, or insult on the other.

Besides resentment and gratitude, I mentioned just now forgiveness. This is a rather unfashionable subject in moral philosophy at present; but to be forgiven is something we sometimes ask, and forgiving is something we sometimes say we do. To ask to be forgiven is in part to acknowledge that the attitude displayed in our actions was such as might properly be resented and in part to repudiate that attitude for the future (or at least for the immediate future); and to forgive is to accept the repudiation and to forswear the resentment.

We should think of the many different kinds of relationship which we can have with other people—as sharers of a common interest; as members of the same family; as colleagues; as friends; as lovers; as chance parties to an enormous range of transactions and encounters. Then we should think, in each of these connections in turn, and in others, of the kind of importance we attach to the attitudes and intentions towards us of those who stand in these relationships to us, and of the kinds of *reactive* attitudes and feelings to which we ourselves

are prone. In general, we demand some degree of goodwill or regard on the part of those who stand in these relationships to us, though the forms we require it to take vary widely in different connections. The range and intensity of our *reactive* attitudes towards goodwill, its absence or its opposite vary no less widely. I have mentioned, specifically, resentment and gratitude; and they are a usefully opposed pair. But, of course, there is a whole continuum of reactive attitude and feeling stretching on both sides of these and—the most comfortable area—in between them.

The object of these commonplaces is to try to keep before our minds something it is easy to forget when we are engaged in philosophy, especially in our cool, contemporary style, viz. what it is actually like to be involved in ordinary interpersonal relationships, ranging from the most intimate to the most casual.

IV

It is one thing to ask about the general causes of these reactive attitudes I have alluded to; it is another to ask about the variations to which they are subject, the particular conditions in which they do or do not seem natural or reasonable or appropriate; and it is a third thing to ask what it would be like, what it is like, not to suffer them. I am not much concerned with the first question; but I am with the second; and perhaps even more with the third.

Let us consider, then, occasions for resentment: situations in which one person is offended or injured by the action of another and in which—in the absence of special considerations—the offended person might naturally or normally be expected to feel resentment. Then let us consider what sorts of special considerations might be expected to modify or mollify this feeling or remove it altogether. It needs no saying now how multifarious these considerations are. But, for my purpose, I think they can be roughly divided into two kinds. To the first group belong all those which might give occasion for the employment of such expressions as 'He didn't mean to', 'He hadn't realized', 'He didn't know'; and also all those which might give occasion for the use of the phrase 'He couldn't help it', when this is supported by such phrases as 'He was pushed', 'He had to do it', 'It was

the only way', 'They left him no alternative', etc. Obviously these various pleas, and the kinds of situations in which they would be appropriate, differ from each other in striking and important ways. But for my present purpose they have something still more important in common. None of them invites us to suspend towards the agent, either at the time of his action or in general, our ordinary reactive attitudes. They do not invite us to view the *agent* as one in respect of whom these attitudes are in any way inappropriate. They invite us to view the *injury* as one in respect of which a particular one of these attitudes is inappropriate. They do not invite us to see the *agent* as other than a fully responsible agent. They invite us to see the *injury* as one for which he was not fully, or at all, responsible. They do not suggest that the agent is in any way an inappropriate object of that kind of demand for goodwill or regard which is reflected in our ordinary reactive attitudes. They suggest instead that the fact of injury was not in this case incompatible with that demand's being fulfilled, that the fact of injury was quite consistent with the agent's attitude and intentions being just what we demand they should be.[3] The agent was just ignorant of the injury he was causing, or had lost his balance through being pushed or had reluctantly to cause the injury for reasons which acceptably override his reluctance. The offering of such pleas by the agent and their acceptance by the sufferer is something in no way opposed to, or outside the context of, ordinary interpersonal relationships and the manifestation of ordinary reactive attitudes. Since things go wrong and situations are complicated, it is an essential and integral element in the transactions which are the life of these relationships.

The second group of considerations is very different. I shall take them in two subgroups of which the first is far less important than the second. In connection with the first subgroup we may think of such statements as 'He wasn't himself', 'He has been under very great strain recently', 'He was acting under post-hypnotic suggestion'; in

3. Perhaps not in every case just what we demand they should be, but in any case not just what we demand they should not be. For my present purpose these differences do not matter.

connection with the second, we may think of 'He's only a child', 'He's a hopeless schizophrenic', 'His mind has been systematically perverted', 'That's purely compulsive behaviour on his part'. Such pleas as these do, as pleas of my first general group do not, invite us to suspend our ordinary reactive attitudes towards the agent, either at the time of his action or all the time. They do not invite us to see the agent's action in a way consistent with the full retention of ordinary inter-personal attitudes and merely inconsistent with one particular attitude. They invite us to view the agent himself in a different light from the light in which we should normally view one who has acted as he has acted. I shall not linger over the first subgroup of cases. Though they perhaps raise, in the short term, questions akin to those raised, in the long term, by the second subgroup, we may dismiss them without considering those questions by taking that admirably suggestive phrase, 'He wasn't himself', with the seriousness that—for all its being logically comic—it deserves. 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 as he does under abnormal stresses, as we should have felt towards him had he acted as he did under normal stresses.

