# The Sources of Normativity

*CHRISTINE M. KORSGAARD*

THE TANNER LECTURES ON HUMAN VALUES

Delivered at

Clare Hall, Cambridge University
November 16 and 17, 1992

CHRISTINE M. KORSGAARD is currently Professor of Philosophy at Harvard University. She was educated at the University of Illinois at Urbana-Champaign, and at Harvard, where she received her Ph.D. degree in philosophy in 1979. She has taught at several schools in the University of California system, including UC Santa Barbara, UCLA, and UC Berkeley, and at the University of Chicago. She is a member of the American Philosophical Association, the North American Kant Society, the Hume Society, and the American Society for Political and Legal Philosophy. She has published and lectured extensively on Immanuel Kant, including "Kant," in *Ethics in the History of Western Philosophy,* edited by Cavalier, Gouinlock, and Sterba (1989), "Kant's Analysis of Obligation," in *The Monist* (1989), and "Aristotle and Kant on the Source of Value," in *Ethics* (1986). In addition, her articles "Immanuel Kant," "John Rawls," and "Richard Price," were published in *The Garland Encyclopedia of Ethics,* edited by Becker (1992). A longer version of her Tanner Lecture, *The Sources of Normativity,* with commentary by G. A. Cohen, Raymond Geuss, Thomas Nagel, and Bernard Williams, and edited by Onora O'Neil, is forthcoming from Cambridge University Press.

LECTURE I: THE NORMATIVE QUESTION

*Introduction*

In 1625, in his book *On the Law of War and Peace,* Hugo Grotius asserted that human beings would have obligations "even if we should concede that which cannot be conceded without the utmost wickedness, that there is no God, or that the affairs of men are of no concern to Him."[1]  But two of his followers, Thomas Hobbes and Samuel Pufendorf, thought that Grotius was wrong. However socially useful moral conduct might be, they argued, it is not really *obligatory* unless some sovereign authority, backed by the power of sanctions, lays it down as the law.[2]  Others in turn disagreed with them, and so the argument began.

Ever since then, modern moral philosophers have been engaged in a debate about the "foundations" of morality. We need to be shown, it is often urged, that morality is "objective." The early rationalists, Samuel Clarke and Richard Price, thought that they knew exactly what they meant by this.[3] Hobbes had said that there is no right or wrong in the state of nature, and to them, this implied that rightness is mere invention or convention, not something real.[4] Hobbes meant that individuals are not obligated to obey the laws of social cooperation in the absence of a sovereign who can impose them on everyone.[5] But the rationalists took him

---

[1] Grotius, *On the Law* of *War and Peace,* Schneewind I, p. 92. I owe a great debt to Jerome Schneewind for drawing my attention to this stretch of the historical debate, and especially for encouraging me to read Pufendorf.

[2.] See Hobbes, especially *Leviathan;* and Pufendorf, *On the Law of Nature and of Nations* and *On the Duty* of *Man and Citizen according to Natural Law.*

[3] See Clarke*, A Discourse concerning the Unchangeable Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation: The Boyle Lectures 1705;* and Price, *A Review* of *the Principal Questions in Morals.*

[4] Hobbes, *Leviathan,* 1.13, p. 90.

[5] Ibid., 1.15, p. 110.

to mean what Bernard Mandeville had later ironically asserted: that virtue is just an invention of politicians, used to keep their human cattle in line.[6]

But what exactly is the problem with that? Showing that something is an invention is not a way of showing that it is not real. Moral standards exist, one might reply, in the only way standards of conduct *can* exist: people believe in such standards and therefore regulate their conduct in accordance with them. Nor are these facts difficult to explain. We all know in a general way how and why we were taught to follow moral rules and that it would be impossible for us to get on together if we didn't do something along these lines. We are social animals, and probably the whole thing has a biological basis. So what's missing here, that makes us seek a philosophical "foundation"?

The answer lies in the fact that ethical standards are *normative*. They do not merely describe a way in which we in fact regulate our conduct. They make *claims* on us: they command, oblige, recommend, or guide. Or at least, when we invoke them, we make claims on one another. When I say that an action is right I am saying that you ought to *do* it; when I say that something is good I am recommending it as worthy of your choice. The same is true of the other concepts for which we seek philosophical foundations. Concepts like knowledge, beauty, and meaning, as well as virtue and justice, all have a normative dimension, for they tell us what to think, what to like, what to say, what to do, and what to be. And it is the force of these normative claims — the right of these concepts to give laws to us — that we want to understand.

And in ethics, the question can become urgent, for the day will come, for most of us, when what morality commands, obliges, or recommends is *hard:* that we share decisions with people whose

---

[6] See Mandeville, *The Fable of the Bees: or, Private Vices, Public Benefits,* especially the section "An Enquiry into the Origin of Moral Virtue," pp. 41–57. Mandeville himself denied that he meant either that virtue is unreal or that it is not worth having. See for instance "A Vindication of the Book," pp. 384ff.; and also *An Enquiry into the Origin of Honor*, Schneewind II, pp. 396–98.

intelligence and integrity don't inspire our confidence; that we assume grave responsibilities to which we feel inadequate; that we sacrifice our lives or voluntarily relinquish what makes them sweet. And then the question *why?* will press, and rightly so. Why should I be moral? This is not, as H. A. Prichard supposed, a misguided request for a demonstration that morality is in our interest (although that may be one answer to the question).[7] It is a call for philosophy, the examination of life. Even those who are convinced that "it is right" must be in itself a sufficient reason for action may request an account of rightness that this conviction will survive. The trouble with a view like Mandeville's is not that it is not a reasonable explanation of how moral practices came about, but rather that our commitment to these practices would not survive our belief that it was true.[8] Why give up your heart's desire, just because some politician wants to keep you in line? When we seek a philosophical foundation for morality we are not looking merely for an explanation of moral practices. We are asking what *justifies* the claims that morality makes on us. This is what I will call "the normative question."

Now it is often thought that the normative question poses a special problem for *modern* moral philosophers. The Modern Scientific World View is supposed to be somehow inimical to ethics, while, in different ways, the teleological metaphysics of the the ancient Greek world and the religious systems of medieval Europe seemed friendlier to the subject. It is a little hard to put the point clearly and in a way that does not give rise to obvious objections, but both of these earlier outlooks seem to support the idea that human life has a purpose that is or only can be fulfilled

---

[7] Prichard, "Does Moral Philosophy Rest on a Mistake?" and "Duty and Interest." Prichard's argument is discussed in detail below.

[8] Actually, as Hume and Hutcheson both argued, there are also problems about the explanatory adequacy of Mandeville's view. For Hume's discussion, see the *Enquiry concerning the Principles of Morals* (1751), p. 214. For Hutcheson's, see the *Inquiry concerning the Original of Our Ideas of Beauty and Virtue* (1725), Raphael I, p. 291. Neither Hume nor Hutcheson names Mandeville, but he is clearly their target.

by those who live up to ethical standards and meet moral demands. And this is supposed to be sufficient to establish that ethics is really normative, that its demands on us are justified. They are justified in the name of life's purpose. The Modern Scientific World View, in depriving us of the idea that the world has a purpose, has taken this justification away.

Whether this is true or not, the moral philosophy of the modern period can be read as a search for the source of normativity. Philosophers in the modern period have come up with four successive answers to the question of what makes morality normative. In brief, they are these:

(1) Voluntarism. According to this view, moral obligation derives from the command of someone who has legitimate authority over the moral agent and so can make laws for her. You must do the right thing because God commands it, say, or because a political sovereign whom you have agreed to obey makes it law. Normativity springs from a legislative will. This is the view of Pufendorf and of Hobbes.

(2) Realism. According to this view, moral claims are normative if they are true, and true if there are intrinsically normative entities or facts that they correctly describe. Realists try to establish the normativity of ethics by arguing that values or obligations or reasons really exist or, more commonly, by arguing against the various forms of skepticism about them. This kind of argument has been found in the work of rational intuitionists ever since the eighteenth century. It was advanced vigorously by Clarke and Price in the eighteenth century and by Prichard, G. E. Moore, and W. D. Ross in the early twentieth century.[9] It is also found in the

---

[9] Clarke, *A Discourse concerning the Unchangeable Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation: The Boyle Lectures 1705*; Price, *A Review of the Principal Questions in Morals;* Prichard, *Moral Obligation and Duty and Interest: Essays and Lectures by H. A. Prichard;* Moore, *Principia Ethica;* and Ross, *The Right and the Good.*

work of some contemporary moral realists, including Thomas Nagel.[10]

(3) I call the third view "Reflective Endorsement." This view is favored by philosophers who believe that morality is grounded in human nature. The philosopher's first job is to explain what the source of morality in human nature is, why we use moral concepts and feel ourselves bound by them. When an explanation of our moral nature is in hand, we can then raise the normative question: all things considered, do we have reason to accept the claims of our moral nature or should we reject them? The question is not "are these claims true?" as it is for the realist. The reasons sought here are practical reasons; the idea is to show that morality is good for us. Arguments with this structure can be found in the tradition in the work of Francis Hutcheson, David Hume, and John Stuart Mill, and in contemporary philosophy in the work of Bernard Williams.“

(4) The Appeal to Autonomy. This kind of argument is found in Immanuel Kant and contemporary Kantian constructivists, especially John Rawls.[12]  Kantians believe that the source of the normativity of moral claims must be found in the agent's own will, in particular in the fact that the laws of morality are the laws of the agent's own will and that its claims are ones she is prepared to make on herself. The capacity for self-conscious reflection about

---

[10] In *The Possibility of Altruism* and *The View from Nowhere.* But see note 44 below.

[11] See Hutcheson, *Inquiry concerning the Original of our Ideas of Beauty and Virtue* and *Illustrations on the Moral Sense;* Hume, *A Treatise of Human Nature* and *Enquiry concerning the Principles of Morals;* Mill, *Utilitarianism;* and Williams, *Ethics and the Limits* of *Philosophy.* More specific references for Hutcheson, Hume, and Williams will be found in Lecture 2. Mill's argument appears in chapter 3, "Of the Ultimate Sanction of the Principle of Utility."

[12] See Kant, *Foundations of the Metaphysics of Morals* and *Critique* of *Practical Reason;* Rawls, *A Theory* of *Justice* and "Kantian Constructivism in Moral Theory: The Dewey Lectures 1980."

our own actions confers on us a kind of authority over ourselves, and it is this authority that gives normativity to moral claims.

During the modern period, each of these accounts of normativity developed in response to the prior one, sometimes as a result of criticism, more often when the implications of the earlier view were pressed a little harder. In this lecture and the next one I am going to describe this historical process, comparing earlier versions of these accounts with those on the contemporary scene. The Kantian account was the culmination of this historical development. In the third lecture I will present an updated version of that account that I believe to be true.

In the rest of this lecture I will discuss the first two theories of normativity: voluntarism and moral realism.

## *Voluntarism*

As I mentioned at the beginning of this lecture, Grotius asserted that human beings would have obligations even if God did not exist to give us laws. Because of that remark, he is often identified as the first *modern* moral philosopher.[13] But the credit for that should really go to Hobbes and Pufendorf. For they were the first to identify clearly the special challenge that the Modern Scientific World View presents to ethics and to try to construct ethical theories in the face of that challenge.

According to Pufendorf, the actions of human beings, like every other form of physical motion, are in themselves morally indifferent. Values are not found in the world of nature at all. Instead, Pufendorf says, intelligent beings must *impose* moral values on nature. He tells us that what he calls "moral entities" — values and obligations — are "superadded" to physical entities — such as actions — at "the will of intelligent entities."[14] Hobbes opens his most famous ethical treatise with the apparently unpromising reflection that since to be alive is simply to be a self-

[13] I owe this point to Schneewind. See Schneewind I, pp. 88–89.
[14] Pufendorf, *The Law of Nature and of Nations,* Schneewind I, p. 171.

moving object, we may as well say that watches and engines and other self-moving objects have an artificial life, and that we ourselves in turn are just a kind of machine.[15]  And he proceeds to construct a completely mechanistic explanation of how human beings work and an ethics that is based upon it.

Their question is how nature, an indifferent and mechanical world of matter in motion, can come to be imbued with moral properties.  Interestingly, both Pufendorf and Hobbes traced obligation ultimately to divine command, not because they hung on to a medieval or religious conception of the world, but rather because they had *adopted* the Modern Scientific World View.  They believed that it *takes* God or a Godlike sovereign to impose moral properties on the indifferent world of nature.  Pufendorf held that "since . . . moral necessity . . . and turpitude . . . are affections of human actions arising from their conformity or non-conformity to some norm or law, and law is the bidding of a superior, it does not appear that [they] . . . can be conceived to exist before law, and without the imposition of a superior."[16]  And Hobbes of course maintained that there is no obligation until a sovereign capable of enforcing the "laws of nature" is in power.  Obligation must come from law, and law from the will of a legislating sovereign; morality only comes into the world when laws are made.

Pufendorf and Hobbes shared two other views of which their critics sometimes failed to see the importance.  First, voluntarism is often criticized on the ground that the sovereign can make anything right or wrong.  And many theological voluntarists have held that that is true.  But Pufendorf and Hobbes thought that the *content* of morality is given by reason independently of the legislative will. They agreed that good and evil, prudence and imprudence, and in a way even justice and injustice, are objectively identifiable attributes of states of affairs and of the actions that produce them. What is good is what is naturally beneficial to a person; what is

[15] Hobbes, *Leviathan,* introduction, p. 9.
[16] Pufendorf, *The Law of Nature and of Nations,* Schneewind I, p. 175.

right and just is what makes harmonious social life possible. So no legislator is needed to give content, at least in a general way, to the ideas of the good and the right. Most human beings in most circumstances have reason to want what is good and, at least as a group, to do what *is* right, independently of law or obligation. But in the absence of God, Pufendorf wrote, the precepts of morality might "be observed for their utility, like the prescriptions doctors give to regulate health" but ". . . would not be *laws.*"[17] And Hobbes, after laying out his laws of nature, says: "These dictates of Reason, men use to call by the name of Lawes; but improperly: for they are but Conclusions, or Theorems concerning what conduceth to the conservation and defence of themselves; whereas Law, properly is the word of him that by right hath command over others." [18] So the role of the legislator is to make what is *in any case* a good idea into *law.*

Second, both Pufendorf and Hobbes believed that no one could be a legislator without the power to impose sanctions to enforce his law. And it is frequently inferred that the point of these sanctions is to provide the subjects of the law with motives to obey it. Actually, however, both of these philosophers thought that morally good action is action that proceeds from what we would now call the motive of duty." Morally good actions are done from what Pufendorf calls an "intrinsic motive" rather than from interest or fear." Pufendorf says that this marks the difference between obligation and compulsion; and Hobbes, similarly, that it marks the difference between mere counsel and command." A just man, as

---

[17] Pufendorf, *On the Duty* of *Man and Citizen,* p. 36.

[18] Hobbes, *Leviathan,* 1.15, p. 111.

[19] While Pufendorf is almost ignored by contemporary moral philosophers, there is a great deal of controversy about Hobbes's views on moral motivation and obligation and substantial recent literature on the topic. For references, see Tuck's Introduction to *Leviathan,* p. xliii. While a complete defense of the view I set forward here would require taking on the issues raised by that controversy, this is not the place for that.

[20] Pufendorf, *On the Law* of  *Nature and of Nations,* Schneewind I, p. 180.

[21] Ibid.; Hobbes, *Leviathan,* II.25, pp. 176-79.

Hobbes put it, is one whose will is "framed" by justice, not by fear or benefit to himself.[22] One does the right thing because it is the right thing, because it is the law, and for no other reason.

Why, then, are sanctions needed? The answer is that they are necessary to establish the authority of the legislator. Pufendorf and Hobbes thought that the legislator's power to enforce the law is necessary to give moral commands the special force of *requirement*. A homely example will illustrate their point. Suppose you are a student in my department. Then my colleagues and I are in a position to require you to take a course in logic. We are in this position because we have authority over you, and we have authority over you in part because we can impose a sanction on you. If you refuse to take the logic course, you will not get a degree from us. Now I want you to notice several things about this. First of all, the scenario does not in the least imply that our decision to make you study logic is arbitrary. It may be a very good idea for philosophy students to study logic, and that may be why we require it. If we are good at our jobs and worthy of our authority, we will have some such reason. In a similar way the laws that God or the Hobbesian sovereign requires us to obey are precepts of reason, determined independently of any arbitrary legislative will. Yet it is not merely their reasonableness that obligates us to obey them, just as it is not merely the benefit of studying logic that obligates students in my department to take the logic course. For if you are a philosophy student but are not in my department, I can give you all sorts of excellent reasons why you should take a course in logic, and you will not thereby be required to take one. And that is why authority requires a sanction.

Let me play out the analogy a moment longer. Suppose again that you are a student in my department and consider your motive for taking the logic course. There are three possibilities. First, you might take it because you grasp the reasons why we require it. You see that it is a good idea and you are moved by that fact.

[22] Hobbes, *Leviathan,* I.15, p. 104.

Second, even if you think the requirement arbitrary and unnecessary, you may take the course out of fear of being denied your degree — because of the sanction. Or, third, you may take it simply because it is a required course. The important point is that the third motive is appropriate here. While you may very well grasp the reasons why we require the course, and it may even be true that for those reasons you would have taken it anyway, there is something a little odd about saying that this is your motive. Since it is required you would have to take it in any case. But there is no reason to suppose that therefore you only take it out of fear of being denied your degree, as it were cringingly. It's being a required course is, under the circumstances, itself a reason. This is the picture of obligation, and of what it is to act from the moral motive, that Hobbes and Pufendorf have in mind. And according to this picture neither moral obligation nor its proper and characteristic motive, the motive of duty, are possible unless there is a legislator backed by the power of sanctions who can lay down the law.

Let me sum up. Hobbes and Pufendorf believed that the content of morality is given by natural reason. What morality demands of us is what it is reasonable for us, at least as a group, to do. The rules of morality are the rules that make social life possible, and social life is necessary for human beings. Hobbes and Pufendorf clearly supposed that in many cases this consideration could be motivationally sufficient as well. Pufendorf, especially, says that in the absence of obligation we would still do what is right because it is useful. The legislator is not invoked to supply the content of morality or to explain why people are often motivated to do what is right. The legislator is necessary to make *obligation* possible, that is, to make morality normative.

### Realism

Samuel Clarke, the first defender of realism, was quick to spot what he took to be a fatal flaw in the view I have just described.

Hobbes, Clarke complains, tries to derive obligation from the so-
cial contract, from our agreement to obey the laws of a sovereign
who will make social cooperation possible. But why are we obli-
gated to conform to the social contract? Clarke says: "To make
these *compacts* obligatory [Hobbes] is forced . . . to recur to an
antecedent *law of nature:* and this destroys all that he had before
said. For the same law of nature which obliges men to *fidelity,
after* having made a compact; will unavoidably, upon all the same
accounts, be found to oblige them, *before* all compacts, to *content-
ment* and mutual *benevolence* . . .[23] If the need to establish a
cooperative system can obligate us to conform to a social contract,
why doesn't that same need obligate us to behave ourselves in
cooperative ways in the first place? Or, if we say obligation comes
from the fact that the laws have been made by the sovereign, then
what are we to say about why we are obligated to obey the sov-
ereign? Again Clarke complains that "compacts ought to be faith-
fully performed, and obedience to be duly paid to civil powers:
the obligation *these things* [Hobbes] is forced to deduce entirely
from the internal reason and fitness of the things themselves. . ."[24]

Pufendorf tries to *explain* why we are obligated to obey the
sovereign, by defining a notion of legitimate authority. He stipu-
lates that the superior who is able to obligate us must have these
two attributes: "not only the strength to inflict some injury upon
the recalcitrant but also just cause to require us to curtail the lib-
erty of our will at his discretion."[25] He goes on to explain that
another has the right to claim our obedience if he has conferred
exceptional benefits on us; or if he is able to look out for us much
better than we can look out for ourselves; or of course if we have
contracted to obey him. So the authority of the legislator springs
not only from his power to impose sanctions, but also from our

[23] Clarke, *A Discourse concerning the Unchangeable Obligations of Natural
Religion,* Raphael I, p. 219.

[24] Ibid., p. 221.

[25] Pufendorf, *On the Duty of Man and Citizen,* p. 28.

gratitude for his benefits or from his benevolent wisdom or from
our own contractual acts. But the difficulty with this solution is
obvious. If we have no antecedent obligation to be grateful to
benefactors, or to submit to the guidance of benevolent wisdom,
or to honor our agreements, how can these things confer legitimate
authority on the legislator? And if we do have a natural obliga-
tion to these things, then why may we not have other natural obli-
gations as well? The very notion of a legitimate authority is al-
ready a normative one and cannot be used to answer the norma-
tive question.

The problem here is a general one, which applies to any
Hobbes has a way of avoiding this last problem, but it is at a
serious cost. He says flatly that God's authority does not depend
on our gratitude or on His graciousness, but simply on His irre-
sistible power.[26] And he concludes that this is true of the authority
of the political sovereign as well. But this gives rise to a problem.
The sovereign's authority now consists entirely in his ability to
punish us. Although sanctions are not our motive for obedience,
they are the source of the sovereign's authority and so of our obli-
gations. I am obligated to do what is right only because the sov-
ereign can punish me if I do not. Well, suppose I commit a crime
and I get away with it. Then the sovereign was not able to punish
me. And if my obligation sprang from his ability to punish me,
then I had no obligation. So a crime I get away with is no crime
at all. If irresistible power is just power unsuccessfully resisted,
then authority is nothing more than the successful exercise of
power, and things always turn out right. For no one can ever do
what he lacks the power to do.[27]

The problem here is a general one, which applies to any
attempt to derive normativity from a natural source of power.
Suppose the authority of obligation derives from the power of our

---

[26] Hobbes, *Leviathan,* 11.31, p. 246.

[27] Strictly speaking, crime is still possible. If the sovereign catches me and
punishes me, then I did something wrong. But wrongdoing is always punished, for
if it is not, then it was not wrongdoing after all. So although not everything that
happens is right, in one sense everything turns out all right.

sympathetic motives. Then if you lack sympathetic motives, you lack obligations. Your obligations vary along with your motives, and so you can do no wrong. Suppose, as Hume sometimes seemed to think, that the authority of our reasons for action must be derived from the strength of our desires. Then you will always do what you have reason to do, and you can do no wrong. As Joseph Butler would later point out, this sort of argument shows that authority cannot be reduced to any kind of power. And the relation in which moral claims stand to us is a relation of authority, not one of power.[28]

So we are faced with a dilemma. If we try to derive the authority of morality from some natural source of power, it will evaporate in our hands. If we try to derive it from some supposedly normative consideration, such as gratitude or contract, we must in turn explain why that consideration is normative, or where its authority comes from. Either its authority comes from morality, in which case we have argued in a circle, or it comes from something else, in which case the question arises again, and we are faced with an infinite regress.

The realist's response is to dig in his heels. The notion of normativity or authority is an irreducible one. It is a mistake to try to explain it. Obligation is simply there, part of the nature of things. We must suppose certain actions to be obligatory in themselves if anything is. According to Clarke, it is a fact about certain actions that they are *fit to be done.* Richard Price argues that unless we may say that some actions are *in themselves* right or wrong it is impossible that we should have any obligations; and in turn that if some actions *a r e* intrinsically right or wrong it is senseless to ask why we are obligated to do or avoid them.[29] Because of these views, Clarke and Price were primarily polemical writers.

---

[28] See Butler, "Upon Human Nature," Sermon 2 of the *Fifteen Sermons Preached at the Rolls Chapel* and of the *Five Sermons,* pp. 39-40.

[29] These positions are defended throughout in Clarke, *A Discourse concerning the Unchangeable Obligations of Natural Religion;* and Price, *A Review of the Principal Questions in Morals.*

They could not prove that obligation was real and instead devoted their efforts to rebutting what they took to be skeptical attacks.

Early twentieth-century rational intuitionism, represented by the work of Prichard, Ross, and Moore, follows a similar pattern. It is clearest in Prichard's classic essays: "Does Moral Philosophy Rest on a Mistake?" and "Duty and Interest." Prichard argues that it makes no sense to ask why you should be moral. If I give you a moral reason — such as, "it is your duty" — then my answer is circular, since it assumes you should be moral. If I give you a self-interested reason — such as, "it will make you happy" — then my answer is irrelevant. That is not the reason why you should be moral; you should be moral because it is your duty. If a question admits only of answers that are either circular or irrelevant then it must be a mistake to ask it. And if that is the question of moral philosophy, Prichard thinks, then moral philosophy rests on a mistake. Obligations just exist, and nobody needs to prove it.

As these arguments show, realism is a metaphysical position in the exact sense criticized by Kant. We can keep asking why: "Why must I do what is right?" — "Because it is commanded by God" — "But why must I do what is commanded by God?" — and so on, in a way that apparently can go on forever. This is what Kant called a search for the unconditioned — in this case, for something that will bring the question "Why must I?" to an end. The unconditional answer must be one that makes it impossible, unnecessary, or incoherent to ask why again. The realist move is to bring this regress to an end by fiat: he declares that some things are *intrinsically* normative. Prichard joins Clarke and Price in asserting this about obligatory actions, while Moore thinks there are intrinsically good states of affairs.[30] The very nature of these intrinsically normative entities is supposed to forbid further questioning. Having discovered that he needs an unconditional answer, the realist straightaway concludes that he has found one.

---

[30] See Moore, *Principia Ethica,* and also "The Conception of Intrinsic Value."

A comparison will help to show why this is metaphysical. Consider the cosmological argument for the existence of God, which purports to prove God's existence by proving that there must be a necessarily existent being. It runs this way: Somewhere there must be an Entity whose existence is necessary in itself. For if an Entity is contingent, it can either exist or not exist. How then can we explain its existence? Well, some other Entity must have brought it into being, have made it exist. What then about this other Entity? Is it necessary or contingent? And if it is contingent then what in turn made it exist? In this way we generate a regress, which can only be brought to an end if some Entity exists necessarily, that is, if there is some Entity about which it is impossible, unnecessary, or incoherent to ask why It exists. So there must be such an Entity, and that is God.

As Hume pointed out in his *Dialogues concerning Natural Religion,* there are two problems here.[31]  First of all, so far as the argument goes, *anything* could be the necessary being. It could be matter, or the universe, or the sun. In placing the necessity in God, the cosmologist has simply placed it where he wanted to find it. And second, unless you assume that even contingent beings must in some sense be necessary — that is, that there must be an explanation that shows that they must have existed —the argument cannot even get started.[32]

Moral realism is like that. Having discovered that obligation cannot exist unless there are actions that it is necessary to do, the realist concludes that there must be such actions and that they are the very ones that we have always thought were necessary, the

---

[31] Hume, *The Dialogues concerning Natural Religion,* part IX.

[32] It may not be obvious that Hume makes this second argument, but it is implied by one he does make. Hume has Cleanthes say, "In such a chain too, or succession of objects, each part is caused by that which preceded it, and causes that which succeeds it. Where then is the difficulty?" (p. 190). That of course amounts to a denial that the items in the "chain" need be in any sense necessary. It is worth noting that the cosmologist Cleanthes explicitly quotes in the course of his criticism is Samuel Clarke.

traditional moral duties. And the same two problems exist. The realist like the cosmologist places the necessity where he wanted to find it. And the argument cannot even get started, unless you assume that there are some actions that are necessary to do.

But when the normative question is raised, these are the exact points that are in contention — whether there is really *anything* I must do, and if so whether it is *this*. So it is a little hard to see how realism can help.

Yet realism is seen by many as the only hope for ethics, the only option to skepticism, relativism, subjectivism, and all the various ways of thinking that the subject is hopeless. There are, I think, two reasons for this. One is clear from the arguments that I have just reviewed. It can look as if granting the existence of intrinsically normative entities is the only way to bring the endless question "why" to an end and still save obligation. The other is based on a confusion. Realism may be defined in a way that makes it look like the *logical* opposite of skepticism — say, for instance, as the existence of moral truth. But considered as a substantive position, realism actually involves more than that.

Let me explain. There is a trivial sense in which everyone who thinks that ethics isn't hopeless is a realist. I will call this *procedural* moral realism, and I will contrast it to what I will call *substantive* moral realism. Procedural moral realism is the view that there are answers to moral questions; that is, that there are right and wrong ways to answer them. Substantive moral realism is the view that there are answers to moral questions *because* there are moral facts or truths, which those moral questions ask *about.*

To see the difference, it helps to consider normative realism more generally. The procedural normative realist thinks that when we ask practical questions like "What must I do?" or "What is best in this case?" or "How should I live?" there are correct and incorrect things to say. This is not just a view about morality. Suppose the correct answer to the question "How should I live?" is "Just as you like." Then people deluded by duty who don't live

as they like would be making a *mistake.* The view that there is *no* normative truth about action is the view that it is impossible to fail to do what you have reason to do, or should do, or ought to do: it is the view, more or less, that it doesn't matter what you do. Procedural realism isn't completely trivial, for it does have an opposite, but that opposite is a kind of nihilism. The denial of procedural normative realism says that there is no ought, should, must, or reason at all.

But procedural realism does not require the existence of intrinsically normative entities, either for morality or for any other kind of normative claim. It is consistent with the view that moral conclusions are the dictates of practical reason, or the projections of human sentiments, or the results of some constructive procedure like the argument from John Rawls's original position.[33] As long as there is some correct or best procedure for answering moral questions, there is some way of applying the concepts of the right and the good. And as long as there is some way of applying the concepts of the right and the good, we will have moral and more generally normative truth. Statements employing moral concepts will be true when those concepts are applied correctly.

Perhaps an example will help here. Most people suppose that the means/end relation is normative, in the sense that the fact that a certain action is a means to your end provides you with a reason to do it. Very few people have ever supposed that this requires an adjustment in the metaphysics of the Modern Scientific World View, say, by the introduction of intrinsically normative entities into our ontology. But how then do we establish that this relation is normative? One plausible answer comes from Kant. Kant tells us that the means/end relation is normative because of a principle of practical reason that he calls the hypothetical imperative. The hypothetical imperative tells us that if we will an end, we have a

---

[33] See *A Theory* of *Justice,* part I. Rawls characterizes his conception of justice as a "Kantian constructivist" one in "Kantian Constructivism in Moral Theory: The Dewey Lectures 1980."

reason to will the means to that end. This imperative, in turn, is
not based on the recognition of a normative fact or truth, but
simply on the nature of the will. To will an end, rather than just
wishing for it or wanting it, is to set yourself to be its cause. And
to set yourself to be its cause is to set yourself to take the available
means to get it.[34]  So the argument goes from the nature of the
rational will to a principle that describes a procedure according to
which such a will must operate, and from there to an application
of that principle that yields a conclusion about what one has a rea-
son to do. And Kant of course thought that in a similar way moral
principles could be shown to be principles of practical reasoning
that are based on the nature of the will and yield conclusions about
what we ought to do. There are then facts, moral truths, about
what we ought to do, but that is not because the actions are in-
trinsically normative. They inherit their normativity from prin-
ciples that spring from the nature of the will — the principles of
practical reasoning.

What distinguishes substantive from procedural realism is a
view about the relationship between the answers to moral ques-
tions and our procedures for arriving at those answers. The pro-
cedural moral realist thinks that there are answers to moral ques-
tions *because* there are correct procedures for arriving at them.
But the substantive moral realist thinks that there are correct pro-
cedures for answering moral questions *because* there are moral
truths or facts that exist independently of those procedures, which
those procedures track.[35]  Substantive realism conceives the pro-
cedures for answering normative questions as ways of *finding out*
about a certain part of the world, the normative part. To that
extent, substantive moral realism is distinguished not by its view
about what kind of truths there are, but by its view of what kind

----

[34] *Foundations of the Metaphysics of Morals,* pp. 414–17;  in Beck's translation,
pp. 31–35.

[35] Substantive realism is a version of procedural realism, of course; what dis-
tinguishes it is its account of *why* there is a correct procedure for answering moral
questions.

of subject ethics is. It conceives ethics as a branch of knowledge: knowledge of the normative part of the world.

Substantive moral realism has been criticized in many ways. It has been argued that we have no reason to believe in intrinsically normative entities or objective values. They are not harmonious with the Modern Scientific World View, nor are they needed for giving scientific explanations. Since the time of Hume and Hutcheson, it has been argued that there is no reason why such entities should motivate us, disconnected as they are from our natural sources of motivation. Many of these criticisms have been summed up in John Mackie's famous "Argument from Queerness." Here it is in Mackie's own words:

> If there were objective values, then they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe. Correspondingly, if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else. . . .
>
> Plato's Forms give a dramatic picture of what objective values would have to be. The Form of the Good is such that knowledge of it provides the knower with both a direction and an overriding motive; something's being good both tells the person who knows this to pursue it and makes him pursue it. An objective good would be sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it. Similarly, if there were objective principles of right and wrong, any wrong (possible) course of action would have not-to-be-doneness somehow built into it.[36]

And nothing, Mackie suggests, could be like that.

Of course Mackie doesn't really prove that such entities couldn't exist. But he does have a point, although I think it is not the point

---

[36] J. L. Mackie, *Ethics: Inventing Right and Wrong,* pp. 38 and 40,

he meant to make. If someone falls into doubt about whether obligations really exist, it doesn't help to say, "Ah, but indeed they do. They are *real* things." To see this, imagine a case where morality requires you to face death rather than do a certain action. You ask the normative question: you want to know whether this terrible claim on you is justified. Is it really true that this is what you must do? The realist's answer to this question is simply "Yes." That is, *all* he can say is that it is *true* that this is what you ought to do. This is of course especially troublesome when the rightness of the action is supposed to be self-evident and known through intuition, so that there is nothing more to say about it. If the realist is not an intuitionist he can go back and get you to review the reasons why the action is required. Prichard says explicitly that it is only because people sometimes need to do this before they can see the necessity of an action that the question "Why should I be moral?" appears to make sense when actually it does not.[37] So we need to remind ourselves that the action promotes pleasure, or is called for by a universalizability criterion, or fosters social life. But this answer appears to be off the mark. It addresses someone who has fallen into doubt about whether the action is really required by morality, not someone who has fallen into doubt about whether moral requirements are really normative.

Now, to be fair to Prichard, it is clear from his essays that he takes words like "right" and "obligatory" to imply normativity by definition. These terms, as he sees it, are normatively loaded, so that it is incorrect to *say* that an action is right or obligatory unless we are already sure that we really have to do it. In one sense, that's fine: it is six of one, half a dozen of the other, whether we ask, "Is this action really obligatory?" or "Is this obligation really normative?" If we take obligation to imply normativity, then the first question is the same as the second. The trouble with Prichard's way of talking about these matters is more a heuristic one. The question "Is this action really obligatory?" can be understood as

[37] See Prichard, "Does Moral Philosophy Rest on a Mistake?" p. 8.

a question about whether moral concepts have been applied cor-
rectly in this case —whether, for instance, the requirement can
really be derived from the categorical imperative or the principle
of utility or some other moral principle. And that is a different
question from the question how this obligation or any obligation
can be normative. Prichard's way of approaching the matter there-
fore leads us to confuse the question of correct application with the
question of normativity. And this actually happened to Prichard
himself. For it led him to think that once we have settled the
question of correct application, there can be nothing more to say
about the normative question.[38]

And that is the problem with realism: it refuses to answer the
normative question. It is a way of saying that it cannot be done.
Or rather, more commonly, it is a way of saying that it need not be
done. For of course if I *do* feel confident that certain actions really
are required of me, I might *therefore* be prepared to believe that
those actions are intrinsically obligatory or objectively valuable,
that just is a property they have. Just listen to what Samuel Clarke
says: "These things are so notoriously plain and self-evident, that
nothing but the extremest stupidity of mind, corruption of man-
ners, or perverseness of spirit, can possibly make any man entertain
the least doubt concerning them." [39] Well, obviously *he* isn't wor-
ried. But suppose you are? Perhaps his confidence will make you
take heart, but it is hard to see how else this could help.

The difficulty here is plain. The metaphysical view that in-
trinsically normative entities or properties exist must be *supported
by* our confidence that we really do have obligations. It is because
we are confident that obligation is real that we are prepared to
believe in the existence of some sort of objective values. But for
that very reason the appeal to objective values cannot be used to

[38] See Lecture 2, note 30, for discussion of a parallel problem in Prichard's
attitude toward skepticism about belief. The point is perhaps even clearer in that
case.