The second and more important subgroup of cases allows that the circumstances were normal, but presents the agent as psychologically abnormal—or as morally undeveloped. The agent was himself; but he is warped or deranged, neurotic or just a child. When we see someone in such a light as this, all our reactive attitudes tend to be profoundly modified. I must deal here in crude dichotomies and ignore the ever-interesting and ever-illuminating varieties of case. What I want to contrast is the attitude (or range of attitudes) of involvement or participation in a human relationship, on the one hand, and what might be called the objective attitude (or range of attitudes) to another human being, on the other. Even in the same situation, I must add, they are not altogether *exclusive* of each other; but they are, profoundly, *opposed* to each other. To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps

precautionary account, of; to be managed or handled or cured or trained; perhaps simply to be avoided, though *this* gerundive is not peculiar to cases of objectivity of attitude. The objective attitude may be emotionally toned in many ways, but not in all ways: it may include repulsion or fear, it may include pity or even love, though not all kinds of love. But it cannot include the range of reactive feelings and attitudes which belong to involvement or participation with others in inter-personal human relationships; it cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other. If your attitude towards someone is wholly objective, then though you may fight him, you cannot quarrel with him, and though you may talk to him, even negotiate with him, you cannot reason with him. You can at most pretend to quarrel, or to reason, with him.

Seeing someone, then, as warped or deranged or compulsive in behaviour or peculiarly unfortunate in his formative circumstances—seeing someone so tends, at least to some extent, to set him apart from normal participant reactive attitudes on the part of one who so sees him, tends to promote, at least in the civilized, objective attitudes. But there is something curious to add to this. The objective attitude is not only something we naturally tend to fall into in cases like these, where participant attitudes are partially or wholly inhibited by abnormalities or by immaturity. It is also something which is available as a resource in other cases too. We look with an objective eye on the compulsive behaviour of the neurotic or the tiresome behaviour of a very young child, thinking in terms of treatment or training. But we *can* sometimes look with something like the same eye on the behaviour of the normal and the mature. We *have* this resource and can sometimes use it; as a refuge, say, from the strains of involvement; or as an aid to policy; or simply out of intellectual curiosity. Being human, we cannot, in the normal case, do this for long, or altogether. If the strains of involvement, say, continue to be too great, then we have to do something else—like severing a relationship. But what is above all interesting is the tension there is, in us, between the participant attitude and the objective attitude. One is tempted to say: between our humanity and our intelligence. But to say this would be to distort both notions.

What I have called the participant reactive attitudes are essentially natural human reactions to the good or ill will or indifference of others towards us, as displayed in *their* attitudes and actions. The question we have to ask is: What effect would, or should, the acceptance of the truth of a general thesis of determinism have upon these reactive attitudes? More specifically, would, or should, the acceptance of the truth of the thesis lead to the decay or the repudiation of all such attitudes? Would, or should, it mean the end of gratitude, resentment, and forgiveness; of all reciprocated adult loves; of all the essentially *personal* antagonisms?

But how can I answer, or even pose, this question without knowing *exactly* what the thesis of determinism is? Well, there is one thing we do know; that if there is a coherent thesis of determinism, then there must be a sense of 'determined' such that, if that thesis is true, then all behaviour whatever is determined in that sense. Remembering this, we can consider at least what possibilities lie formally open; and then perhaps we shall see that the question can be answered *without* knowing exactly what the thesis of determinism is. We can consider what possibilities lie open because we have already before us an account of the ways in which particular reactive attitudes, or reactive attitudes in general, may be, and, sometimes, we judge, should be, inhibited. Thus I considered earlier a group of considerations which tend to inhibit, and, we judge, should inhibit, resentment, in particular cases of an agent causing an injury, without inhibiting reactive attitudes in general towards that agent. Obviously this group of considerations cannot strictly bear upon our question; for that question concerns reactive attitudes in general. But resentment has a particular interest; so it is worth adding that it has never been claimed as a consequence of the truth of determinism that one or another of *these* considerations was operative in every case of an injury being caused by an agent; that it would follow from the truth of determinism that anyone who caused an injury *either* was quite simply ignorant of causing it *or* had acceptably overriding reasons for acquiescing reluctantly in causing it *or* . . . , etc. The prevalence of this happy state of affairs would not be a consequence of the reign of universal determinism, but of the reign of universal goodwill. We cannot, then, find here the possibility of an affirmative answer to our question, even for the particular case of resentment.

Next, I remarked that the participant attitude, and the personal reactive attitudes in general, tend to give place, and it is judged by the civilized should give place, to objective attitudes, just in so far as the agent is seen as excluded from ordinary adult human relationships by deep-rooted. psychological abnormality—or simply by being a child. But it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition.

Now this dismissal might seem altogether too facile; and so, in a sense, it is. But whatever is too quickly dismissed in this dismissal is allowed for in the only possible form of affirmative answer that remains. We can sometimes, and in part, I have remarked, look on the normal (those we rate as 'normal') in the objective way in which we have learned to look on certain classified cases of abnormality. And our question reduces to this: could, or should, the acceptance of the determinist thesis lead us always to look on everyone exclusively in this way? For this is the only condition worth considering under which the acceptance of determinism could lead to the decay or repudiation of participant reactive attitudes.

It does not seem to be self-contradictory to suppose that this might happen. So I suppose we must say that it is not absolutely inconceivable that it should happen. But I am strongly inclined to think that it is, for us as we are, practically inconceivable. The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical 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 inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question.

This, then, is a part of the reply to our question. A sustained objectivity of inter-personal attitude, and the human isolation which that would entail, does not seem to be something of which human beings would be capable, even if some general truth were a theoretical ground for it. But this is not all. There is a further point, implicit in the foregoing, which must be made explicit. Exceptionally, I have said, we can have direct dealings with human beings without any

degree of personal involvement, treating them simply as creatures to be handled in our own interest, or our side's, or society's—or even theirs. In the extreme case of the mentally deranged, it is easy to see the connection between the possibility of a wholly objective attitude and the impossibility of what we understand by ordinary interpersonal relationships. Given this latter impossibility, no other civilized attitude is available than that of viewing the deranged person simply as something to be understood and controlled in the most desirable fashion. To view him as outside the reach of personal relationships is already, for the civilized, to view him in this way. For reasons of policy or self-protection we may have occasion, perhaps temporary, to adopt a fundamentally similar attitude to a 'normal' human being; to concentrate, that is, on understanding 'how he works', with a view to determining our policy accordingly, or to finding in that very understanding a relief from the strains of involvement. Now it is certainly true that in the case of the abnormal, though not in the case of the normal, our adoption of the objective attitude is a consequence of our viewing the agent as *incapacitated* in some or all respects for ordinary interpersonal relationships. He is thus incapacitated, perhaps, by the fact that his picture of reality is pure fantasy, that he does not, in a sense, live in the real world at all; or by the fact that his behaviour is, in part, an unrealistic acting out of unconscious purposes; or by the fact that he is an idiot, or a moral idiot. But there is something else which, *because* this is true, is equally certainly *not* true. And that is that there is a sense of 'determined' such that (1) if determinism is true, all behaviour is determined in this sense, and (2) determinism might be true, i.e. it is not inconsistent with the facts as we know them to suppose that all behaviour might be determined in this sense, and (3) our adoption of the objective attitude towards the abnormal is the result of a prior embracing of the belief that the behaviour, or the relevant stretch of behaviour, of the human being in question *is* determined in this sense. Neither in the case of the normal, then, nor in the case of the abnormal is it true that, when we adopt an objective attitude, we do so *because* we hold such a belief. So my answer has two parts. The first is that we cannot, as we are, seriously envisage ourselves adopting a thoroughgoing objectivity of attitude to others as a result of theoretical conviction of the

truth of determinism; and the second is that when we do in fact adopt such an attitude in a particular case, our doing so is not the consequence of a theoretical conviction which might be expressed as 'Determinism in this case', but is a consequence of our abandoning, for different reasons in different cases, the ordinary interpersonal attitudes.

It might be said that all this leaves the real question unanswered, and that we cannot hope to answer it without knowing exactly what the thesis of determinism is. For the real question is not a question about what we actually do, or why we do it. It is not even a question about what we would *in fact* do if a certain theoretical conviction gained general acceptance. It is a question about what it would be *rational* to do if determinism were true, a question about the rational justification of ordinary inter-personal attitudes in general. To this I shall reply, first, that such a question could seem real only to one who had utterly failed to grasp the purport of the preceding answer, the fact of our natural human commitment to ordinary interpersonal attitudes. This commitment is part of the general framework of human life, not something that can come up for review as particular cases can come up for review within this general framework. And I shall reply, second, that if we could imagine what we cannot have, viz., a choice in this matter, then we could choose rationally only in the light of an assessment of the gains and losses to human life, its enrichment or impoverishment; and the truth or falsity of a general thesis of determinism would not bear on the rationality of *this* choice.[4]

4. The question, then, of the connection between rationality and the adoption of the objective attitude to others is misposed when it is made to seem dependent on the issue of determinism. But there is another question which should be raised, if only to distinguish it from the misposed question. Quite apart from the issue of determinism, might it not be said that we should be nearer to being purely rational creatures in proportion as our relation to others was in fact dominated by the objective attitude? I think this might be said; only it would have to be added, once more, that if such a choice were possible, it would not necessarily be rational to choose to be more purely rational than we are.

V

The point of this discussion of the reactive attitudes in their relation—or lack of it—to the thesis of determinism was to bring us, if possible, nearer to a position of compromise in a more usual area of debate. We are not now to discuss reactive attitudes which are essentially those of offended parties or beneficiaries. We are to discuss reactive attitudes which are essentially not those, or only incidentally are those, of offended parties or beneficiaries, but are nevertheless, I shall claim, kindred attitudes to those I have discussed. I put resentment in the centre of the previous discussion. I shall put moral indignation—or, more weakly, moral disapprobation—in the centre of this one.

The reactive attitudes I have so far discussed are essentially reactions to the quality of others' wills towards us, as manifested in their behaviour: to their good or ill will or indifference or lack of concern. Thus resentment, or what I have called resentment, is a reaction to injury or indifference. The reactive attitudes I have now to discuss might be described as the sympathetic or vicarious or impersonal or disinterested or generalized analogues of the reactive attitudes I have already discussed. They are reactions to the qualities of others' wills, not towards ourselves, but towards others. Because of this impersonal or vicarious character, we give them different names. Thus one who experiences the vicarious analogue of resentment is said to be indignant or disapproving, or morally indignant or disapproving. What we have here is, as it were, resentment on behalf of another, where one's own interest and dignity are not involved; and it is this impersonal or vicarious character of the attitude, added to its others, which entitle it to the qualification 'moral'. Both my description of, and my name for, these attitudes are, in one important respect, a little misleading. It is not that these attitudes are essentially vicarious—one can feel indignation on one's own account—but that they are essentially capable of being vicarious. But I shall retain the name for the sake of its suggestiveness; and I hope that what is misleading about it will be corrected in what follows.