[39] Clarke, *A Discourse concerning the Unchangeable Obligations of Natural Re-
ligion,* Raphael I, p. 194; Schneewind I, p. 296.

support our confidence. And the normative question arises when our confidence has been shaken, whether by philosophy or by the exigencies of life. So realism cannot answer the normative question.

Some contemporary realists, such as Thomas Nagel, have argued that realism need not commit us to the existence of curious metaphysical objects like Plato's Forms or Moore's nonnatural intrinsic values. According to Nagel, we need only determine whether certain natural human interests, like our interest in having pleasure and avoiding pain, have the normative character that they appear to us to have. The point is not to look for some sort of specially normative *object,* but to look *more objectively* at the apparently normative considerations that present themselves in experience. That you are, say, in pain, *seems* like a reason to change your situation; the question is whether it is one.[40] Utilitarianism itself can be seen as a naturalistic form of realism, and versions of it have been defended as such by contemporary realists like David Brink and Peter Railton.[41] Contemporary realists argue that there is no need to make the right and the good into mysterious entities. Nothing seems more obviously normative than pleasures and pains, or desires and aversions, or our natural interests. So the realist need not assume, as Mackie supposes, that believing in objective values is believing in some sort of peculiar entities. We need only believe that reasons themselves exist.[42]

But if we take Mackie's point in the way that I have suggested, this leaves the problem in place. For how do we determine that these reasons exist? Like his rationalist predecessors, Nagel asserts that all we can do is rebut the skeptical arguments against the reality of reasons and values. Once we have done that, there is no special reason to doubt they exist.[43] And then when you see some-

---

[40] Nagel, *The View from Nowhere,* p. 157.

[41] Brink, *Moral Realism and the Foundations of Ethics,* especially chapter 8; and Railton, "Moral Realism," pp. I89ff.

[42] Nagel, *The View* from *Nowhere,* p. *144.*

[43] Ibid., pp. 143–44. Nagel says: "It is very difficult to argue for such a possibility [the reality of values], except by refuting arguments against it" (p. 143).

thing that appears to be a reason, such as, say, your desire to avoid pain, the best explanation of this appearance is that that's what it is — it's a reason.[44]

And there's nothing wrong with that. But it is an expression of confidence and nothing more. Just listen to what Thomas Nagel says: "In arguing for this claim, I am somewhat handicapped by the fact that I find it self-evident."[45] Nagel's manners are better than Samuel Clarke's, but his predicament is the same. He isn't worried.

Now I'd like to pause for a moment and say something that I hope will be helpful about why the normative question slips so easily through our fingers. Earlier I said that in a sense Prichard *is* asking the normative question. For him "obligation" is a normatively loaded word. If "obligation" is a normatively loaded word, then the normative question is whether certain actions are really obligatory. If "reason" is the normatively loaded word, as Nagel thinks, then the normative question is whether obligations give us reasons, or more generally whether we have any moral reasons. If "objective" is a normatively loaded word, as Mackie seems to

[44] Ibid., p. 141. He actually says: "The method is to begin with the reasons that appear to obtain from my own point of view and those of other individuals; and ask what the best perspectiveless account of those reasons is." Because Nagel believes in the existence of reasons, rather than Forms or Non-Natural properties, it would be easy to suppose that he is only what I have here called a "procedural realist." Actually the issue is a bit complicated. I categorize him here as a substantive realist because he seems to believe, as the passage quoted shows, that our relation to reasons is one of seeing or knowing that they are there. As I have just argued, there is a way in which this view of ethics as an epistemological subject is the essential characteristic of substantive realism. But in §II of my paper "The Reasons We Can Share: An Attack on the Distinction between Agent-Relative and Agent-Neutral Values" I argue that it is *Possible* to understand the projects Nagel prosecutes in both *The Possibility of Altruism* and *The View from Nowhere* as constructivist projects and that Nagel himself wavers between that way and a realist way of construing his own work. If we read Nagel as a constructivist then he is only a procedural realist.

[45] Nagel, *The View from Nowhere,* pp. 159–60. Actually he says this about the idea that pain and pleasure provide "agent-neutral" rather than "agent-relative'' reasons. But he says things pretty much like this about whether reasons exist at all. For instance on p. 157 he says that if there is no special reason to doubt the existence of reasons then denying that pain provides a reason to change your situation "seems meaningless."

think, then the normative question is whether obligations are objective, and so on.

Discussions of normativity often founder because of unexamined assumptions about the normatively loaded word. There are two problems here. First, philosophers making different assumptions about which is the normatively loaded word may fail to understand each other. The second and perhaps more serious problem is that all of the ways of formulating the normative question that I have just mentioned suffer from the fact that they are readily confused with different questions. As I pointed out in my discussion of Prichard, the question whether the action is "really obligatory" can be confused with the question whether the moral concept really applies. In a similar way, the question whether an obligation really provides a reason can be confused with the question whether it provides an adequate motive. Again, the question whether the obligation is *objective* can be confused with the question whether the moral concept is one whose application is determinate or sufficiently "world-guided." In all of these cases, the philosopher is led to think that settling the other question, whatever it is, is a way of settling the normative question. And in all of these cases it is not.

Part of what I have tried to do in this lecture is to raise the normative question in a way that is independent of our more ordinary normative concepts and words. No doubt this has sometimes been confusing as I have tried to describe and compare the views of philosophers who use different terms to imply normativity. The point is not that I think that there is no normatively loaded word. Of course we will have to use some words to imply normativity, but we can choose any of the above ways of talking or others. All that matters there is that we agree, so that we will understand each other. But the interesting question is not how we decide to talk about the issue. The interesting question is why there should be such an issue: that is, why human beings need normative concepts and words. And substantive realism — to get back now to my argument — is not merely the view that "obligation" (as Prichard

thinks) or "good" (as Moore thinks) or "reason" (as Nagel thinks) are normative words that we know how to apply correctly. It is a view — and a false one — about why human beings have normative words.

What is really wrong with substantive realism is its view about the source of normativity. Why do we use normative concepts like good, right, reason, obligation? According to the substantive realist, it is because we grasp that there are things that have normative properties. Some things *appear* normative, and there is no reason to doubt that they are what they seem. We have normative concepts because we've spotted some normative entities, as it were wafting by.

According to substantive realism, then, ethics is really a theoretical or epistemological subject. When we ask ethical questions, or normative questions generally, there is something about the world that we are trying to find out. The world contains a realm of inherently normative entities, whose existence we have noticed, and the business of ethics, or of practical philosophy more generally, is to investigate them further, to learn about them in a more systematic way. But isn't ethics supposed to be a practical subject, a guide to action? Well, the realist will grant that the eventual point is to apply all this knowledge in practice. Look at the result of that view: according to the substantive realist, the moral life is the most sublime feat of technical engineering, the application of theoretical knowledge to the solution of human problems. And in general human life and action consist in the application of theories, theories about what is good. Now that is surely wrong.

I've just been criticizing moral realism for asserting that we have moral concepts because we have noticed some moral entities in the universe. There's another argument on the contemporary scene that makes what looks like a similar criticism, but takes this criticism as a reason for moral skepticism. Since I am not arguing for skepticism, I want to say something about that. This other argument is that we have no reason to believe in the existence of

moral entities or facts, because we do not need to assume the existence of such entities or facts in order to explain the moral phenomena. We need to assume that physical entities and facts exist in order to explain our observations of and beliefs about the "external world," but we do not need to assume that moral facts or entities exist in order to explain our moral beliefs and motives. Explanations of those can proceed in entirely psychological terms. So, the argument suggests, the best explanation of why I see a rock is that there is one. But the best explanation of why I disapprove of killing is that I was brought up in a certain way.[46]

A more carefully formulated version of this argument has some force against substantive moral realism, and this is a point I will come back to. But I want to start by saying what I think is wrong with this argument. As it is stated, this argument looks as if it should work against *any* form of normative realism. It should have just as much force against the existence of *theoretical* normative truth (that x is a reason to *believe* y) as it does against *practical* normative truth (that x is a reason to *do* y). We can after all explain the occurrence of people's beliefs merely in terms of the causes of those beliefs and leave their reasons out of it. Even if people's beliefs are caused by their thoughts about what reasons they have, we can explain the beliefs simply as caused by those thoughts. This does not commit us to saying that the reasons that appear in the contents of those thoughts are real. I may tell the truth because I think lying wrong, but in order to explain my honesty you need not suppose that my reason is real. It is enough that I think so. In the same way, I think that I am mortal because I am human, but in order to explain why I believe I am mortal you need not suppose that my reason is real. Again it is enough that I think so. So we don't need to assume that theoretical reasons exist in order to explain the occurrence of beliefs.[47] But we

[48] The locus classicus is perhaps Gilbert Harman, *The Nature of Morality: An Introduction to Ethics,* chapter 1.

[47] Actually, however, there is a problem explaining how human beings could come to have the illusion that there are such things as theoretical and practical rea-

cannot coherently take that fact as a *reason* to doubt that there is any such thing as a reason for belief. For if there is no such thing as a reason for belief, there is *ipso facto* no reason for believing this argument. And —to echo Clarke himself —if instead we admit that there are reasons for belief, then why not admit that there are reasons for action as well?

The trouble with drawing skeptical conclusions from the fact that a belief in normative truth is not needed to explain what people think or do is that it assumes that explanation and description of the phenomena is the sole or primary function of human concepts. That amounts to supposing that the business of human life is the construction and application of theories. And the reason the argument has some force against substantive realism is that substantive realism implicitly shares that assumption. The substantive realist assumes we have normative concepts because we are aware that the world contains normative phenomena, and we are inspired by that awareness to construct theories about them.

But that is not why we have normative concepts. The very enterprise we are engaged in right now shows why we have those: it is because we have to figure out what to believe and what to do. Normative concepts exist because human beings have normative problems. And we have normative problems because we are self-conscious rational animals, capable of reflection about what we ought to believe and to do. That is why the normative question can be raised in the first place: because even when we are inclined to believe that something is right and to some extent feel ourselves moved to do it we can still *always* ask: "But is this really true?" and "Must I really do this?"

Normative concepts like right, good, obligation, and reason are our names for the solutions to normative problems, for what it is we are looking for when we face them. And if we sometimes succeed in solving those problems, then there will be normative

---

sons if no such things exist at all. But the reason why we have the concept of a "reason" does not therefore have to be that we *notice* that they exist.

truths: that is, statements that employ normative concepts correctly. So it is true that the assumption of a realm of inherently normative entities or objective values is not needed to explain the existence of normative concepts or the resulting existence of a category of normative truths. It is not because we notice normative entities in the course of our experience, but because we are normative animals who can question our experience, that normative concepts exist.

## Conclusion

Contemporary defenses of substantive moral realism almost always arise in the same way. They are always initiated by *somebody else,* a self-proclaimed spokesperson for the Modern Scientific World View. Armed with the distinction between facts and values, or brandishing Ockham's razor like a club, the spokesperson for the Modern Scientific World View declares that there cannot be ethical knowledge, that we can explain the moral phenomena without positing the existence of moral entities or facts, or that intrinsically normative entities are just too queer to exist. And the moral philosopher, frantic with the sense of impending loss, rushes to the defense of ethical knowledge. And nobody pauses to ask whether ethical knowledge, or indeed any sort of knowledge at all, is what we really want here in the first place.

Is the normative question a request for knowledge? To raise the normative question is to ask whether our more unreflective moral beliefs and motives can withstand the test of reflection. The Platonic realist thinks that we can answer that question by taking a closer look at the *objects* of our beliefs and motives, to discover whether they are really the True and the Good. Nagel thinks we should take a closer look at the beliefs and motives themselves, to discover whether they are really reasons. But no such *discovery* is ever made. The realist's belief in the existence of normative entities is not based on any discovery. It is based on his *confidence* that his beliefs and desires are indeed normative. But if confidence

can support a metaphysics that in turn is supposed to support the claims of morality, why can't confidence support the claims of morality more directly?

In the next lecture I will examine the views of some philosophers who reject the idea that knowledge is what we need for normativity and put something more like confidence in its place. According to these philosophers, morality is not grounded in our apprehension of truths about objective values. It is grounded in human nature and certain natural human sentiments. The normative question is then whether it is good to have such a nature and to yield to its claims. Normativity will be established, not by knowledge, but by our own reflective endorsement of our moral nature.

## LECTURE II: REFLECTIVE ENDORSEMENT

*Introduction*

At the end of the last lecture I argued that normativity is a problem for human beings because of our reflective nature. Even if we are inclined to believe that an action is right and even if we are inclined to be motivated by that fact, it is always possible for us to call our beliefs and motives into question. This is why, after all, we seek a philosophical foundation for ethics in the first place: because we are afraid that the true explanation of why we have moral beliefs and motives might not be one that sustains them. Morality might not survive reflection.

The view I am going to describe in this lecture takes its starting point from that thought. It applies one of the best rules of philosophical methodology: that a clear statement of the problem is also a statement of the solution. If the problem is that morality might not survive reflection, then the solution is that it might. If we find upon reflecting on the true moral theory that we still are inclined to endorse the claims that morality makes on us, then morality will be normative. I call this way of establishing normativity the "reflective endorsement" method.

The reflective endorsement method has its natural home in theories that reject realism and ground morality in human nature. In the modern period it makes its first appearance in the work of the sentimentalists of the eighteenth century. They explicitly rejected the realism of the rationalists and argued that the moral value of actions and objects is a projection of human sentiments. As Hume famously says:

> Take any action allow'd to be vicious: Willful murder, for instance. Examine it in all lights, and see if you can find that matter of fact, or real existence, which you call vice. In whichever way you take it, you find only certain passions, motives, volitions and thoughts. There is no other matter of fact in the case. The vice entirely escapes you, as long as you consider the object. You can never find it, till you turn your reflexion into your own breast, and find a sentiment of disapprobation, which arises in you, towards this action. Here is a matter of fact; but 'tis the object of feeling, not of reason. It lies in yourself, not in the object.[1]

Strictly speaking, we do not disapprove the action because it is vicious; instead, it is vicious because we disapprove it. Since morality is grounded in human sentiments, the normative question cannot be whether its dictates are true. Instead, it is whether we have reason to be glad that we have such sentiments and to allow ourselves to be governed by them. The question is whether morality is a good thing for us.

Of course the sentimentalists were not the first to ground morality in human nature. Some of the classical Greek philosophers, in particular Aristotle, did so as well. So it is not surprising that the reflective endorsement method has reemerged in some recent moral thought of Aristotelian inspiration, namely that of Bernard Williams.[2]  Like Hume, Williams rejects realism and defends in

---

[1] Hume, *Treatise of Human Nature,* III.l.i, p. 469.

[2] These remarks will naturally raise the question whether Aristotle himself used the reflective endorsement method. In chapter 3 Williams makes a good case for the

its place a theory that grounds morality in human dispositions. And like Hume, he finds that the answer to the normative question rests in whether those dispositions are ones we have reason to endorse.

My purpose in this lecture is to explain this method of establishing normativity in more detail and to defend it against certain natural objections that arise from the realist camp. My aim will not be to criticize this view. Instead, I will end by saying why I think the logical consequence of Hume and Williams's theory of normativity is the moral philosophy of Kant.

## David Hume

The choice of Hume as the major traditional representative of a theory of normativity might seem perverse. The pose Hume strikes in his moral philosophy is that of the scientist, whose task is to explain the origin of moral ideas. In his essay "Of the Different Species of Philosophy," Hume firmly separates two different ways of treating moral philosophy, which we may call "theoretical" and "practical." Theoretical or "abstruse" philosophers regard human nature as a subject of speculation and are concerned to discover the principles that regulate our understanding, excite our sentiments, and cause us to approve and disapprove as we do.[3] Practical philosophers, by contrast, are interested in inciting us to good conduct. Their work, as Hume puts it, is to paint virtue in "amiable colors, borrowing all helps from poetry and eloquence, and treating their subject in an easy and obvious manner, and such as is best fitted to please the imagination and engage the affections."[4] Hume compares the theoretical philosopher to an anato-

---

claim that reflective endorsement is at least involved in Aristotle's method of justifying morality. (See below). But Aristotle's teleological conception of the world adds another element to his conception of normativity. In these lectures I am addressing modern methods of establishing normativity, so I have not discussed Aristotle's views directly. What I think about them will, however, become apparent in the course of Lecture III.

[3] Hume, *Enquiry concerning Human Understanding,* p.   6.

[4] Ibid., p. 5.

mist and the practical philosopher to a painter.[5] The business of
the anatomist is to explain what causes us to approve of virtue;
the business of the painter is to make virtue appealing. And Hume
styles himself a theoretical philosopher: his aim is to reveal the
elements of the mind's "anatomy" that make us approve and dis-
approve as we do.

The odd thing about this way of dividing up the philosophical
enterprise is that the normative question seems to fall between
the cracks. Neither the anatomist nor the painter seems to be
interested in the *justification* of morality's claims. The theoretical
philosopher is concerned only with providing a true explanation
of the origin of moral concepts. The practical philosopher is a
preacher or a Mandevillian politician. His task is to get people to
behave themselves in socially useful ways, and he is prepared to
use "all helps from poetry and eloquence." So we have explana-
tion on the one hand and persuasion on the other, but no branch
of moral philosophy that is concerned with justification.

It is not that Hume takes it for granted that morality's claims
can be justified to the individual. He explicitly denies that the
truth of his theoretical account depends at all on "its tendency
to promote the interests of society." He thinks it is conceivable
that knowledge of the true moral theory would undermine the
commitment of individuals to moral conduct. Yet he also asserts
that "a man has but a bad grace, who delivers a theory, however
true, which . . . leads to a practice dangerous and pernicious." As
he says: "The ingenuity of your researches may be admired, but
your systems will be detested; and mankind will agree, if they
cannot refute them, to sink them, at least, in eternal silence and
oblivion. Truths which are *pernicious* to society, if any such there
be, will yield to errors which are salutary and *advantageous."* But
although he admits that this could happen, he thinks that it doesn't.
Although he is not supposed to be a practical philosopher, Hume

---

[5] Ibid., pp, 9-10; *A Treatise of Human Nature,* III.iii.6, pp. 620–21. I owe a
debt to Charlotte Brown for many useful discussions of this issue.

cannot resist pointing out that his account of the origin of moral ideas *does* make virtue attractive. According to his theory, he points out, virtue asks nothing of us but "gentleness, humanity, beneficence, and affability." And he urges: "She talks not of useless austerities and rigours, suffering and self-denial. She declares that her sole purpose is to make her votaries and all mankind, during every instant of their existence, if possible, cheerful and happy..." [6] So Hume thinks that his account of morality, though itself theoretical and abstruse, can be used by the practical philosopher to good effect.