The personal reactive attitudes rest on, and reflect, an expectation of, and demand for, the manifestation of a certain degree of goodwill

or regard on the part of other human beings towards ourselves; or at least on the expectation of, and demand for, an absence of the manifestation of active ill will or indifferent disregard. (What will, in particular cases, *count* as manifestations of good or ill will or disregard will vary in accordance with the particular relationship in which we stand to another human being.) The generalized or vicarious analogues of the personal reactive attitudes rest on, and reflect, exactly the same expectation or demand in a generalized form; they rest on, or reflect, that is, the demand for the manifestation of a reasonable degree of goodwill or regard, on the part of others, not simply towards oneself, but towards all those on whose behalf moral indignation may be felt, i.e. as we now think, towards all men. The generalized and non-generalized forms of demand, and the vicarious and personal reactive attitudes which rest upon, and reflect, them are connected not merely logically. They are connected humanly; and not merely with each other. They are connected also with yet another set of attitudes which I must mention now in order to complete the picture. I have considered from two points of view the demands we make on others and our reactions to their possibly injurious actions. These were the points of view of one whose interest was directly involved (who suffers, say, the injury) and of others whose interest was not directly involved (who do not themselves suffer the injury). Thus I have spoken of personal reactive attitudes in the first connection and of their vicarious analogues in the second. But the picture is not complete unless we consider also the correlates of these attitudes on the part of those on whom the demands are made, on the part of the agents. Just as there are personal and vicarious reactive attitudes associated with demands on others for oneself and demands on others for others, so there are self-reactive attitudes associated with demands on oneself for others. And here we have to mention such phenomena as feeling bound or obliged (the 'sense of obligation'); feeling compunction; feeling guilty or remorseful or at least responsible; and the more complicated phenomenon of shame.

All these three types of attitude are humanly connected. One who manifested the personal reactive attitudes in a high degree but showed no inclination at all to their vicarious analogues would appear as an abnormal case of moral egocentricity, as a kind of moral

solipsist. Let him be supposed fully to acknowledge the claims to regard that others had on him, to be susceptible of the whole range of self-reactive attitudes. He would then see himself as unique both as one (*the* one) who had a general claim on human regard and as one (*the* one) on whom human beings in general had such a claim. This would be a kind of moral solipsism. But it is barely more than a conceptual possibility; if it is that. In general, though within varying limits, we demand of others for others, as well as of ourselves for others, something of the regard which we demand of others for ourselves. Can we imagine, besides that of the moral solipsist, any other case of one or two of these three types of attitude being fully developed, but quite unaccompanied by any trace, however slight, of the remaining two or one? If we can, then we imagine something far below or far above the level of our common humanity—a moral idiot or a saint. For all these types of attitude alike have common roots in our human nature and our membership of human communities.

Now, as of the personal reactive attitudes, so of their vicarious analogues, we must ask in what ways, and by what considerations, they tend to be inhibited. Both types of attitude involve, or express, a certain sort of demand for inter-personal regard. The fact of injury constitutes a prima facie appearance of this demand's being flouted or unfulfilled. We saw, in the case of resentment, how one class of considerations may show this appearance to be mere appearance, and hence inhibit resentment, *without* inhibiting, or displacing, the sort of demand of which resentment can be an expression, without in any way tending to make us suspend our ordinary inter-personal attitudes to the agent. Considerations of this class operate in just the same way, for just the same reasons, in connection with moral disapprobation or indignation; they inhibit indignation without in any way inhibiting the sort of demand on the agent of which indignation can be an expression, the range of attitudes towards him to which it belongs. But in this connection we may express the facts with a new emphasis. We may say, stressing the moral, the generalized aspect of the demand: considerations of this group have no tendency to make us see the agent as other than a morally responsible agent; they simply make us see the injury as one for which he was not morally responsible. The offering and acceptance of such exculpatory pleas as are here in

question in no way detracts in our eyes from the agent's status as a term of moral relationships. On the contrary, since things go wrong and situations are complicated, it is an essential part of the life of such relationships.

But suppose we see the agent in a different light: as one whose picture of the world is an insane delusion; or as one whose behaviour, or a part of whose behaviour, is unintelligible to us, perhaps even to him, in terms of conscious purposes, and intelligible only in terms of unconscious purposes; or even, perhaps, as one wholly impervious to the self-reactive attitudes I spoke of, wholly lacking, as we say, in moral sense. Seeing an agent in such a light as this tends, I said, to inhibit resentment in a wholly different way. It tends to inhibit resentment because it tends to inhibit ordinary interpersonal attitudes in general, and the kind of demand and expectation which those attitudes involve; and tends to promote instead the purely objective view of the agent as one posing problems simply of intellectual understanding, management, treatment, and control. Again the parallel holds for those generalized or moral attitudes towards the agent which we are now concerned with. The same abnormal light which shows the agent to us as one in respect of whom the personal attitudes, the personal demand, are to be suspended, shows him to us also as one in respect of whom the impersonal attitudes, the generalized demand, are to be suspended. Only, abstracting now from direct personal interest, we may express the facts with a new emphasis. We may say: to the extent to which the agent is seen in this light, he is not seen as one on whom demands and expectations lie in that particular way in which we think of them as lying when we speak of moral obligation; he is not, to that extent, seen as a morally responsible agent, as a term of moral relationships, as a member of the moral community.

I remarked also that the suspension of ordinary inter-personal attitudes and the cultivation of a purely objective view is sometimes possible even when we have no such reasons for it as I have just mentioned. Is this possible also in the case of the moral reactive attitudes? I think so; and perhaps it is easier. But the motives for a total suspension of moral reactive attitudes are fewer, and perhaps weaker: fewer, because only where there is antecedent personal involvement can

there be the motive of seeking refuge from the strains of such involvement; perhaps weaker, because the tension between objectivity of view and the moral reactive attitudes is perhaps less than the tension between objectivity of view and the personal reactive attitudes, so that we can in the case of the moral reactive attitudes more easily secure the speculative or political gains of objectivity of view by a kind of setting on one side, rather than a total suspension, of those attitudes.

These last remarks are uncertain; but also, for the present purpose, unimportant. What concerns us now is to inquire, as previously in connection with the personal reactive attitudes, what relevance any general thesis of determinism might have to their vicarious analogues. The answers once more are parallel; though I shall take them in a slightly different order. First, we must note, as before, that when the suspension of such an attitude or such attitudes occurs in a particular case, it is *never* the consequence of the belief that the piece of behaviour in question was determined in a sense such that all behaviour *might* be, and, if determinism is true, all behaviour *is*, determined in that sense. For it is not a consequence of any general thesis of determinism which might be true that nobody knows what he's doing or that everybody's behaviour is unintelligible in terms of conscious purposes or that everybody lives in a world of delusion or that nobody has a moral sense, i.e. is susceptible of self-reactive attitudes, etc. In fact no such sense of 'determined' as would be required for a general thesis of determinism is ever relevant to our actual suspensions of moral reactive attitudes. Second, suppose it granted, as I have already argued, that we cannot take seriously the thought that theoretical conviction of such a general thesis would lead to the total decay of the personal reactive attitudes. Can we then take seriously the thought that such a conviction—a conviction, after all, that many have held or said they held—would nevertheless lead to the total decay or repudiation of the vicarious analogues of these attitudes? I think that the change in our social world which would leave us exposed to the personal reactive attitudes but not at all to their vicarious analogues, the generalization of abnormal egocentricity which this would entail, is perhaps even harder for us to envisage as a real possibility than the decay of both kinds of attitude together. Though

there are some necessary and some contingent differences between the ways and cases in which these two kinds of attitudes operate or are inhibited in their operation, yet, as general human capacities or pronenesses, they stand or lapse together. Finally, to the further question whether it would not be *rational*, given a general theoretical conviction of the truth of determinism, so to change our world that in it all these attitudes were wholly suspended, I must answer, as before, that one who presses this question has wholly failed to grasp the import of the preceding answer, the nature of the human commitment that is here involved: it is *useless* to ask whether it would not be rational for us to do what it is not in our nature to (be able to) do. To this I must add, as before, that if there were, say, for a moment open to us the possibility of such a god-like choice, the rationality of making or refusing it would be determined by quite other considerations than the truth or falsity of the general theoretical doctrine in question. The latter would be simply irrelevant; and this becomes ironically clear when we remember that for those convinced that the truth of determinism nevertheless really would make the one choice rational, there has always been the insuperable difficulty of explaining in intelligible terms how its falsity would make the opposite choice rational.

I am aware that in presenting the argument as I have done, neglecting the ever-interesting varieties of case, I have presented nothing more than a schema, using sometimes a crude opposition of phrase where we have a great intricacy of phenomena. In particular the simple opposition of objective attitudes on the one hand and the various contrasted attitudes which I have opposed to them must seem as grossly crude as it is central. Let me pause to mitigate this crudity a little, and also to strengthen one of my central contentions, by mentioning some things which straddle these contrasted kinds of attitude. Thus parents and others concerned with the care and upbringing of young children cannot have to their charges either kind of attitude in a pure or unqualified form. They are dealing with creatures who are potentially and increasingly capable both of holding, and being objects of, the full range of human and moral attitudes, but are not yet truly capable of either. The treatment of such creatures must therefore represent a kind of compromise, constantly shifting in one direction, between objectivity of attitude and developed

human attitudes. Rehearsals insensibly modulate towards true performances. The punishment of a child is both like and unlike the punishment of an adult. Suppose we try to relate this progressive emergence of the child as a responsible being, as an object of non-objective attitudes, to that sense of 'determined' in which, if determinism is a possibly true thesis, all behaviour *may* be determined, and in which, if it is a true thesis, all behaviour *is* determined. What bearing *could* such a sense of 'determined' have upon the progressive modification of attitudes towards the child? Would it not be grotesque to think of the development of the child as a progressive or patchy emergence from an area in which its behaviour is in this sense determined into an area in which it isn't? Whatever sense of 'determined' is required for stating the thesis of determinism, it can scarcely be such as to allow of compromise, borderline-style answers to the question, 'Is this bit of behaviour determined or isn't it?' But in this matter of young children, it is essentially a borderline, penumbral area that we move in. Again, consider—a very different matter—the strain in the attitude of a psychoanalyst to his patient. *His* objectivity of attitude, *his* suspension of ordinary moral reactive attitudes, is profoundly modified by the fact that the aim of the enterprise is to make such suspension unnecessary or less necessary. Here we may and do naturally speak of restoring the agent's freedom. But here the restoring of freedom means bringing it about that the agent's behaviour shall be intelligible in terms of conscious purposes rather than in terms only of unconscious purposes. *This* is the object of the enterprise; and it is in so far as *this* object is attained that the suspension, or half-suspension, of ordinary moral attitudes is deemed no longer necessary or appropriate. And in this we see once again the *irrelevance* of that concept of 'being determined' which must be the central concept of determinism. For we cannot both agree that this object is attainable and that its attainment has this consequence and yet hold (1) that neurotic behaviour is determined in a sense in which, it may be, all behaviour is determined, and (2) that it is because neurotic behaviour is determined in this sense that objective attitudes are deemed appropriate to neurotic behaviour. Not, at least, without accusing ourselves of incoherence in our attitude to psychoanalytic treatment.

VI

And now we can try to fill in the lacuna which the pessimist finds in the optimist's account of the concept of moral responsibility, and of the bases of moral condemnation and punishment; and to fill it in from the facts as we know them. For, as I have already remarked, when the pessimist himself seeks to fill it in, he rushes beyond the facts as we know them and proclaims that it cannot be filled in at all unless determinism is false.