One can, of course, take Hume to be saying merely that his theory is a gold mine for practical philosophers. But I think he has something more in mind. Normativity is not the provenance of either the theoretical or the practical philosopher because it will emerge, if it does emerge, in the way the two sides of philosophy interact. If the true account of our moral nature were one that made us want to reject its claims, then practical philosophers, as the guardians of social order, would have to make sure that the truth was not known. But if practical philosophers can get people to accept the claims of morality simply by telling them the truth about the nature of morality, then the claims of morality are *justified.* Hume is claiming that his theory is normative — or so I will now argue.

According to Hume, moral judgments are based on sentiments of approval and disapproval that we feel when we contemplate a person's character from what he calls "a general point of view."[7] Taking up the general point of view regulates our sentiments about a person in two ways. First, we view the person not through the eyes of our own interests, but instead through the eyes of our sympathy with the person herself and her friends, family, neighbors, and colleagues.[8] We are sympathetically pleased or pained

---

[6]All of the quotations in this paragraph are from Hume, *Enquiry concerning the Principles of Morals,* p. 279.

[7] Hume, *A Treatise of Human Nature,* III.3.i, pp. 581–82.

[8] Ibid., p. 582.

by the good or bad effects of her character on those with whom she usually associates, the people Hume calls her "narrow circle." [9] Second, we judge her characteristics according to the usual effects of such characteristics, rather than according to their actual effects in this or that case. As Hume puts it, we judge according to "general rules." [10]

These two regulative devices bring a kind of objectivity to our moral judgments. Judging in sympathy with the narrow circle and according to general rules, we are able to reach agreement, in the sense of a convergence of sentiments, about a person's character. We all approve and disapprove of the same characteristics, and as a result we come to share an ideal of good character. A person of good character, one whom we judge to have the virtues, is one who is useful and agreeable to herself and her friends. Since people love those who have useful and agreeable qualities, and since the perception of a lovable quality in ourselves causes pride, virtue is a natural cause of pride, and vice in the same way of humility. And since pride is a pleasing sentiment and humility a painful one, we have a natural desire to be proud of ourselves and to avoid the causes of humility. This gives us a natural desire to acquire the virtues and avoid the vices. The normative question, then, is whether we really have reason to yield to these desires and to try to be virtuous people.

I think this is the question Hume is raising in the last section of the *Enquiry concerning the Principles of Morals* when he says: "Having explained the moral *approbation* attending merit or virtue, there remains nothing but briefly to consider our interested *obligation* to it, and to inquire whether every man, who has any regard to his own happiness and welfare, will not find his account in the practice of every moral virtue."

---

[9] Ibid., III.3.iii, p. 602.

[10] Ibid., III.3.i, p. 585.

[11] Hume, *Enquiry concerning the Principles of Morals,* p. 278.

Hume proceeds to detail the ways in which the practice of virtue contributes to the moral agent's happiness. His fourfold division of the virtues into qualities useful and agreeable to self and others enables him to do this in very short order. No argument is needed to defend the qualities that make you useful and agreeable to yourself, for those contribute to your happiness by definition. Almost as little is required to defend the qualities that make you *agreeable* to others, for we all want others to like and admire us. To defend the qualities that are *useful* to others, Hume borrows a famous argument from Joseph Butler.[12] In order to be happy, we must have some desires and interests whose fulfillment will bring us satisfaction. And other-directed desires and interests are just as good for this purpose as self-absorbed ones. Indeed, in many ways they are better. Hume reminds us that any desire, "when gratified by success, gives a satisfaction proportioned to its force and violence." But benevolent desires have the additional advantages that their "immediate feeling . . . is sweet, smooth, tender, and agreeable" and that they make others like us and make us pleased with ourselves.[13] To be a morally good person, then, is conducive to your happiness or at least not inconsistent with it.

Now it might be thought that this argument is not intended to show anything about the goodness of being subject to motives of moral *obligation* and that therefore it cannot show anything about the normativity of obligation. For according to Hume's account a *naturally* virtuous person is one who acts, not from the motive of duty or obligation, but simply from some natural motive, such as benevolence, that a spectator would approve. No reason why you are *obligated* to perform virtuous actions has been given by the argument or is required by it; you perform virtuous actions because you have natural motives to do so; and the argument has simply shown that this is a good way for you to be.

---

[12] Butler, "Upon the Love of Our Neighbor," sermon 11 of the *Fifteen Sermons Preached at the Rolls Chapel;* sermon 4 of the *Five Sermons.*

[13] Hume, *Enquiry concerning the Principles of Morals,* p. 282.

But this would not be correct. For first, Hume admits that in a case where a person is aware of lacking a virtuous moral motive, he "may hate himself upon that account, and may perform the action without the motive, from a certain sense of duty." [14] And second, it turns out that in the case of what Hume calls the *artificial* virtues such as justice, this sense of duty is the motive that is normally operative." According to Hume, the first or natural motive for participating in a system of justice is self-interest. But this is not the usual motive for performing just *actions,* for just actions, taken singly, do not necessarily or even usually promote self-interest. What promotes self-interest is the existence of the *system* of justice. But the connection between individual just actions and the system is too "remote" to sustain interested motivation.[16] Instead, Hume argues, sympathy with the public interest causes us to disapprove of all unjust actions on account of their general tendency to bring down the system.[17] And this sympathy grounds a sense of duty that motivates us to avoid injustice. We avoid injustice because we would disapprove of ourselves —that is, we would feel humility —if we did not.

Furthermore, there are cases in which this sense of duty is the *only* available motive, for it can happen that an action, while it is of the type that tends to bring down the system of justice, will not in fact do that system any harm at all, and that the agent knows that. This is the plight of the famous "sensible knave" who poses the most difficult challenge to Hume's account of "interested obligation." The sensible knave, as Hume describes him, "may think that an act of iniquity or infidelity will make a considerable addition to his fortune, without causing any considerable breach in the social union and confederacy." [18] So why shouldn't he do it?

---

[14] Hume, *A Treatise of Human Nature*, 111.24 p.479.

[15] Ibid.

[16] Ibid., 111.2.ii, p. 499.

[17] Ibid., pp. 499–500.

[18] Hume, *Enquiry concerning the Principles of Morals,* p. 282.

This is, of course, a version of the familiar free-rider problem. The sensible knave wants to know why he should not profit from injustice when it will not damage his interests by endangering the system of justice. And here is Hume's surprising answer:

> I must confess that, if a man think that this reasoning much requires an answer, it will be a little difficult to find any which will appear to him satisfactory and convincing. If his heart rebel not against such pernicious maxims, if he feel no reluctance to the thoughts of villainy or baseness, he has indeed lost a considerable motive to virtue; and we may expect that his practice will be answerable to his speculation. . . . Inward peace of mind, consciousness of integrity, a satisfactory review of our own conduct; these are circumstances, very requisite to happiness, and will be cherished and cultivated by every honest man, who feels the importance of them.[19]

There's an old joke about a child who's glad he doesn't like spinach, since then he'd eat it, and he hates the disgusting stuff. Hume appears at first sight to be giving us that sort of reason for being glad we don't like injustice. *Of course* integrity will be cherished by honest people who feel the importance of it. But the sensible knave is questioning exactly that importance. The fact that we disapprove of injustice and therefore of ourselves when we engage in it can hardly be offered as a reason for endorsing our own disapproval of injustice.

Actually, however, in Hume's theory it can. Hume's theory of sympathy allows him to argue that an individual is likely to experience humility when he acts unjustly regardless of whether or not he believes that there is good reason to disapprove of the unjust action in the case at hand. For it follows from Hume's account of sympathy that the sentiments of others are contagious to us. And their sentiments about ourselves, in particular, have a tendency to get under our skins. So the fact that *other people* will disapprove

[19] Ibid., p. 283.

and dislike the sensible knave will be sufficient to provide him
with feelings of disapproval and dislike of himself. Of course a
knave will try to keep his knavish actions secret. But unless he is
very hardened indeed, even the knowledge that others *would* hate
him if they knew what he is up to will be enough to produce hu-
mility and self-hatred when he acts unjustly. As Hume says:

> By continual and earnest pursuit of a character, a name, a repu-
> tation in the world, we bring our own deportment and conduct
> frequently in review, and consider how they appear in the eyes
> of those who approach and regard us. This constant habit of
> surveying ourselves as it were, in reflection, keeps alive all the
> sentiments of right and wrong, and begets, in noble natures, a
> certain reverence for themselves as well as others, which is the
> surest guardian of every virtue.[20]

So Hume's reply to the sensible knave is not circular. Morality
provides a set of pleasures of its own, a set of pleasures that the
knave loses out on. Because of sympathy, the sense that you are
lovable and worthy in the eyes of others makes you lovable and
worthy in your own. For the same reason, the sense that you are
detestable in the eyes of others makes you detestable in your own.
And morality provides these feelings regardless of whether you
think that morality is justified or not. This fact enables Hume to
add the familiar claim that virtue is its own reward to his list of
the ways in which virtue promotes self-interest without any circu-
larity at all. Together, all of these arguments establish what Hume
calls our "interested obligation" to be moral.

The arguments I've just detailed give rise to two closely related
criticisms, which issue from the realist camp. First, you might
think that Hume is not giving an account of the normativity of
morality, but simply an account of our motives to be moral, and
one that falls afoul of Prichard's famous argument at that.[21] We

---

[20] Ibid., p. 276.

[21] For a discussion of Prichard's argument, see "Realism" in Lecture I.

should not practice virtue because it is in our interest, but rather for its own sake, so Hume's argument is irrelevant. But it is clear that Hume is not saying that we should perform *particular* virtuous or obligatory actions because it serves our own interest to do so. He is saying that it is in our interest to be *people who practice virtue for its own sake.* This is especially clear in the Butlerian argument used to defend the virtues that are useful to others. Neither the immediately agreeable sensations of benevolence nor its gratifications are available to anyone who is not genuinely and wholeheartedly concerned about others. The Butlerian argument is not meant to show that morality promotes some set of interests you already have, but rather that moral interests are good ones to have. What the argument establishes is the harmony of two potentially normative points of view, morality and self-interest.[22]

The second realist objection carries Prichard's worry to a higher level. This time the objector grants that Hume's argument is not offered to us as a wrongheaded theory of moral motivation, but rather as an attempt to establish normativity by showing that morality is good. But it says that even as such it fails. An argument that shows that virtue is good from the point of view of self-interest only shows that morality is extrinsically good or extrinsically normative. But what we need for normativity is to show that morality is intrinsically good or intrinsically normative. And now we come back to a thought familiar from our encounter with realism: that only something intrinsically normative can satisfy the demand for unconditional justification.

At this point it will help to turn to an earlier view Hume held about normativity. The arguments I have been detailing until now are for the most part from the *Enquiry concerning the Principles of Morals.* In *A Treatise of Human Nature,* Hume appealed to a

---

[22] The argument can therefore be seen as establishing what Rawls calls "congruence." See *A Theory of Justice,* p. 399. Rawls's own argument that justice is a good for the just person, in §86 of that work, is a congruence argument. On the use of congruence arguments among the eighteenth-century British Moralists, see Charlotte Brown, "Hume against the Selfish Schools and the Monkish Virtues."

more specific version of the reflective endorsement account, which I call "normativity as reflexivity." This view can help to answer the realist's worry.

Since Hume does not set this view out explicitly, I will start by explaining the grounds on which I attribute it to him. Book 1 of *A Treatise of Human Nature* ends in a mood of melancholy despair and skepticism, while book 3 concludes in a mood of triumphant affirmation. And this is because at the end of book 1 Hume finds that "the understanding, when it acts alone, and according to its most general principles, entirely subverts itself, and leaves not the lowest degree of evidence in any proposition, either in philosophy or common life." [23] Whereas at the end of book 3 Hume concludes that the moral sense "must certainly acquire new force, when reflecting on itself, it approves of those principles, from whence it is deriv'd, and finds nothing but what is great and good in its rise and origin." [24] The understanding, when it reflects on itself, falls into doubt about and so subverts itself. But the moral sense approves of and so reinforces itself. Therefore skepticism about the understanding is in order, but skepticism about morality is not.

These facts suggest that Hume is relying on an account of normativity that is completely general, applying to any kind of purportedly normative claim. Let me define two terms that will help express the view. Call a purportedly normative judgment a "verdict" and the mental operation that gives rise to it a "faculty." The faculty of understanding gives rise to beliefs, which are verdicts of conviction. The moral sense gives rise to moral sentiments or verdicts of approval and disapproval. The faculty of taste gives rise to verdicts of beauty. According to this theory, a faculty's verdicts are normative if the faculty meets the following test: *when the faculty takes itself and its own operations for its object,* **it** *gives a positive verdict.*

[23] Hume, *A Treatise of Human Nature,* I.4.vii, pp. 267–68.
[24] Ibid., III.3.vi, p. 619.

Now Hume clearly thinks that the understanding fails this test. A belief, according to Hume, is a sentiment of conviction, a lively idea of the thing believed. He argues that the harder we press the question whether we ought to believe our beliefs or whether they are likely to be true, the more the degree of our conviction — that is, the liveliness or vivacity of the ideas — will tend to diminish. So the more we reason about whether reasoning is likely to lead us to the truth, the less confidence in the results of reasoning we will end up having.[25]  The understanding in this way "subverts itself" when it reflects on its own operations.

But the moral sense passes the reflexivity test. In the conclusion of the *Treatise,* Hume asserts that, in explaining our moral judgments as arising from sympathy, he has traced them to a "noble source" and has given us a "just notion both of the generosity and capacity of our nature." He says:

> It requires but very little knowledge of human affairs to perceive, that a sense of morals is a principle inherent in the soul, and one of the most powerful that enters into the composition. *But this sense must certainly acquire new force, when reflecting on itself, it approves of those principles, from whence it is deriv'd, and finds nothing but what is great and good in its rise and origin.* . . . not only virtue must be approv'd of, but also the sense of virtue. And not only that sense, but also the principles from whence it is deriv'd. So that nothing is presented on any side, but what is laudable and good.[26]

Reflection on the origin of our moral sentiments only serves to strengthen those sentiments. The moral sense approves of its own origins and workings and so it approves of *itself*.

I believe that Hume got the idea for this theory of normativity from the moral sense theorist Francis Hutcheson. In his *Illustrations on the Moral Sense,* Hutcheson imagines a rationalist who

[25] See ibid., I.4.i, pp. 180–85.
[26] Ibid., III.3.vi, p. 619 (my emphasis).

objects that judgments of good and evil cannot come from a moral sense, because we judge our senses themselves to be good or evil.[27] For instance, we approve of a benevolence-approving moral sense, while we would deplore a malice-approving moral sense. These judgments would be trivial if they came from the benevolence-approving moral sense itself. The argument is a variant on one familiar argument against theological voluntarism —that if God determines what is good and evil then we cannot significantly judge God himself to be good —and like that argument it is intended to drive us to realism. Hutcheson replies that the goodness of a sense must be assessed from some point of view from which we make judgments of good and bad and that we have a limited number of such points of view to which we can appeal. We can judge the moral sense from the point of view of the moral sense itself; we can judge it from the point of view of benevolence toward others; or we can judge it from the point of view of our own self-interest.[28] What we cannot do is get outside of all of the points of view from which we judge things to be good or bad and still coherently ask whether something is good or bad. There is no place outside of our normative points of view from which normative questions can be asked.

The same argument can of course be made about the normativity of the verdicts of the understanding. If we fall into doubt about whether we really ought to believe what we find ourselves inclined to believe —that is, if we fall into doubt about whether our beliefs are true —we cannot dispel the doubt by comparing our beliefs to the world to see whether they are true. We have no access to the world except through the verdicts of the understanding itself, just as we have no access to the good except through the verdicts of the various points of view from which we make judgments of goodness. The only point of view from which we can

---

[27] Hutcheson, *lllustrations on the Moral Sense,* p. 133.

[28] Ibid., pp. 133–34.

assess the normativity of the understanding is therefore that of the understanding itself.

It is this line of thought, I believe, that gave Hume the idea for the reflexivity test. It is, of course, complicated in the moral case by the fact that there is more than one point of view from which we can assess things as good or bad. This is what, in the later work, leads Hume to use the more general reflective endorsement test instead. But we can see reflexivity and reflective endorsement as working together. For one of the reasons that the moral sense approves of itself *is* that morality contributes to our happiness, and the moral sense approves of anything that contributes to people's happiness.

Now let's go back to the more general form of the realist's objection. This was that the reflective endorsement test only shows that morality is extrinsically normative, whereas what we want to show is that it is intrinsically normative. The addition of the reflexivity test does show that or, rather, shows something that is very close. It shows that *human nature,* including our moral nature, is intrinsically normative, in a negative version of the sense required by the realist argument: there is *no intelligible challenge* that can be made to its claims. Within human nature, morality can coherently be challenged from the point of view of self-interest, and self-interest from the point of view of morality. Outside of human nature, there is no normative point of view from which morality can be challenged. But morality can meet the internal challenge that is made from the point of view of self-interest, and it also approves of itself. It is human nature to be governed by morality, and from every point of view, including its own, morality earns its right to govern us. We have therefore no reason to reject our nature and can allow it to be a law to us. Human nature, including moral government, is therefore normative and has authority for us.

Perhaps a comparison will make this thought seem more familiar. According to the teleological ethics of the ancient world, to

be virtuous is to realize our true nature, to be the best version of what we are. So it is to let our own nature be a law to us. And the Greeks thought that, since our own good would be realized in being the best version of what we are, we have every reason to be virtuous. Sentimentalism can be seen as a kind of negative surrogate of the teleological ethics of the ancient world. According to the sentimentalists, we have *no reason not to be* the best version of what we are.

## Bernard Williams

This brings us to a recent attempt to revive the virtue-oriented ethics of the ancient world. In chapters 8 and 9 of *Ethics and the Limits of Philosophy,* Bernard Williams argues that there is a contrast between the kind of objectivity we can hope to find in science and that which we can hope to achieve in ethics. Williams accepts a form of realism in the case of science, but rejects it in the case of ethics.

Williams frames this contrast in terms of convergence, that is, in terms of what might lead us to the best kind of agreement. In science, the ideal form of convergence would be this: we come to agree with one another in our beliefs because we are all converging on the way the world really is. In ethics, this sort of convergence is unavailable, and so another must be found.[29] This, as we will see, is where reflective endorsement comes in.

Williams begins by solving a problem in the formulation of his contrast. The problem is essentially the same as the one that drove Hume to suppose that only a reflexivity test could establish the normativity of belief: we can't go outside of our beliefs in order to determine whether they match the world or whether they correctly capture "the way the world really is." Williams puts the problem this way. We have a certain way of conceptualizing the world, a conceptual scheme. One thing we might mean in talking about "the way the world really is" is whether we have applied

---

[29] Williams, *Ethics and the Limits of Philosophy,* p. 136.

our concepts correctly. If we say that grass is green we have and if we say that it is pink we have not. This notion is unproblematic, but it leaves us no room to query our way of conceptualizing the world itself.[30]   Is our conceptual scheme adequate? Is it the correct one or the best one or the one that captures the most or the one that captures what is "really true" about the world? Philosophers will of course disagree on whether any of these questions are coherent and, if so, which one of them is the right one to ask. But since science leads us to modify our conceptual scheme, and we think of these modifications as improvements, it does appear that some such question is in order.

Williams proposes that we can capture the distinction between the way the world really is and the way it seems to us by the formation of a kind of limiting conception that he calls "the 'absolute conception' of the world." [31]   The idea involves a contrast between concepts that are more and less dependent on the particular perspective from which we view the world. For instance, we use color categories because we are visual, so color concepts like "green" and "pink" are dependent on something about our own perspective. The concept of a certain wavelength of light might be less dependent.

Williams associates two other properties with a concept's greater independence from our particular perspectives. First, our use of concepts that are more dependent on our own perspectives will be explained in terms of a theory that employs concepts that

---

[30] It is interesting that Prichard (on pp. 14–15 of "Does Moral Philosophy Rest on a Mistake?") argues that this is correct — there *is* no room to query our way of conceptualizing the world. Just as the only way to resolve a doubt about whether we are "really obligated" — whether obligation is normative — is to review the reasons why the action is right, so the only way to resolve a doubt about whether our beliefs are true is to review the reasons for those beliefs — in the language I am using here, to make sure the concept has been applied correctly. The problem here is the same as the one I discussed in "Realism" in Lecture I. By asking the normative question in the form "Is my belief really true?" Prichard is led to confuse it with the question whether my concepts have been applied correctly. But the normative question is a question about the status of the concepts, not about whether they have been correctly applied.

[31] Williams, *Ethics and the Limits of Philosophy,*  p. 139.

are less dependent. So, for instance, our use of color concepts might be explained by a theory of vision that employs wavelength concepts. Relatedly, and importantly, this theory (or some yet more absolute theory in which it is embedded) will also *justify* our belief that color vision is a form of *perception,* that is, a way of learning about the world, by the way that it explains it.[32] Color vision is a way of learning about the world because it gives us information about wavelengths, or something yet more ultimate, which we take to be part of reality. Second, the more independent of our own perspective a concept is, the more likely it is that it could be shared by investigators who were unlike us in their ways of learning about the world. Suppose that there are rational creatures on Jupiter who cannot see colors but do something more like hear them or perhaps feel them in the form of vibrations. They could not use color concepts, but they might be able to use wavelength concepts. The more independent concepts are more shareable.

Williams thinks that the nearest thing we have to a conception of the way the world really is is the conception of the world that is maximally independent of our own perspective. And if we and the alien investigators actually began to converge on such a conception (and of course to agree on what judgments are correct within it) then we would have reason to believe we were converging on what the world is really like. This would be the best case of convergence for science: our theories would come to converge with the theories of other investigators because all of us were converging on the way the world is.

Now consider what the parallel would be in ethics. Here too we must deal with a possible objection — namely that there is nothing analogous to perceptual judgments in ethics. Seeing the facts is one thing, and evaluating them in a certain way is another. This sort of argument was popular among early and mid-twentieth-century emotivists and prescriptivists. To counter it, Williams

[32] Ibid., p. 149.

notices, we may appeal to the existence of what he calls "thick" as opposed to "thin" ethical concepts, Thin ethical concepts —like right and good and ought —do not appear to be world-guided, in the sense that their application does not appear to be guided by the facts. Pure in their normativity, they are like those little gold stars you can stick on anything. But thick ethical concepts — Williams's own examples are coward, lie, brutality, and gratitude —are world-guided and action-guiding at the same time.[33] Only an action that is motivated in some way by fear can be called cowardly, and yet to call an action cowardly is to suggest that it ought not to be done.[34]

Of course the prescriptivist or emotivist has his own account of these concepts. He thinks that their world-guidedness is one thing and that their action-guidingness is another. The facts tell us which actions are motivated by fear, and when we disapprove of those actions or want to discourage others from doing them, we project our pejorative feelings onto them. So the word "cowardly" is just a pejorative way of describing an act motivated by fear, used when we want to express our feelings or influence our neighbors.

The difficulty with this analysis is that it suggests that it would be possible to use a thick ethical concept with perfect accuracy even if you were completely incapable of appreciating the value it embodies. Williams argues that this is implausible. Of course he does not mean that we can only use evaluative concepts when we ourselves actually endorse the values in question. But we apply such concepts by entering imaginatively into the world of those who have the values, not merely by applying a set of factual criteria.[35] We have to see the world through their eyes, This makes

---

[33] Ibid., pp. 140–41.

[34] Williams says that thick concepts often provide reasons for action (or refraining), but of course, strictly speaking, this is not true of "cowardly." To say that an action is cowardly is to suggest that there is a reason not to do it but not to mention what that reason is. Something in the situation is worth overcoming human fearfulness for, but the term doesn't tell us what. This is because courage is a so-called executive virtue. Williams's other examples are of more directly reason-providing concepts.

[35] Williams, *Ethics and the Limits of Philosophy,* pp. 141–42.

it natural to think of judgments employing thick ethical concepts as perceptual ones. And that in turn makes it natural to think that, like other perceptual judgments, they are a kind of knowledge.

I say that the sky is blue, and my visitor from Jupiter says that it makes a humming noise. Are we agreeing? Certainly we don't *mean* the same thing, since I am talking about how the sky looks and she is talking about how it sounds. Yet when we reflect on these views we find that the things we both say have implications that are expressible in terms of a more absolute concept, that of wavelengths. And when we look at those implications our judgments are found to converge. Here we find grounds for confidence that both of our perceptions are guiding us rightly: they are ways of knowing about the world. Now take this case. The medicine man says that killing the black snake will charm away the evil spirit. And we take "charming away the evil spirit" to have implications expressible in terms of what *we* take to be a more absolute concept, let's say that of curing an illness. And probably we think he is wrong: killing snakes is not a way of curing illnesses.[36]

What would the parallels be in ethics? They might look something like this. The monk says that lying is sinful, and the knight says that it is dishonorable. Certainly they do not *mean* exactly the same thing, for the monk is saying something about the lie's effect on his soul and about how it relates him to his God, while the knight is saying something about the lie's effect on his reputation —on his "character" in the older, more public sense of that word —and how it relates him to his social world. But we take both of their remarks to have implications for what *we* think is a more absolute concept—the lie is wrong and ought not to be told —

---

[36] He *might* be right, of course. There might be some story to tell about placebo effects —perhaps killing the black snake really works because the patient believes it will. Or perhaps the patient knows that if killing the black snake doesn't work the medicine man will try to frighten the evil spirit off by doing something dreadful to the patient, and this prospect frightens the patient into getting well. We don't know enough about medicine to know— and all that matters for the point is that we know roughly how such stories would have to go in order for us to be convinced by them.

and here we find that they converge. And we may think, in this case, that the convergence shows that their concepts are guiding them toward what we take to be a moral truth or that they correctly reflect a moral reality: say, that there are certain kinds of actions that you cannot do without being personally diminished or disfigured, and that this is related to their wrongness.

On the other hand, suppose the knight says that he will be dishonored unless he fights a duel with the man who has insulted him. If we take this to have the implication that trying to kill someone who has hurt your feelings is required, or even all right, we shall have to disagree. But now this is a conclusion that we should be uncomfortable with, and this is precisely because there is a world-guided side to the idea of dishonor. The knight's reputation, his position in his social world, may be damaged in *exactly* the ways that he foresees and has in mind when he says he will be dishonored. What is *for him* his identity may be diminished and disfigured just as it would have been by telling the lie. Facts of this sort should give us pause about whether he is, after all, using the idea of dishonor in a way that has implications for what is morally right or wrong in *our* sense of those words.

Thinking about such cases may lead us to conclude that after all the analogy with the scientific case doesn't hold. We may see the medicine man as trying to cause health, but we should not see the knight as trying to figure out what it is morally right to do. We should not even, according to Williams, assume that we share with the knight any general sense of what it is right or all right to do, about which our views and the knight's both have implications. Instead Williams proposes a different way in which we might look at the ethical beliefs of others:

> On the other model we shall see their judgments as part of their way of living, a cultural artifact they have come to inhabit (although they have not consciously built it). On this, nonobjectivist, model, we shall take a different view of the relations between that practice and critical reflection. We shall

not be disposed to see the level of reflection as implicitly already there, and we shall not want to say that their judgments have, just as they stand, these implications [that is, implications about what it is right or all right to do].[37]

The proposal is that we should see their values not as their best approximations of the truth about value, but rather as a kind of *habitation.* Their values form a part of the structure of the social world in which they live.

But this does not mean that we cannot make any evaluative judgments about their values. We can ask whether their social world — that is, the world that is made of those values — is a good place for human beings to live. This is still, in a broad sense, an ethical question, but our resources for answering it are not tied to any particular system of values. Questions about the suitability of a habitat are answered with reference to the health and flourishing of the creatures who live in it. Williams suggests that a theory of human nature, drawing on the resources of the social as well as the physical sciences, could guide our reflections about what makes for human flourishing. And those reflections in turn could enable us to assess whether a given system of values promoted human flourishing.[38] Williams mentions psychoanalytic theory as one such resource, and of course it is impossible not to think of Freud in this context, with his gloomy view that "the cultural superego . . . does not trouble itself enough about the facts of the mental constitution of human beings." [39] It does seem natural to say that societies in which girls wish passionately that they had been born boys, or in which suicide motivated by feelings of personal worthlessness is common, or in which large segments of the population are sexually dysfunctional are suffering from their values.

Williams proposes that if we did find that a social world promoted the best life or at least a flourishing life for human beings,

---

[37] Williams, *Ethics and the Limits of Philosophy,* p. 147.

[38] Ibid., pp. 45ff, 152–53.

[39] Ibid., p. 45; Freud, *Civilization and Its Discontents,* p. 90.

this would justify the values embodied in that social world. The structure of justification would be very different from the realist structure he thinks we can find in the case of scientific belief. The justification would not be that we find upon reflection that the values are true or that they are reliable guides to the truth about morally right action the way colors are reliable guides to wave-lengths. Williams suggests that the only ethical belief that might survive at the reflective level would be the belief "that a certain kind of life was best for human beings." [40]   The justification of other ethical beliefs would be that it is good for human beings to lead a life that is guided and governed by those beliefs.

So far, in detailing Williams's view, I have been talking, as Williams does, as if from the point of view of an outside observer of an alien society. But when we imagine this same reflective exercise being carried out by a member of the society in question, it becomes clear that the structure of justification here is one of reflective endorsement. Hume, as we saw earlier, reverses the realist ordering of things and argues that vice is bad because we disapprove of it. In a similar way, Williams thinks that ethical value is projected onto the world by our ethical beliefs. Both would deny that it is coherent to ask whether our values are true independently of our own moral or ethical sentiments. The only question left to ask is whether it is good for us to have those senti-ments, and that question must be answered from the perspective of the other practical claims our nature makes on us. Where Hume establishes normativity by showing that morality is congruent with self-interest, Williams asserts that it would have to be established by congruence with human flourishing.

Like Hume, Williams entertains the possibility that this will not be the result. But the prospect is in one way a more alarming one for Hume. Hume believes that he is talking about a set of evaluative concepts that are deeply grounded in human nature and human psychology. He supposes that, if reflection yielded the re-

[40] Williams, *Ethics and the Limits of Philosophy,* p. 154.

sult that morality is bad for the individual, the truth would have to be sunk in "eternal silence and oblivion" in the interests of social order. Williams, by contrast, supposes that different cultures provide us with different sets of values. He sees the reflective test more as a method for choosing among them. When cultures come into what he calls "real confrontation," their members, forced by that confrontation to reflect on the value of their values, may lose confidence in them and come to the conclusion that some other values would lead to a better way of life.[41]   The result will not be that they will decide that their old beliefs were false, or even that after all they did not know what, say, sin or honor was. It will be that they will stop using those concepts altogether.

   In one case, a case of our own, this description of changing values rings true. Consider the uneasy fate of the evaluative concepts "masculine" and "feminine." People who have fallen into doubt about the values embodied in these concepts and the way of life to which they once led us do not argue about whether they track the ethical truth. People who have already decided against these values do not run around telling us that masculinity and femininity are false or wrong. If someone says that aggressiveness is not feminine the response will not be that aggressiveness *is* feminine or that aggressiveness is great. The response is "Let's not talk that way." The complaint that has been launched against these values is not that they were false or misleading but that they were straitjackets, stunting everybody's growth. It is that people who hold themselves and others to these ideals do not flourish. They must therefore be abandoned or revised.

   There is also an element of *reflexivity* in Williams's view. Williams borrows the idea that morality is a projection of human dispositions from Aristotle rather than from Hume. Now Aristotle believed that an ethically good life must be good for the person whose life it is. And Aristotle, again like Hume, has been accused of harboring some form of egoism under this assumption. In de-

---

   [41] Ibid., pp. 160ff.

fending Aristotle against this charge, Williams points out that the Aristotelian agent will reflect on his ethical dispositions from an *ethical* point of view. Or, if he does try to reflect on his ethical dispositions from a point of view outside of those dispositions, from the point of view of his other needs and capacities, the important question will be whether there is any conflict between the demands of those needs and capacities and the demands of his ethical nature?' Aristotle argued that there would not be such a conflict. Again, the conclusion is that our ethical dispositions are judged good from every point of view that makes practical claims on us, including their own point of view. And in this way normativity is established.

## The Reflective Agent

Reflection, Williams tells us, can destroy knowledge.[43] History illustrates the point, for when Bentham reflected on Hume's theory of the virtues, he became a utilitarian.[44] Unfortunately, it looks as if there is a clear route from Hume to Bentham. And it is a route that leads through reflection — in particular, through the reflection of *agents*.

We have seen that in Hume's theory just actions are done from the motive of obligation. Sympathy with the public interest inspires us with a sentiment of disapproval when we think of injustice, and this motivates us to avoid it ourselves. Now let us consider a slightly more attractive version of Hume's sensible knave. Our knave is the lawyer for a rich client who has recently died, leaving his money to medical research. In going through the client's papers the lawyer discovers a will of more recent date,

[42] Ibid., pp. 51–52.

[43] Ibid., p. 148.

[44] This is by Bentham's own report. In a well-known footnote in *A Fragment on Government* (1776), Bentham reports that when he read Humes's *Treatise,* "I felt as if the scales had fallen from my eyes" (p. 50n). What he learned from Hume was "that *utility* was the test and measure of all virtue; . . . and that the obligation to minister to general happiness, was an obligation paramount to and inclusive of every other" (p. 51n).

made without the lawyer's help but in due form, leaving the money instead to the client's worthless nephew, who will spend it all on beer and comic books. The lawyer could easily suppress this new will, and she is tempted to do so. She is also a student of Hume and believes the theory of the virtues that we find in *A Treatise of Human Nature.* So what does she say to herself?

Well, she says to herself that she would disapprove of herself if she did this. She hates unjust actions and the people who perform them. But since the lawyer knows Hume's theory she also knows *why* she would disapprove of herself. She would disapprove of herself because unjust actions have a general tendency to bring down the system of justice. But she also knows that her distaste for such actions is caused by their general tendency, not their actual effects. As Hume has shown, our moral sentiments are influenced by "general rules." And our lawyer knows that this particular unjust action will have no actual effects but good ones. It will not bring down the system of justice, and it will bring much-needed money to medical research.

The lawyer believes that her disapproval of this action depends on the fact that actions of this kind usually have bad effects that this one does not have. It is almost inconceivable that believing this will have no effect on her disapproval itself. Her own feeling of disapproval may seem to her to be, in this case, poorly grounded and therefore in a sense irrational. And this may lead her to set it aside or, if she can't, to resist its motivational force. She may say to herself: since I approve of just actions because they are, generally speaking, useful, why not simply do what will be useful? And then of course she is not a Humean anymore; she is a utilitatarian.[45]

Hume has a defense against this point, but it is a defense of the wrong kind. Consider once more the original sensible knave.

---

[45] There might be arguments of a familiar rule-utilitarian kind against the action she is considering, but if she is moved by those arguments she is still now a utilitarian and not a Humean, at least not in the sense of the *Treatise.*

What does he lose by his knavery? According to Hume, he loses his character with himself, his pleasing sense of self-worth. As I argued earlier, this does not depend on his moral beliefs or on whether he endorses the claims of morality. Since sympathy makes him see himself through the eyes of others, who would disapprove of him for his injustice, it will happen anyway. But that is exactly the problem. If Hume is right, the lawyer may find that she cannot destroy a valid will without intense feelings of humility or self-hatred. These may or may not be strong enough to cause her to desist. But even if they are there will have been normative failure. The lawyer does not believe that the claims her moral feelings make on her in this case are well-grounded. If she could cure herself of them then that is what she would do.

The difficulty in this case is not, strictly speaking, a difficulty with the reflective endorsement strategy. It arises most immediately from something particular to Hume's view: the fact that the moral sentiments are supposed to be influenced by "general rules," rules that do not hold in every case. Such rules cause us to disapprove of certain dispositions or character traits, which are themselves tendencies of a general kind. But that disapproval will be transferred to each and every exercise of the disposition in question only if we forget that the rules that cause it are merely general.