Yet a partial sense of the facts as we know them is certainly present to the pessimist's mind. When his opponent, the optimist, undertakes to show that the truth of determinism would not shake the foundations of the concept of moral responsibility and of the practices of moral condemnation and punishment, he typically refers, in a more or less elaborated way, to the efficacy of these practices in regulating behaviour in socially desirable ways. These practices are represented solely as instruments of policy, as methods of individual treatment and social control. The pessimist recoils from this picture; and in his recoil there is, typically, an element of emotional shock. He is apt to say, among much else, that the humanity of the offender himself is offended by *this* picture of his condemnation and punishment.

The reasons for this recoil—the explanation of the sense of an emotional, as well as a conceptual, shock—we have already before us. The picture painted by the optimists is painted in a style appropriate to a situation envisaged as wholly dominated by objectivity of attitude. The only operative notions invoked in this picture are such as those of policy, treatment, control. But a thoroughgoing objectivity of attitude, excluding as it does the moral reactive attitudes, excludes at the same time essential elements in the concepts of *moral* condemnation and *moral* responsibility. This is the reason for the conceptual shock. The deeper emotional shock is a reaction, not simply to an inadequate conceptual analysis, but to the suggestion of a change in our world. I have remarked that it is possible to cultivate an exclusive objectivity of attitude in some cases, and for some reasons, where the object of the attitude is not set aside from developed interpersonal and moral attitudes by immaturity or abnormality. And the

suggestion which seems to be contained in the optimist's account is that such an attitude should be universally adopted to all offenders. This is shocking enough in the pessimist's eyes. But, sharpened by shock, his eyes see further. It would be hard to make *this* division in our natures. If to all offenders, then to all mankind. Moreover, to whom could this recommendation be, in any real sense, addressed? Only to the powerful, the authorities. So abysses seem to open.[5]

But we will confine our attention to the case of the offenders. The concepts we are concerned with are those of responsibility and guilt, qualified as 'moral', on the one hand—together with that of membership of a moral community; of demand, indignation, disapprobation and condemnation, qualified as 'moral', on the other hand—together with that of punishment. Indignation, disapprobation, like resentment, tend to inhibit or at least to limit our goodwill towards the object of these attitudes, tend to promote an at least partial and temporary withdrawal of goodwill; they do so in proportion as they are strong; and their strength is in general proportioned to what is felt to be the magnitude of the injury and to the degree to which the agent's will is identified with, or indifferent to, it. (These, of course, are not contingent connections.) But these attitudes of disapprobation and indignation 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, as the holding of objective attitudes does, involve as a part of itself 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, rather, 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 essential part of punishment is all of a piece with this whole range of attitudes of which I have been speaking. It is not only moral reactive

5. Peered into by Mr. J. D. Mabbott, in his article 'Freewill and Punishment', published in *Contemporary British Philosophy*, 3rd ser., 1956.

attitudes towards the offender which are in question here. We must mention also the self-reactive attitudes of offenders themselves. Just as the other-reactive attitudes are associated with a readiness to acquiesce in the infliction of suffering on an offender, within the 'institution' of punishment, so the self-reactive attitudes are associated with a readiness on the part of the offender to acquiesce in such infliction *without* developing the reactions (e.g. of resentment) which he would normally develop to the infliction of injury upon him; i.e. with a readiness, as we say, to accept punishment[6] as 'his due' or as 'just'.

I am not in the least suggesting that these readinesses to acquiesce, either on the part of the offender himself or on the part of others, are always or commonly accompanied or preceded by indignant boilings or remorseful pangs; only that we have here a continuum of attitudes and feelings to which these readinesses to acquiesce themselves belong. Nor am I in the least suggesting that it belongs to this continuum of attitudes that we should be ready to acquiesce in the infliction of injury on offenders in a fashion which we saw to be quite indiscriminate or in accordance with procedures which we knew to be wholly useless. On the contrary, savage or civilized, we have some belief in the utility of practices of condemnation and punishment. But the social utility of these practices, on which the optimist lays such exclusive stress, is not what is now in question. What is in question is the pessimist's justified sense that to speak in terms of social utility alone is to leave out something vital in our conception of these practices. The vital thing can be restored by attending to that complicated web of attitudes and feelings which form an essential part of the moral life as we know it, and which are quite opposed to objectivity of attitude. Only by attending to this range of attitudes can we recover from the facts as we know them a sense of what we mean, i.e. of *all* we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice. But we *do* recover it from the facts as we know them. We do not have to go beyond them. Because the optimist neglects or misconstrues these attitudes, the pessimist rightly claims to find a lacuna in his account. We can fill the

6. Of course not any punishment for anything deemed an offence.

lacuna for him. But in return we must demand of the pessimist a surrender of his metaphysics.

Optimist and pessimist misconstrue the facts in very different styles. But in a profound sense there is something in common to their misunderstandings. Both seek, in different ways, to over-intellectualize the facts. Inside the general structure or web of human attitudes and feelings of which I have been speaking, there is endless room for modification, redirection, criticism, and justification. But questions of justification are internal to the structure or relate to modifications internal to it. The existence of the general framework of attitudes itself is something we are given with the fact of human society. As a whole, it neither calls for, nor permits, an external 'rational' justification. Pessimist and optimist alike show themselves, in different ways, unable to accept this.[7] The optimist's style of over-intellectualizing the facts is that of a characteristically incomplete empiricism, a one-eyed utilitarianism. He seeks to find an adequate basis for certain social practices in calculated consequences, and loses sight (perhaps wishes to lose sight) of the human attitudes of which these practices are, in part, the expression. The pessimist does not lose sight of these attitudes, but is unable to accept the fact that it is just these attitudes themselves which fill the gap in the optimist's account. Because of this, he thinks the gap can be filled only if some general metaphysical proposition is repeatedly verified, verified in all cases where it is appropriate to attribute moral responsibility. This proposition he finds it as difficult to state coherently and with intelligible relevance as its determinist contradictory. Even when a formula has been found ('contra-causal freedom' or something of the kind) there still seems to remain a gap between its applicability in particular cases and its supposed moral consequences. Sometimes he plugs this gap with an intuition of fittingness—a pitiful intellectualist trinket for a

7. Compare the question of the justification of induction. The human commitment to inductive belief-formation is original, natural, non-rational (not *ir*rational), in no way something we choose or could give up. Yet rational criticism and reflection can refine standards and their application, supply 'rules for judging of cause and effect'. Ever since the facts were made clear by Hume, people have been resisting acceptance of them.

philosopher to wear as a charm against the recognition of his own humanity.

Even the moral sceptic is not immune from his own form of the wish to over-intellectualize such notions as those of moral responsibility, guilt, and blame. He sees that the optimist's account is inadequate and the pessimist's libertarian alternative inane; and finds no resource except to declare that the notions in question are inherently confused, that 'blame is metaphysical'. But the metaphysics was in the eye of the metaphysician. It is a pity that talk of the moral sentiments has fallen out of favour. The phrase would be quite a good name for that network of human attitudes in acknowledging the character and place of which we find, I suggest, the only possibility of reconciling these disputants to each other and the facts.

There are, at present, factors which add, in a slightly paradoxical way, to the difficulty of making this acknowledgement. These human attitudes themselves, in their development and in the variety of their manifestations, have to an increasing extent become objects of study in the social and psychological sciences; and this growth of human self-consciousness, which we might expect to reduce the difficulty of acceptance, in fact increases it in several ways. One factor of comparatively minor importance is an increased historical and anthropological awareness of the great variety of forms which these human attitudes may take at different times and in different cultures. This makes one rightly chary of claiming as essential features of the concept of morality in general, forms of these attitudes which may have a local and temporary prominence. No doubt to some extent my own descriptions of human attitudes have reflected local and temporary features of our own culture. But an awareness of variety of forms should not prevent us from acknowledging also that in the absence of *any* forms of these attitudes it is doubtful whether we should have anything that *we* could find intelligible as a system of human relationships, as human society. A quite different factor of greater importance is that psychological studies have made us rightly mistrustful of many particular manifestations of the attitudes I have spoken of. They are a prime realm of self-deception, of the ambiguous and the shady, of guilt-transference, unconscious sadism and the rest. But it is an exaggerated horror, itself suspect, which would make us unable to

acknowledge the facts because of the seamy side of the facts. Finally, perhaps the most important factor of all is the prestige of these theoretical studies themselves. That prestige is great, and is apt to make us forget that in philosophy, though it also is a theoretical study, we have to take account of the facts in *all* their bearings; we are not to suppose that we are required, or permitted, as philosophers, to regard ourselves, as human beings, as detached from the attitudes which, as scientists, we study with detachment. This is in no way to deny the possibility and desirability of redirection and modification of our human attitudes in the light of these studies. But we may reasonably think it unlikely that our progressively greater understanding of certain aspects of ourselves will lead to the total disappearance of those aspects. Perhaps it is not inconceivable that it should; and perhaps, then, the dreams of some philosophers will be realized.

If we sufficiently, that is *radically*, modify the view of the optimist, his view is the right one. It is far from wrong to emphasize the efficacy of all those practices which express or manifest our moral attitudes, in regulating behaviour in ways considered desirable; or to add that when certain of our beliefs about the efficacy of some of these practices turn out to be false, then we may have good reason for dropping or modifying those practices. What *is* wrong is to forget that these practices, and their reception, the reactions to them, really *are* expressions of our moral attitudes and not merely devices we calculatingly employ for regulative purposes. Our practices do not merely exploit our natures, they express them. Indeed the very understanding of the kind of efficacy these expressions of our attitudes have turns on our remembering this. When we do remember this, and modify the optimist's position accordingly, we simultaneously correct its conceptual deficiencies and ward off the dangers it seems to entail, without recourse to the obscure and panicky metaphysics of libertarianism.

ACKNOWLEDGMENTS

THIS BOOK has been helped by many people over many years. Special thanks are due to Lucy Allais, David Beglin, Maudemarie Clarke, Justin D'Arms, Daniela Dover, Stephen Finlay, Laura Gillespie, David Goldman, Barbara Herman, John F. Horty, Andrew Hsu, Daniel Jacobson, Mark C. Johnson, A. J. Julius, Gavin Lawrence, Janice McDowell, Victoria McGreer, Michael McKenna, Benjamin Mitchell-Yellin, Herbert Morris, Calvin Normore, Phillip Pettit, Jacob Ross, T. M. Scanlon, Kieran Setiya, Seana Valentine Shiffrin, David Shoemaker, Angela M. Smith, Scott Soames, David Sobel, Galen Strawson, Sharon Street, Gary Watson, Roger White, Gideon Yaffe, and audiences at the 50th Anniversary of P. F. Strawson's "Freedom and Resentment" College of William and Mary; Ethicists Network, Princeton University; New Orleans Invitational Seminar in Ethics, Tulane University; Syracuse University; MIT; Colgate University; the University of Southern California; and Harvard University.

The final manuscript was greatly improved by input and extensive commentary from Sarah Buss, Matt Rohal and the editorial team at Princeton University Press, and an anonymous reviewer.