But the difficulty does show us something important about the reflective endorsement method. Consider again the knavish lawyer. She has asked herself whether her feeling of disapproval is really a reason — and now I mean a normative reason — not to do the action, and in this case she has found that it is not. She only disapproves of injustice because it is usually counterproductive. But this act, isolated and secret, will be useful in every way. So now she thinks she has a reason to do it.

Or does she? Why should her reflection stop there? We said that she was a convinced Humean, so she rejects realism. She therefore does not think the fact that an action is useful is in and of itself a reason for doing it — that is, she does not think that

utility is an intrinsically normative consideration. So why should she be moved by utility, any more than by disapproval? Perhaps she now finds that she is *inclined* to be moved by the thought of utility, but that is no more a reason than the fact that she was *inclined* to be moved by disapproval before. She can also ask whether this new inclination is really a reason for action. What is to stop her from continuing to ask that question, from pushing reflection as far as it will go?

If the reflective endorsement of our dispositions is what establishes the normativity of those dispositions, then what we need to establish the normativity of more particular motives and inclinations is the reflective endorsement of those. That after all is the whole point of using the reflective endorsement method to justify morality: we are supposing that, when we reflect on the things that we find ourselves inclined to do, we can then accept or reject the authority those inclinations claim over our conduct and act accordingly.

But what I have just described is exactly the process of thought that, according to Kant, characterizes the deliberations of the autonomous moral agent. According to Kant, as each impulse to action presents itself to us, we should subject it to the test of reflection, to see whether it really is a *reason* to act. Since a reason is supposed to be intrinsically normative, we test a motive to see whether it is a reason by determining whether we should allow it to be a *law* to us. And we do that by asking whether the maxim of acting on it can be willed as a law.

Hume and Williams see the test of reflective endorsement as a philosophical exercise, used to establish the normativity of our moral dispositions and sentiments. But according to Kant, it is not merely that. The test of reflective endorsement is the test used by actual moral agents to establish the normativity of all their particular motives and inclinations. So the reflective endorsement test is not merely a way of justifying morality. *It is morality itself.* In the next lecture, I will elaborate this view.

LECTURE III: THE AUTHORITY OF REFLECTION

*Introduction*

Over the course of the last two lectures I have sketched the way in which the normative question took shape in the debates of modern moral philosophy. Voluntarism tries to explain normativity in what is in some sense the most natural way: we are subject to laws, including the laws of morality, because we are subject to lawgivers. But when we ask why we should be subject to those lawgivers, an infinite regress threatens. Realism tries to block that regress by postulating the existence of entities —objective values, reasons, or obligations —whose intrinsic normativity forbids further questioning. But why should we believe in these entities? In the end, it seems we will be prepared to assert that such entities exist only because —and only if —we are already confident that the claims of morality are justified.

The reflective endorsement theorist tries a new tack. Morality is grounded in human nature. Obligations and values are projections of our own moral sentiments and dispositions. To say that these sentiments and dispositions are justified is not to say that they track the truth, but rather to say that they are good. We are the better for having them, for they perfect our social nature and promote our self-interest.

But the normative question is one that arises in the heat of action. So it is not just our dispositions, but rather the particular motives and impulses that spring from them, that must seem to us to be normative. It is this line of thought that presses us toward Kant. Kant, like the realist, thinks we must show that particular actions are right and particular ends are good. Each impulse as it offers itself to the will must pass a kind of test for normativity before we can adopt it as a reason for action. But the test that it must pass is not the test of knowledge or truth. For Kant, like Hume and Williams, thinks that morality is grounded in human nature and that moral properties are projections of human dispositions. So the test is one of reflective endorsement.

In what follows I will lay out the elements of a theory of normativity. This theory derives its main inspiration from Kant, but with some modifications that I have come to think are needed. What I say will necessarily be sketchy, and sketchily argued. My attention here will be focused on four points: first, that autonomy is the source of obligation, and in particular of our ability to obligate ourselves; second, that we have *moral* obligations, by which I mean obligations to humanity as such; third, that since we can obligate ourselves, we can also be obligated by other people; and fourth, that we have obligations to other living things. I will have little to say about the content of any of these obligations. And it will be no part of my argument to suggest either that all obligations are moral or that obligations can never conflict. My aim is to show you where obligation comes from. Exactly which obligations we have and how to negotiate among them is a topic for another day.

## *The Problem*

The human mind is self-conscious. Some philosophers have supposed that this means that our minds are internally luminous, that their contents are completely accessible to us, that we always can be certain what we are thinking and feeling and wanting, and so that introspection yields certain knowledge of the self. Like Kant, and many philosophers nowadays, I do not think that this is true. Our knowledge of our own mental states and activities is no more certain than anything else.

But the human mind *is* self-conscious in the sense that it is essentially reflective. I'm not talking about being *thoughtful,* which of course is an individual property, but about the structure of our minds that makes thoughtfulness possible. A lower animal's attention is fixed on the world. Its perceptions are its beliefs and its desires are its will. It is engaged in conscious activities, but it is not conscious *of* them. That is, they are not the objects of its attention. But we human animals turn our attention on to our per-

ceptions and desires themselves, and we are conscious *of* them. That is why we can think *about* them.

And this sets us a problem no other animal has. It is the problem of the normative. For our capacity to turn our attention onto our own mental activities is also a capacity to distance ourselves from them and to call them into question. I perceive, and I find myself with a powerful impulse to believe. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn't dominate me and now I have a problem. Shall I believe? Is this perception really a *reason* to believe? I desire and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a certain distance. Now the impulse doesn't dominate me and now I have a problem. Shall I act? Is this desire really a *reason* to act? The reflective mind cannot settle for perception and desire, not just as such. It needs a *reason*. Otherwise, at least as long as it reflects, it cannot commit itself or go forward.

If the problem springs from reflection then the solution must do so as well. If the problem is that our perceptions and desires might not withstand reflective scrutiny, then the solution is that they might. We need reasons because our impulses must be able to withstand reflective scrutiny. We have reasons if they do. The normative word "reason" refers to a kind of reflective success. If "good" and "right" are also taken to be intrinsically normative words then they too must refer to reflective success. And they do. Think of what they mean when we use them as *exclamations:* "Good!" "Right!" There they mean: I'm satisfied, I'm happy, I'm committed, you've convinced me, let's go. They mean the work of reflection is done.

"Reason" then means reflective success. So if I decide that my desire is a reason to act, I must decide that on reflection I endorse that desire. And here we find the problem. For how do I decide that? Is the claim that I look at the desire and see that it is intrinsically normative or that its object is? Then all of the argu-

ments against realism await us. Does the desire or its object in-
herit its normativity from something else? Then we must ask what
makes that other thing normative, what makes it the source of a
reason. And now of course the usual regress threatens. So what
brings reflection to an end?

Kant described this same problem in terms of freedom. It is
because of the reflective structure of the mind that we must act,
as he puts it, under the idea of freedom. He says, "We cannot
conceive of a reason which consciously responds to a bidding from
the outside with respect to its judgments."[1] If the bidding from
outside is desire, then his point is that the reflective mind must
endorse the desire before it can act on it — it must say to itself
that the desire is a reason. We must, as he puts it, *make it our
maxim* to act on the desire. And this is something we must do of
our own free will.

Kant defines a free will as a rational causality that is effective
without being determined by any alien cause. Anything outside
of the will counts as an alien cause, including the desires and
inclinations of the person. The free will must be entirely self-
determining. Yet, because the will is a causality, it must act ac-
cording to some law or other. Kant says, "Since the concept of a
causality entails that of laws . . . it follows that freedom is by no
means lawless . . ."[2] Alternatively, we may say that since the will
is practical reason, it cannot be conceived as acting and choosing
for no reason. Since reasons are derived from principles, the free
will must have a principle. But because the will is free, no law
or principle can be imposed on it from outside. Kant concludes
that the will must be autonomous: that is, it must have its *own*
law or principle. And here again we arrive at the problem. For
where is this law to come from? If it is imposed on the will from
outside then the will is not free. So the will must adopt the law

[1] Kant, *Foundations of the Metaphysics of Morals*, p. 448; in Beck's translation,
p. 66.

[2] Ibid., p. 446; in Beck's translation, p. 65.

for itself. But until the will has a law or principle, there is nothing from which it can derive a reason. So how can it have any reason for adopting one law rather than another?

Well, here is Kant's answer. The Categorical imperative tells us to act only on a maxim that we could will to be a law. And *this,* according to Kant, *is* the law of a free will. To see why, we need only compare the problem faced by the free will with the content of the Categorical imperative. The problem faced by the free will is this: the will must have a law, but because the will is free, it must be its own law. And nothing determines what that law must be. *All that it has to be is a law.* Now consider the content of the Categorical imperative. The Categorical imperative simply tells us to choose a law. Its only constraint on our choice is that it have the form of a law. And nothing determines what that law must be. *All that it has to be is a law.*

Therefore the categorical imperative is the law of a free will. It does not impose any external constraint on the free will's activities, but simply arises from the nature of the will. It describes what a free will must do in order to be what it is. It must choose a maxim it can regard as a law.[3]

Now I'm going to make a distinction that Kant doesn't make. I am going to call the law of acting only on maxims you can will to be laws "the Categorical imperative." And I am going to distinguish it from what I will call "the moral law." The moral law, in the Kantian system, is the law of what Kant calls the Kingdom of Ends, the republic of all rational beings. The moral law tells us to act only on maxims that all rational beings could agree to act on together in a workable cooperative system. Now the Kantian argument that I have just described establishes that *the categorical imperative* is the law of a free will. But it does not establish that *the moral law* is the law of a free will. Any law is universal, but

---

[3]This is a reading of the argument Kant gives in ibid., pp. 446–48; in Beck's translation, pp. 64–67; and in *The Critique of Practical Reason* under the heading "Problem II," p. 29; in Becks translation, pp. 28–29. It is explained in greater detail in my "Morality as Freedom."

the argument doesn't settle the question of the *domain* over which the law of the free will must range. And there are various possibilities here. If the law is the law of acting on the desire of the moment, then the agent will treat each desire as it arises as a reason, and her conduct will be that of a wanton.[4] If the law ranges over the interests of an agent's whole life, then the agent will be some sort of egoist. It is only if the law ranges over every rational being that the resulting law will be the moral law, the law of the Kingdom of Ends.

Because of this, it has sometimes been claimed that the categorical imperative is an empty formalism. And this in turn has been conflated with another claim, that the moral law is an empty formalism. Now that second claim is false.[5] But it is true that the argument that shows that we are bound by the categorical imperative does not show that we are bound by the moral law. For that we need another step. The agent must think of *herself* as a Citizen of the Kingdom of Ends.

## The Solution

Those who think that the human mind is internally luminous and transparent to itself think that the term "self-consciousness" is appropriate because what we get in human consciousness is a direct encounter with the self. Those who think that the human mind

---

[4] I have a reason for saying that her behavior will be that of a wanton rather than simply saying that she will be a wanton. Harry Frankfurt, from whom I am borrowing the term, defines a wanton as someone who has no second-order volitions. An animal, whose desire is its will, is a wanton. I am arguing here that a person cannot be like that, because of the reflective structure of human consciousness. A person must act on a reason, and so the person who acts like a wanton must be treating the desire of the moment as a reason. That commits her to the principle that the desire of the moment is a reason, and her commitment to that principle counts as a second-order volition. See Frankfurt, "Freedom of the Will and the Concept of a Person," especially the discussion on pp. 16–19, The affinity of my account with Frankfurt's will be evident.

[5] Bradley and others understood Hegel's famous objection this way, and if it is taken this way it is a mistake. I argue for this in my paper "Kant's Formula of Universal Law." In that paper, however, I do not distinguish the categorical imperative from the moral law, and my arguments there actually only show that the moral law has content.

has a reflective structure use the term too, but for a different rea-
son. The reflective structure of the mind is a source of "self-
consciousness" because it forces us to have a *conception* of our-
selves. As Kant argues, this is a fact about what it is *like* to be
reflectively conscious and it does not prove the existence of a meta-
physical self. From a third person point of view, outside of the
deliberative standpoint, it may look as if what happens when
someone makes a choice is that the strongest of his conflicting
desires wins. But that isn't the way it is *for you* when you deliber-
ate. When you deliberate, it is as if there were something over
and above all of your desires, something that is *you,* and that
*chooses* which desire to act on. This means that the principle or
law by which you determine your actions is one that you regard as
being expressive of *yourself.* To identify with such a principle or
law is to be, in St. Paul's famous phrase, a law to yourself.[6]

An agent might think of herself as a Citizen in the Kingdom
of Ends. Or she might think of herself as a member of a family
or an ethnic group or a nation. She might think of herself as the
steward of her own interests, and then she will be an egoist. Or
she might think of herself as the slave of her passions, and then
she will be a wanton. And how she thinks of herself will deter-
mine whether it is the law of the Kingdom of Ends, or the law of
some smaller group, or the law of the egoist, or the law of the
wanton that is the law that she is to herself.

The conception of one's identity in question here is not a theo-
retical one, a view about what as a matter of inescapable scientific
fact you are. It is better understood as a description under which
you value yourself, a description under which you find your life to
be worth living and your actions to be worth undertaking. So I
will call this a conception of your practical identity. Practical
identity is a complex matter and for the average person there will
be a jumble of such conceptions. You are a human being, a woman
or a man, an adherent of a certain religion, a member of an ethnic

6 Romans II:14.

group, someone's friend, and so on. And all of these identities give rise to reasons and obligations. Your reasons express your identity, your nature; your obligations spring from what that identity forbids.

Our ordinary ways of talking about obligation reflect this connection to identity. A century ago a European could admonish another to civilized behavior by telling him to act like a Christian. It is still true in many quarters that courage is urged on males by the injunction "Be a man!" Duties more obviously connected with social roles are of course enforced in this way. "A psychiatrist doesn't violate the confidence of her patients." No "ought" is needed here because the normativity is built right into the role. But it isn't only in the case of social roles that the idea of obligation invokes the conception of practical identity. Consider the astonishing but familiar "I couldn't live with myself if I did that." Clearly there are *two* selves here, me and the one I must live with and so must not fail. Or consider the protest against obligation ignored : "Just who do you think you are ?"

The connection is also present in the concept of integrity. Etymologically, integrity is oneness, integration is what makes something one. To be a thing, one thing, a unity, an entity; to be anything at all: in the metaphysical sense, that is what it means to have integrity. But we use the term for someone who lives up to his own standards. And that is because we think that living up to them is what makes him one, and so what makes him a person at all.

It is the conceptions of ourselves that are most important to us that give rise to unconditional obligations. For to violate them is to lose your integrity and so your identity, and no longer to be who you are. That is, it is no longer to be able to think of yourself under the description under which you value yourself and find your life worth living and your actions worth undertaking. That is to be for all practical purposes dead or worse than dead. When an action cannot be performed without loss of some fundamental

part of one's identity, and an agent would rather be dead, then the obligation not to do it is unconditional and complete. If reasons arise from reflective endorsement, then obligation arises from reflective *rejection.*

But the question how exactly an agent *should* conceive her practical identity, the question which law she should be to herself, is not settled by the arguments I have given. So moral obligation is not yet on the table. To that extent the argument is formal, and in one sense empty.

But in another sense it is not empty at all. What we have established is this. The reflective structure of human consciousness requires that you identify yourself with some law or principle that will govern your choices. It requires you to be a law to yourself. And that is the source of normativity. So the argument shows just what Kant said that it did: that our autonomy is the source of obligation.

It will help to put the point in Joseph Butler's terms, in terms of the distinction between power and authority. We do not always do what upon reflection we would do or even what upon reflection we have already decided to do. Reflection does not have irresistible power over us. But when we do reflect we cannot but think that we ought to do what on reflection we conclude we have reason to do. And when we don't do that we punish ourselves, by guilt and regret and repentance and remorse. We might say that the acting self concedes to the thinking self its right to government. And the thinking self, in turn, tries to govern as well as it can. So the reflective structure of human consciousness establishes a relation here, a relation that we have to ourselves. And it is a relation not of mere power but rather of *authority.* And *that* is the authority that is the source of obligation.

Notice that this means that voluntarism is true after all. The source of obligation is a legislator, one whose authority is beyond question and does not need to be established. But there is only one such authority and it is the authority of your own mind and

will.[7]  So Pufendorf and Hobbes were right.  It is not the bare fact
that it would be a good idea to perform a certain action that obli-
gates us to perform it.  It is the fact that we *command ourselves*
to do what we find it would be a good idea to do.

One more step is necessary.  The acting self concedes to the
thinking self its right to govern.  But the thinking self in turn must
try to govern well.  It is its job to make what is in any case a good
idea into law.  How do we know what is a good idea or what
should be a law?  Kant proposes that we can tell whether our
maxims should be laws by attending not to their matter but to
their form.

To understand this idea, we need to return to its origins, which
are in Aristotle.  According to Aristotle, a thing is composed of a
form and a matter.  The matter is the material, the parts, from
which it is made.  The form of a thing is its functional arrange-
ment.  That is, it is the arrangement of the matter or of the parts
that enables the thing to serve its purpose, to do whatever it does.
For example, the purpose of a house is to be a shelter, so the form
of a house is the way the arrangement of the parts — the walls
and the roof — enables it to serve as a shelter.  "Join the walls at
the corner, put the roof on top, and that's how we keep the weather
out."  That is the form of a house.[8]

Next consider the maxim of an action.  Since every human
action is done for an end, a maxim has two parts, the act and the
end. The form of the maxim is the arrangement of its parts. Take,
for instance, Plato's famous example of the three   maxims.[9]

1. I will keep my weapon, because I want it for myself.

---

[7] This remark needs a qualification, which springs from the fact that we can
unite our wills with the wills of others.  In Kant's theory, this happens when we
are citizens who together form a general will or when we make friends or get mar-
ried.  In those cases it is sometimes the united will that has authority over our con-
duct.  For further discussion, see my "Creating the Kingdom of Ends: Reciprocity
and Responsibility in Personal Relations."

[8] These views are found throughout Aristotle's writings, but centrally discussed
in books VII–IX of the *Metaphysics and in On the Soul.*

[9] Plato, *Republic,* I, 331c., p. 580.