Work on this book was generously supported by a grant from the American Council of Learned Societies and the Center for Advanced Study in the Behavioral Sciences at Stanford University.

BIBLIOGRAPHY

Bilgrami, Akeel. *Self-Knowledge and Resentment*. Cambridge, MA: Harvard University Press, 2006.

Callanan, John J. "Making Sense of Doubt: Strawson's Anti-Scepticism." *Theoria* 77, no. 3 (September 2011): 261–78.

Carnap, Rudolf. "Empiricism, Semantics, and Ontology." *Revue Internationale de Philosophie* 4 (1950): 20–40.

Darwall, Stephen. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press, 2006.

Goldman, David. "Modification of the Reactive Attitudes." *Pacific Philosophical Quarterly* 94, no. 4 (2014): 1–22.

Goodman, Nelson. *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press, 1955.

Grice, H. P., and Peter F. Strawson. "In Defense of a Dogma." *Philosophical Review* 65, no. 2 (April 1956): 141–58.

Hieronymi, Pamela. "Articulating an Uncompromising Forgiveness." *Philosophy and Phenomenological Research* 62, no. 3 (2001): 529–55.

———. "Believing at Will." *Canadian Journal of Philosophy* 35, suppl. (2009): 149–87.

———. "Controlling Attitudes." *Pacific Philosophical Quarterly* 87, no. 1 (March 2006): 45–74.

———. "Fairness, Sanction, and Condemnation." Promised to *Oxford Studies in Agency and Responsibility*, edited by David Shoemaker. Oxford: Oxford University Press, 2020.

———. "The Force and Fairness of Blame." *Philosophical Perspectives* 18, no. 1 (2004): 115–48.

———. "I'll Bet You Think This Blame Is about You." In *Oxford Studies in Agency and Responsibility: Essays on Themes from the Work of Gary Watson*, edited by Justin Coates and Neal Tognazzini, 60–87. Oxford: Oxford University Press, 2019.

———. "Of Metaethics and Motivation: The Appeal of Contractualism." In *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon*, edited by Rahul Kumar,

Samuel Scheffler, and R. Jay Wallace, 101–28. Oxford: Oxford University Press, 2011.

———. "Reflection and Responsibility." *Philosophy and Public Affairs* 42, no. 1 (2014): 3–41.

———. "The Wrong Kind of Reason." *Journal of Philosophy* 102, no. 9 (September 2005): 1–21.

Holmes, Oliver Wendell, Jr. "Early Forms of Liability." Lecture I from *The Common Law* (1909).

Kant, Immanuel. *Critique of Practical Reason*. Cambridge Texts in the History of Philosophy. Translated by Mary Gregor. Cambridge: Cambridge University Press, 1997.

Nietzsche, Friedrich. *On the Genealogy of Morality*. Translated by Maudemarie Clarke and Alan J. Swenson. Indianapolis: Hackett, 1998.

Nowell-Smith, P. H. "Freewill and Moral Responsibility." *Mind* 57, no. 225 (1948): 45–61.

Orwell, George. *Animal Farm*. London: Secker and Warburg, 1945.

Plato. *Republic*. Translated by G. M. A. Grube. Edited by C. D. C. Reeve. Indianapolis: Hackett, 1992.

Rawls, John. *A Theory of Justice*. Cambridge, MA: Belknap, 1971.

Sartre, Jean-Paul. "Existentialism Is a Humanism." Lecture, 1945.

Scanlon, T. M. *Being Realistic about Reasons*. Oxford: Oxford University Press, 2014.

———. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press, 1998.

Shoemaker, David, and Neal Tognazzini, eds. *Freedom and Resentment at 50*. Oxford Studies in Agency and Responsibility, vol. 2. New York: Oxford University Press, 2015.

Smart, J. J. C. "Free Will, Praise, and Blame." *Mind* 70, no. 279 (1961): 291–306.

Strawson, Peter F. "Carnap's Views on Constructed Systems versus Natural Languages in Analytic Philosophy." In *The Philosophy of Rudolf Carnap*, Library of Living Philosophers, edited by Paul A. Schilpp, 503–18. LaSalle, IL: Open Court, 1963.

———. "Freedom and Resentment." *Proceedings of the British Academy* 48 (1962): 1–25.

———. *Individuals: An Essay in Descriptive Metaphysics*. London: Routledge, 1959.

———. *Introduction to Logical Theory*. London: Methuen, 1952.

———. "Replies." In *Philosophical Subjects: Essays Presented to P. F. Strawson*, edited by Zak Van Straaten, 260–96. Oxford: Clarendon, 1980.

———. "Reply to Ernest Sosa." In *The Philosophy of P. F. Strawson*, Library of Living Philosophers, ed. Lewis Edwin Hahn, 370–72. LaSalle, IL: Court Press, 1998.

———. *Skepticism and Naturalism*. New York: Columbia University Press, 1985.

———. "Social Morality and Individual Ideal." *Philosophy* 36, no. 136 (January 1961): 1–17.

Street, Sharon. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127, no. 1 (2006): 109–66.

Stroud, Barry. "Transcendental Arguments." *Journal of Philosophy* 65, no. 9 (1968): 241–56.

Taylor, Charles. "Responsibility for Self." In *The Identities of Persons*, edited by Amélia O. Rorty, 281–99. Berkeley: University of California Press, 1976.

Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard University Press, 1996.

Watson, Gary, ed. *Free Will*. 2nd ed. Oxford: Oxford University Press, 2003.

INDEX

A NOTE ON THE TYPE

This book has been composed in Arno, an Old-style serif typeface in the classic Venetian tradition, designed by Robert Slimbach at Adobe.