2. I will refuse to return your weapon, because I want it for myself.

3. I will refuse to return your weapon, because you have gone mad and may hurt someone.

Maxims 1 and 3 are good; maxim 2 is bad. What makes them so? Not the actions, for maxims 2 and 3 have the same actions; not the purposes, for maxims 1 and 2 have the same purposes. The goodness does not rest in the parts; but rather in the way the parts are combined and related; so the goodness does not rest in the matter, but rather in the form of the maxim. But form is not merely the arrangement of the parts; it is the *functional* arrangement — the arrangement that enables the thing to do what it does. If the walls are joined and roof placed on top *so that* the building can keep the weather out, then the building has the form of a house. So: if the action and the purpose are related to one another *so that* the maxim can be willed as a law, then the maxim is good.

Notice what this establishes. A good maxim is good in virtue of its internal structure. Its internal structure, its form, makes it fit to be willed as a law. A good maxim is therefore an *intrinsically normative entity.* So realism is true after all, and Nagel, in particular, was right. When an impulse presents itself to us, as a kind of candidate for being a reason, we look to see whether it really is a reason, whether its claim to normativity is true.

But this isn't an exercise of intuition or a discovery about what is out there in the world. The test for determining whether an impulse is a reason is whether *we* can will the maxim of acting on that impulse as law. So the test is a test of endorsement.

This completes the first part of my argument, so let me sum up what I've said. What I have shown so far is why there is such a thing as obligation. The reflective structure of human consciousness forces us to act for reasons. At the same time, and relatedly, it forces us to have a conception of our own identity, a conception that identifies us with the source of our reasons. In this way, it makes us laws to ourselves. When an impulse presents itself to us

we ask whether it could be a reason. We answer that question by seeing whether the maxim of acting on it can be willed as a law by a being with the identity in question. If it can be willed as a law, it is a reason, for it has an intrinsically normative structure. If it cannot be willed as a law, we must reject it, and in that case we get obligation.

A moment ago I said that realism is true after all. But that could be misleading. That we obligate ourselves is simply a fact about human nature. But whether a maxim can serve as a law still depends upon the way that we think of our identities. So there is still an element of relativism in the system. In order to establish that there are *moral* obligations we will need another step.

## *Moral Obligation*

There is another way to make the points I have been making, and in approaching the problem of relativism it will be helpful to employ it. We can take as our model the way Rawls employs the concept/conception distinction in *A Theory of Justice.* There, the *concept* of justice refers to a problem, the problem of how the benefits of social cooperation are to be distributed. A *conception* of justice is a principle that is proposed as a solution to that problem.[10]

In the same way, the most general normative concepts, the right and the good, are names for problems — for the normative problems that spring from our reflective nature. "Good" names the problem of what we are to strive for, aim for, and care about in our lives. "Right" names the more specific problem of what we are to do. The "thinness" of these terms, to use Bernard Williams's language, comes from the fact that they are only concepts, names for whatever it is that solves the problems in question.

How do we get from concepts to conceptions? What mediates is a conception of practical identity. In Rawls's argument, we move from concept to conception by taking up the standpoint of

[10] Rawls, *A Theory of Justice,* p. 5.

the pure citizen and asking what principles such a citizen would have reason to adopt. In Kant's argument, we move from concept to conception by taking up the standpoint of a Citizen in the Kingdom of Ends and asking what principles that citizen would have reason to adopt.

Because they are normative, thick ethical concepts stand to thin ones as conceptions to concepts. They represent solutions, or at least reasons that will be weighed in arriving at solutions, to the problems that are set by reflection. And that means that they embody a view about what is right or good. If this is right, then Williams is wrong to say that reflection is not inherent in, or already implied by, thick ethical concepts.[11] As normative concepts, they are essentially reflective.

Furthermore, our thin ethical concepts, although not necessarily our thick ones, will be shared with those alien scientific investigators.[12] For the fact that they are scientific investigators means that they have asked themselves what they ought to believe and that they have decided that the question is worth pursuing. And that in turn means that they are rational and social beings, who face normative problems like our own and sometimes solve them. The exact shape of their problems may be different from ours, and so they may have different conceptions. But if we can see their conceptions as solutions to the normative problems that *they* face, there will even be a kind of convergence.

But this does not eliminate the element of relativism that Williams has sought to preserve. The mediation between concepts and conceptions comes by way of practical identity. And human identity has been differently constituted in different social worlds. Sin, dishonor, and moral wrongness all represent conceptions of what one cannot do without being diminished or disfigured, without loss of identity, and therefore conceptions of what one must not do. But they belong to different worlds in which human beings thought

[11] See "Bernard Williams" in Lecture II.
[12] See "Bernard Williams" in Lecture II.

of themselves and of what made them themselves in very different ways. Where sin is the conception, my identity is my soul and it exists in the eyes of my God. Where dishonor is the conception, my identity is my reputation, my position in some small and knowable social world. The conception of *moral* wrongness as we now understand it belongs to the world *we* live in, the one brought about by the Enlightenment, where one's identity is one's relation to humanity itself. Hume said at the height of the Enlightenment that to be virtuous is to think of yourself as a member of the "party of humankind, against vice or disorder, its common enemy." [13] And that is now true. But we coherently can grant that it was not always so.

But this is not to say that there is nothing to be said in favor of the Enlightenment conception. This sort of relativism has its limits, and they come from two different but related lines of thought.

We have already seen one of them set forward by Bernard Williams. We could, with the resources of a knowledge of human nature, rank different sets of values according to their tendency to promote human flourishing. If values are associated with ways of thinking of what we most fundamentally are, then the point will be that some ways of conceiving one's identity are healthier and better for us than others.

But it is also important to remember that no argument can preserve any form of relativism without on another level eradicating it. This is one of the main faults with one well-known criticism of liberalism, that the conception of the person that is employed in its arguments is an "empty self."[14] It is urged by communitarians that people need to conceive themselves as members of smaller communities, essentially tied to particular others and traditions. This is an argument about how human beings need to constitute our practical identities, and if it is successful what it

---

[13] Hume, *Enquiry concerning the Principles of Morals,* p. 275.
[14] See, for instance, Michael Sandel, *Liberalism and the Limits of Justice.*

establishes is a *universal* fact, namely that our practical identities must be constituted in part by particular ties and commitments. And the communitarian who has reflected and reached this conclusion now has a conception of his own identity that is universal: he is an animal that needs to live in community.

And there is a further implication of this that is important. Once the communitarian sees himself this way, his particular ties and commitments will remain normative for him only if this more fundamental conception of his identity is one that he can see as normative as well. A further stretch of reflection requires a further stretch of endorsement. So he must endorse this new view of his identity. He is an animal that needs to live in community, and he now takes *this* to be a normative identity. He treats it as a source of reasons, for he argues that it matters that he gets what he needs. And this further stretch of endorsement is exactly what occurs. Someone who is moved to urge the value of *having* particular ties and commitments has discovered that part of their normativity comes from the fact that human beings need to have them. He urges that our lives are meaningless without them. That is not a reason that *springs from* one of his own particular ties and commitments. It is a plea on behalf of all human beings. And that means that he is no longer immersed in a normative world of particular ties and commitments. Philosophical reflection does not leave everything just where it was.

This is just a fancy new model of an argument that first appeared in a much simpler form, Kant's argument for his Formula of Humanity. The form of relativism with which Kant began was the most elementary one we encounter —the relativity of value to human desires and interests. He started from the fact that when we make a choice we must regard its object as good. His point is the one I have been making —that being human we must endorse our impulses before we can act on them. Kant asked what it is that makes these objects good, and, rejecting one form of realism, he decided that the goodness was not in the objects themselves.

Were it not for our desires and inclinations, we would not find
their objects good. Kant saw that we take things to be important
because they are important to us — and he concluded that we must
therefore take ourselves to be important. In this way, the value of
humanity itself is implicit in every human choice.[15] If normative
skepticism is to be avoided — if there is any such thing as a reason
for action — then humanity as the source of all reasons and values
must be valued for its own sake.[16]

The point I want to make now is the same. In this lecture I
have offered an account of the source of normativity. I have argued
that a human being is an animal who needs a practical conception
of her own identity, a conception of who she is that is normative
for her. Otherwise she could have no reasons to act, and since she
is reflective she needs reasons to act. But you are a human being
and so if you believe my argument you can now see that *this* is
your identity. You are an animal of the sort I have just described.
And that is not merely a contingent conception of your identity,
which you have constructed or chosen for yourself or could con-
ceivably reject. It is simply the truth. Now that you see that your
need to have a normative conception of yourself comes from the
sort of animal you are, you can ask whether it really matters
whether animals of this kind conform to their normative practical
identities. Does it really matter what human beings do? And here
you have no option but to say yes. Since you are human you *must*
take something to be normative, that is, some conception of practi-
cal identity must be normative for you. If you had no normative
conception of your identity, you could have no reasons for action,

---

[15] Kant, *Foundations of the Metaphysics of Morals,* pp. 427–28; in Beck's trans-
lation, pp. 45–47. I am here summarizing the interpretation of this argument I give
in "Kant's Formula of Humanity."

[16] This implies that you must accept the laws that arise from this more funda-
mental view of your identity, the laws of morality. But it does not imply that the
less fundamental laws no longer exist or that the more fundamental ones always
trump them. The view I have as I have spelled it out so far leaves room for con-
flict. Some account of how such conflicts might be negotiated is desirable, but I do
not mean to be giving or implying any such account here.

and because your consciousness is reflective, you could then not act at all. Since you cannot act without reasons and your humanity is the source of your reasons, you must endorse your own humanity if you are to act at all.

It follows from this argument that human beings are valuable. Enlightenment morality is true.

### Obligating One Another

So far I have argued that the reflective structure of human consciousness gives us legislative authority over ourselves. That is why we are able to obligate ourselves. And just now I argued that once we understand how all of this works, we must concede that our humanity is an end in itself, that human nature as the source of our values is itself a value. This, I should add, is what gives rise to *moral* obligation.

You might suppose that I am claiming that this settles the question of our obligations to others. Since I regard my humanity as a source of value, I must in the name of consistency regard your humanity that way as well. So I must value the things that you value. Or, to put it another way, since I think my humanity is what makes my desires into normative reasons, I must suppose that the humanity of others makes their desires into normative reasons as well.

This is a familiar form of argument. Versions of it appear in Thomas Nagel's book *The Possibility of Altruism,* and in Alan Gewirth's book *Reason and Morality.* And the criticism of this form of argument is always the same. Consistency can force me to grant that your humanity is normative for you just as mine is normative for me. It can force me to acknowledge that your desires have the status of reasons for you, in exactly the same way that mine do for me. But it does not force me to share in your reasons or make *your* humanity normative for me.[17] It could still be true

---

[17] See for instance Williams's criticism of Gewirth in chapter 4 of *Ethics and the Limits of Philosophy.*

that I have my reasons and you have yours, and indeed that they leave us eternally at odds.[18]   Human beings might be egoistic, not in the sense of being concerned only about themselves, but in the sense defined by Nagel in *The Possibility of Altruism*. The egoist thinks that reasons are a kind of private property. We each act on our own private reasons, and we need some special reason, like friendship or contract, for taking the reasons of others into account.

In one sense this objection is correct. Consistency is not what forces us to share our reasons. And even if these arguments did work, they would work in the wrong way. They would show that I have an obligation *to myself* to treat you in ways that respect the value that I place on you. But they would not show that I have obligations *to you*.   So we need something more.

As we have seen, I can obligate myself because I am conscious of myself. So if you are going to obligate me I must be conscious of you. You must be able to intrude on my reflections — you must be able to get under my skin. People suppose that practical reasons are private because they suppose that reflection is a private activity. And they suppose that, in turn, because they believe in the privacy of consciousness. So what we need at this point is some help from Wittgenstein.

Consider the private language argument. As Wittgenstein defines it, a private language would be a language that referred to something essentially private and incommunicable, say for instance a sensation that is yours alone, and cannot be described in any other way than by a name that you give to it. You can't even call it a tickle or an itch, for then it would be communicable. So you just call it 'S.' And whenever you experience it, you say to yourself, "That was S.'"[19]

Wittgenstein argues that there couldn't be any such language. One way to understand his argument goes like this: Meaning is

---

[18] In contemporary jargon, the objection is that the reasons the argument reveals are "agent-relative'' rather than "agent-neutral.''

[19] See Wittgenstein, *Philosophical Investigations,*   §§243ff., pp. 88ff.

relational because it is a *normative* notion: to say that X means Y is to say that one ought to take X for Y ; and this requires two, a legislator to lay it down that one ought to take X for Y and a citizen to obey. And the relation between these two is not merely causal because the citizen can disobey: there must be a possibility of misunderstanding or mistake. Since it is a relation in which one gives a law to another, it takes two to make a meaning. So you cannot peer inwardly at an essentially private and incommunicable sensation and say, "That is what I mean by S" and so in that way mean something. For if that is what you mean by S, then when you call something S it must be *that,* and if you call something else S you must be wrong. But if what you call S is just that sensation that makes you feel like saying "S," and it cannot be identified in any other way, then you cannot be wrong." The idea of a private language is inconsistent with the normativity of meaning.

If we read Wittgenstein that way, there is an obvious similarity between the kind of normativity that he thinks characterizes language and the kind of normativity that I have been attributing to practical reasons. We could make a parallel argument against private reasons: Reasons are relational because reason is a normative notion: to say that R is a reason for A is to say that one should do A because of R; and this requires two, a legislator to lay it down and a citizen to obey. And the relation between them is not just causal because the citizen can disobey: there must be a possibility of irrationality or wrongdoing. Since it is a relation in which one gives a law to another, it takes two to make a reason. And here the two are the two elements of reflective consciousness, the thinking self and the active self: what I have been talking about all along is how you can make laws and reasons for your self.[21]

[20] See especially ibid., §258, p. 92: "But 'I impress it on myself' can only mean: this process brings it about that I remember the connection *right* in the future, But in the present case I have no criterion of correctness. One would like to say: whatever is going to seem right to me is right. And that only means that here we cannot talk about 'right.'"

[21] It may look as if there is a disanalogy here. The private language argument shows that you cannot mean a certain sensation by 'S' just now and never again,

There are two important points here. The first point is that the mistake involved in thinking that a meaning is a mental entity is exactly like that involved in thinking that a reason or a value is a mental entity. To talk about reasons and meanings is not to talk about entities, but to talk in a shorthand way about relations we have with ourselves and one another. The normative demands of meaning and reason are not demands that are made on us by objects, but are demands that we make on ourselves and each other.

The second point concerns privacy. The private language argument does not show that I could not have my own personal language. It shows that I could not have a language that is in principle incommunicable to anybody else. When I make a language, I make its meanings normative for me. As Wittgenstein puts it, I *undertake* to use words in certain ways." And however I go about binding myself to those meanings, it must be possible for me to bind another in exactly the same way.

If I say to you, "Picture a yellow spot!" you will. What exactly is happening? Are you simply cooperating with me? No, because at least without a certain active resistance you will not be able to help it. Is it a causal connection then? No, or at least not merely that, for if you picture a *pink* spot you will be mistaken, wrong. Causal connections cannot be wrong. What kind of necessity is this, both normative and compulsive? It is *obligation*.

Philosophers have been concerned for a long time about how we understand the meanings of words, but we have not paid enough attention to the fact that it is so hard not to. It is nearly impossible to hear the words of a language you know as mere noise. And this has implications for the supposed privacy of human consciousness. For it means that I can always intrude myself

---

because then you could not be wrong. The remark I just made makes it look as if you could have a reason just now and never again — the thinking self could bind the acting self to act a certain way just now. Actually, however, I do not think that is a possibility, since the acting self cannot coherently be taken to exist just at a particular moment. See my "Personal Identity and the Unity of Agency: A Kantian Response to Parfit," pp. 113–14.

    [22] Wittgenstein, Philosophical Investigations, §262, p. 93.

into your consciousness. All I have to do is talk to you in the words of a language you know, and in this way I can force you to think. The space of linguistic consciousness is essentially public, like a town square. You might happen to be alone in yours, but I can get in anytime. Wittgenstein says, "Think in this connection how singular is the use of a person's name to *call* him." [23]

If I call out your name, I make you stop in your tracks. (If you love me, I make you come running.) Now you cannot proceed as you did before. Oh, you can proceed, all right, but not just as you did before. For now if you walk on, you will be ignoring me and slighting me. It will probably be difficult for you, and you will have to muster a certain active resistance, a sense of rebellion. But why should you have to rebel against me? It is because I am a law to you. By calling out your name, I have obligated you. I have given you a reason to stop.[24]

Of course you might not stop. You have reasons of your own, and you might decide, rightly or wrongly, that they outweigh the one I have given you. But that I have given you a reason is clear from the fact that, in ordinary circumstances, you will feel like giving me one back. "Sorry, I must run, I'm late for an appointment." We all know that reasons must be met with reasons, and that is why we are always exchanging them.

We do not seem to need a reason to take the reasons of others into account. We seem to need a reason not to. Certainly we do things because others want us to, ask us to, tell us to, all the time. We give each other the time and directions, open doors and step aside, warn each other of imminent perils large and small. We respond with the alacrity of obedient soldiers to telephones and doorbells and cries for help. You could say that it is because we want to be cooperative, but that is like saying that you understand

---

[23] Ibid., §27, p. 13.

[24] More strictly speaking, the needs and demands of others present us with what Kant calls "incentives," just as our own inclinations do. Incentives come up for automatic consideration as candidates for being reasons. I thank Ulrike Heuer for prompting me to be clearer on this point.

my words because you want to be cooperative. It ignores the same essential point, which is that it is so hard not to.

Now the egoist may reply that this does not establish that other people's reasons are reasons for me. He'll say that I am merely describing a deep psychological fact — that human beings are very susceptible to one another's pressure. We tend to cave in to the demands of others. But nothing I have said so far shows that we have to treat the demands of others as *reasons.* It is at this point that Thomas Nagel's argument, from *The Possibility of Altruism,* comes into its own.

Suppose that we are strangers and that you are tormenting me, and suppose that I call upon you to *stop.* I say, "How would you like it if someone did that to you?" Now you cannot proceed as you did before. Oh, you can proceed all right, but not just as you did before. For I have obligated you to stop.

How does the obligation come about? Just the way that Nagel says that it does. I invite you to consider how you would like it if someone did that to you. You realize that you would not merely dislike it, you would resent it. You would think that the other has a reason to stop — more, that he has an obligation to stop. And that obligation would spring from your own objection to what he does to you. You make yourself an end for others; you make yourself a law to them. But if you are a law to others insofar as you are just a person, just *someone,* then others are also laws to you.[25] By making you think these thoughts, I force you to acknowledge the value of *my* humanity, and I obligate you to act in a way that respects it.

As Nagel observes, the argument does not go through if you fail to see yourself, to identify yourself, as just someone, a person, one person among others who are equally real.[26] The argument invites you to change places with the other, and you cannot do this if you fail to see what you and the other have in common. Suppose you could say, "Someone doing that to *me,* why that would be

[25] See Nagel, *The Possibility of Altruism,* pp. 82–84.

[26] Ibid., chapter 9.

terrible! But then I am *me,* after all." Then the argument would fail of its effect; it would not find a foothold in you. But the argument never really fails in *that* way.

For it to fail in that way, I would have to hear your words as mere noise, not as intelligible speech. And it is impossible to hear the words of a language you know as mere noise. In hearing your words as *words,* I acknowledge that you are *someone.* In acknowledging that I can hear them, I acknowledge that I am *someone.* If I listen to the argument at all, I have already admitted that each of us is *someone.*

Consider an exchange of reasons. A student comes to your office door and says, "I need to talk to you. Are you free now?" You say, "No, I've got to finish this letter right now and then I've got to go home. Could you possibly come around tomorrow, say about three?" And your student says, "Yes, that will be fine. I'll see you tomorrow at three then."

What is happening here? On my view, the two of you are reasoning together, to arrive at a decision, a single shared decision, about what to do. And I take that to be the natural view. But if egoism is true, and reasons cannot be shared, then that is not what is happening. Instead, each of you backs into the privacy of his practical consciousness, reviews his own reasons, comes up with a decision, and then reemerges to announce the result to the other. And the process stops when the results happen to coincide, and the agents know it, because of the announcements they have made to each other.

Now consider an exchange of ideas, rather than an exchange of practical reasons. Here we do not find these two possibilities. If meanings could not be shared, there would be no point in announcing the results of one's private thinking to anybody else. If they can be shared, then it is in principle possible to think the issues through together, and that is what people do when they talk. But if we have to grant that meanings can be shared, why not grant that practical reasons can be shared too?

The egoist may reply that I am leaving out an option. The student/teacher relation is a personal one. People who enter into particular personal relationships have special reasons to take each other's reasons into account. So the exchange I've just described takes place against a background agreement that the parties involved will take each other's reasons into account. The egoist is someone who only acts on his own reasons, not someone who has no concern for others. So you and your student reason together because you have tacitly agreed to, but this does not show that this is what usually happens.

But the objection reemerges within this framework. How are we to understand this personal relationship? If reasons are still private then it goes like this: each of you has a private reason to take the reasons of the other into account. A personal relationship is an interest in one another's interests.[25] This doesn't change the shape of the deliberation —you still back into your private deliberative spaces and then reemerge to announce the results. This only shows why you think there's a point in the exercise at all, why you hope to reach a convergence. But if you are really reasoning together, if you have joined your wills to arrive at a single decision — well, then that can happen, can't it? And why shouldn't it be what usually happens? Why shouldn't language force us to reason practically together, in just the same way that it forces us to think together?

I believe that the myth of egoism will die with the myth of the privacy of consciousness. Now you may object that the way in which I have argued against the privacy of consciousness —by arguing that we can think and reason together —has nothing to do with what philosophers mean when they discuss that privacy. What they mean by privacy is that you don't always know what someone else is thinking or feeling. The way in which you have access to the contents of another person's mind —through words and expressions and other such forms of evidence —doesn't allow

[27] And that's not what a personal relationship is. See note 7 of this lecture.

you to look around in it freely, and make sure that you know what's there and what's not.

But that's not an issue about privacy. If you accept the thesis that consciousness is reflective rather than internally luminous, then you must admit that you don't have access to your *own* mind in *that* way. So that doesn't mark a difference between the kind of relationship you have to yourself and the kind that you have to other people. All we've got here is a matter of degree. You know some people better than others; if you're honest and lucky, you know yourself pretty well.

Human beings are social animals in a deep way. It is not just that we go in for friendship or prefer to live in swarms or packs. The space of linguistic consciousness — the space in which meanings and reasons exist — is a space that we occupy together.

## The Origin of Value and the Value of Life

Pain is an objection. Interestingly, it is an objection to several of the views that I have discussed here. First, for many, pain is the biggest stumbling block to accepting Wittgenstein's views about our mental lives. It seems to them that pain is a sensation and that it is in the mind and therefore that what it is to be in pain is to have a sensation in your mind. And it seems to them that there could be a pain that was private in just the sense that Wittgenstein denied. Second, for many, pain is the biggest temptation to some form of naturalistic realism about normativity. One can have doubts about pleasure, for there are pleasures we deplore, but pain seems obviously to be a normative fact. And, third, if that is so, pain is an objection to Kantian ethics, or to any ethics that makes the value of humanity the foundation of value. For the other animals suffer pain, and if pain is intrinsically normative, then it matters that they do. Animals just as such should have moral standing.

The first two objections are related. Wittgenstein's argument against a private language deploys one of the standard objections

against any form of normative naturalism —that you cannot be wrong. Hobbes said you could only be obligated by the law if the sovereign is able to punish you. But if you break the law and get away with it, then the sovereign was not after all able to punish you and so you were not wrong. Hume says that your reason is your strongest desire. But if you always act from your strongest desire, then you always do what you have reason to do, and you cannot be wrong. Wittgenstein says that if a word just refers to the very sensation that makes you feel like saying that word, then you cannot be wrong.

But both the opponent of Wittgenstein and the normative realist point to pain, and more generally to sensation, as a case where it seems to be no objection to say that we cannot be wrong. In fact it creates a foundation. The utilitarian claims that pleasure and pain are facts that are also values, a place where the natural and the normative are one, and so where ethics can find a foundation in the world. And this is exactly analogous to the epistemological claim that our sensations are the place where the natural and the normative are one, and so where knowledge can find a foundation in the world. Sensations are seen to be intrinsically normative entities, about which we cannot be wrong.

But can't we? "I cannot be wrong about whether I am seeing red." If you mean that the object before you is red, you can certainly be wrong. "No, I mean that I am having a red sensation." And what is that? It is the sensation that makes you feel like saying that a thing is red. You are not describing a condition that explains what you are inclined to say. You are simply announcing what you are inclined to say. In the same way, someone who says he is in pain is not describing a condition that gives him a reason to change his condition. He is announcing that he has a *very* strong impulse to change his condition.

Now that way of putting it, inspired by Wittgenstein, has a problem. People have thought that Wittgenstein was making a point about *language,* to the effect that when people talk about

their own inner states and sensations they must be using language expressively, as if "I am in pain" could only be a cry of pain, and you could not simply be reporting your condition. Of course you can report your condition; once you've mastered the language, you can do anything that you like. His point is rather about mental activities, and whether a way of talking leaves anything for them to *be*. If "I see something red" *means* "I am having a red sensation" then one can never perceive; one can only announce the results of a perception that has already taken place. For what is this "having"? Did the little person in your mind perceive the red sensation? Wittgenstein is attacking a certain picture of what it is like to be conscious, which reduces all mental activity to the contemplation of sensations and ideas. And the language of "having" supports this picture. Does "I am in pain" mean "I am having a horrible sensation"? What here is the form of the "having"? Are you contemplating it? What would be so horrible about that?

But surely, you will reply, a *physical* pain is not just an impulse to change your condition. It *is* a sensation of a certain character. Now I am not denying that when we are in pain part of what is going on is that we are having sensations of a certain character. I am however denying that the painfulness of pain consists entirely in the character of those sensations. The painfulness of pain consists in the fact that these are sensations that we are inclined to fight. You may want to ask: why are we inclined to fight them if they are not horrible in themselves? Well, in some cases we are biologically wired this way; pain could not do its biological job if we were not inclined to fight it. When nature equipped us with pain she was giving us a way of taking care of ourselves, not a *reason* to take care of ourselves. Why do you thrash? Is it as if you were trying to hurl your body away from itself? Why do you say "as if"? Pain really is less horrible if you can curb your inclination to fight it. This is why it helps, in dealing with pain, to take a tranquilizer or to lie down. Ask yourself how, if the painfulness of pain rested just in the character of the sensations, it

could help to lie down? The sensations do not change. Pain wouldn't hurt if you could just relax and enjoy it.

If the painfulness of pain rested in the character of the sensations rather than in our tendency to revolt against them, our belief that physical pain has something in common with grief, rage, and disappointment would be inexplicable. For that matter, what physical pains have in common with each other would be inexplicable, for the sensations are of many different kinds. What do nausea, migraine, menstrual cramps, pinpricks, and pinches have in common that makes us call them all pains? What emotional pains have in common with physical ones is that in these cases too we are in the grip of an overwhelming urge to do battle, not now against our sensations, but against the world. Stoics and Buddhists are right in thinking that we could put an end to pain if we could just stop fighting. The person who cared only for his own virtue, if there could *be* such a person, would be happy on the rack.[28] They are wrong if they conclude that we should therefore stop fighting. Many pains are worth having; one may even say that they are true. Pain is not the condition that is a reason to change your condition, the condition in which the natural and the normative are one. It is our *perception* that we have a reason to change our condition." Pain itself is not a reason at all.

---

[28] Of course there could not be such a person, or at least, he could not have the virtues that were the only things he cared about. To have the virtues is in part to care about certain external things.

[29] When you feel pity for someone, why does it strike you as a reason to help him? Why don't you just take a tranquilizer? Hutcheson says, "If our sole Intention, in Compassion or Pity, was the Removal of our Pain, we should run away, shut our Eyes, divert our Thoughts from the miserable Object, to avoid the Pain of Compassion, which we seldom do . . ." (this passage is not in Raphael; one may find it in Selby-Bigge, *British Moralists,* p. 93). The point is reiterated by Nagel: "Sympathy is not, in general, just a feeling of discomfort produced by the recognition of 'distress in others, which in turn motivates one to relieve their distress, Rather, it is the pained awareness of their distress as *something to be relieved'* (*The Possibility of Altruism,* p. 80n). Wittgenstein says, "How am I filled with pity for *this man?* How does it come out what the object of my pity is? (Pity, one may say, is a form of conviction that someone else is in pain)" (*Philosophical Investigations* §287, p. 98). Pity is painful because it is the perception of *another's* pain, and so the perception that there is a reason to change *his* condition.

But pain is the perception of a reason. Since animals have pain, and until now I have seemed to suggest that only human beings have reasons, this will take a moment to explain.

The best account of what an animal is comes from Aristotle. We have already seen that Aristotle thought that the form of a thing is the organization or arrangement of its parts that allows it to be what it is, to do what it does, to do its job. Now Aristotle thought that a *living* thing is a thing with a special kind of form. A living thing is so designed as to maintain and reproduce itself. It has what we might call a self-maintaining form. So it is its own end; its job is just to keep on being what it is. Its business in life is to preserve its own *identity.* And its organs and activities are arranged to that end.[30]

If a living thing is an animal, if it is conscious, then part of the way it preserves its own identity is through its sensations. And this is where pain comes in. When something is a threat to its physical existence, or would be if it went on long enough, the animal perceives that fact and revolts against it. The animal is moved to take action to fix what is wrong. Suppose for instance that the animal needs nourishment. It perceives that by getting hungry. It finds this unpleasant and is moved to get something to eat. Don't be confused here: it is not that the pain is an unpleasant sensation that gives the animal a reason to eat. The animal has a reason to eat, which is that it will die if it does not. It does not know that it has that reason, but it does perceive it. The sensation in question is the sensation of hunger, not of pain. But an animal is designed to perceive and revolt against threats to the preservation of its identity, such as hunger. When it does that, it is in pain.

Now consider this comparison. A human being is an animal whose nature it is to construct a practical identity that is normative for her. She is a law to herself. When some way of acting is a threat to her practical identity and reflection reveals that fact, the

---

[30] This account of the nature of an animal is based primarily on *On the Soul,* book II.

person finds that she must reject that way of acting, and act in another way. In that case, she is obligated.

A living thing is an entity whose nature it is to preserve and maintain its physical identity. It is a law to itself. When something it is doing is a threat to that identity and perception reveals that fact, the animal finds that it must reject what it is doing and do something else instead. In that case, it is in pain.

Obligation is the reflective rejection of a threat to your identity. Pain is the *unreflective* rejection of a threat to your identity. So pain is the *perception* of a reason, and that is why it seems normative.

To say that life is a value is almost a tautology. Since a living thing is a thing for which the preservation of identity is imperative, life is a form of morality. Or to put the point less strangely and in a way that has been made more familiar to us by Aristotle, morality is just the form that *human life* takes.

From here the argument proceeds as it did in the case of other people. I won't spell out the details here. Roughly it will look like this: I first point out to you that your animal nature is a fundamental form of identity on which your human identity depends. A further stretch of reflection requires a further stretch of endorsement. If you don't value your animal nature, you can value nothing. So you must endorse its value. Perhaps that by itself doesn't show us that we have obligations to the other animals, since the value could still be private. To show us that we have obligations, animals must have a way of impressing their value upon us, the way we impress our value on each other when we ask, "How would you like it if someone did that to you?" They must be able to intrude into our consciousness and make us think.

But that isn't a problem, is it? The cries of an animal are no more mere noise than the words of a person. An animal's cries express pain, and they mean that there is a reason to change its condition. Another animal can obligate you in exactly the same way another person can. It is a way of being *someone* that you share. So of course we have obligations to animals.

*Conclusion*

I hope by now it is clear that all of the accounts of normativity that I have discussed in these lectures are true.

Voluntarists like Pufendorf and Hobbes held that normativity must spring from the commands of a legislator. A good legislator commands us to do only what it is in any case a good idea to do, but the bare fact that an action is a good idea cannot make it a requirement. For that, it must be made law by someone in a position to command us.

As we saw, that view is true, What it describes is the relation in which we stand to ourselves. The fact that we must act in the light of reflection gives us a double nature. The thinking self has the power to command the acting self, and it is only its command that can make action obligatory. A good thinking self commands the acting self only to do what is good, but the acting self must in any case do what it says.

Realists like Nagel think that reasons are intrinsically normative entities and that what we should do when a desire presents itself is to look at it more objectively, to see whether it is such an entity. This view is also true. What it describes is the activity of the thinking self as it assesses the impulses that present themselves to us, the legislative proposals of our nature.

Reflection has the power to compel obedience and to punish us for disobedience. It in turn is bound to govern us by laws that are good. Together these facts yield the conclusion that the relation of the thinking self to the acting self is the relation of legitimate authority. That is to say, the necessity of acting in the light of reflection makes us authorities over ourselves. And insofar as we have authority over ourselves, we can make laws for ourselves, and those laws will be normative. So Kant's view is also true. Autonomy is the source of obligation.

Once we see this, we can see that the reflective endorsement theory is true on another level as well. In the end, nothing can be normative unless we endorse our own nature, unless we place a

value upon ourselves. Reflection reveals to us that the normativity of our values springs from the fact that we are animals of a certain kind, autonomous moral animals. That is, in the Aristotelian sense, our human form. If we do not place a value on being such animals, then nothing will be normative at all.

That means realism is true on another level too. To see this, recall once again John Mackie's famous "argument from queerness." [31]  According to Mackie, it is fantastic to think that the world contains objective values or intrinsically normative entities. For in order to do what values do, they would have to be entities of a very strange sort, utterly unlike anything else in the universe. The way that we know them would have to be different from the way that we know ordinary facts. Knowledge of them, Mackie says, would have to provide the knower with both a direction and a motive. For when you met an objective value, according to Mackie, it would have to be —and I'm nearly quoting now — able both to tell you what to do and to make you do it. And nothing is like that.

But Mackie is wrong and realism is right. Of course there are entities that meet these criteria. It's true that they are queer sorts of entities and that knowing them isn't like anything else. But that doesn't mean that they don't exist. John Mackie must have been alone in his room with the Scientific World View when he wrote those words. For it is the most familiar fact of human life that the world contains entities that can tell us what to do and make us do it. They are people, and the other animals.[32]

[31] See "Realism" in Lecture I,

[32] I would like to thank Charlotte Brown, Peter Hylton, Arthur Kuflik, Andrews Reath, Amelie Rorty, Thomas Scanlon, Jay Schleusener, and my commentators on the occasion of the lectures (listed below) for comments on earlier versions of these lectures. A longer version of the lectures, together with commentary by Gerald Cohen, Raymond Geuss, Thomas Nagel, and Bernard Williams, is forthcoming from Cambridge University Press.

## BIBLIOGRAPHY

Aristotle. *The Complete Works of Aristotle.* Edited by Jonathan Barnes. Princeton: Princeton University Press, 1984.

Bentham, Jeremy. *A Fragment on Government; with an Introduction to the Principles of Morals and Legislation* (1776, 1789). Edited by Wilfrid Harrison. Oxford: Basil Blackwell, 1948.

Brink, David O. *Moral Realism and the Foundations of Ethics.* Cambridge: Cambridge University Press, 1989.

Brown, Charlotte. "Hume against the Selfish Schools and the Monkish Virtues." Delivered at the meetings of the Hume Society, 1989.

Butler, Joseph. *Fifteen Sermons Preached at the Rolls Chapel* (1726). The most influential of these are collected in Butler, *Five Sermons Preached at the Rolls Chapel and A Dissertation upon the Nature of Virtue.* Edited by Stephen Darwall. Indianapolis: Hackett, 1983.

Clarke, Samuel. *A Discourse concerning the Unchangeable Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation: The Boyle Lectures 1705.* (I have quoted from both J. B. Schneewind [ed.], *Moral Philosophy from Montaigne to Kant,* and D. D. Raphael [ed.], *British Moralists 1650–1800.)*

Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68 (January 1971). Reprinted in Frankfurt, *The Importance of What We Care About.* Cambridge: Cambridge University Press, 1988.

Freud, Sigmund. *Das Unbehagen in der Kultur (Civilization and Its Discontents,* 1930). Translated and edited by James Strachey. New York: W. W. Norton and Company, 1961.

Gewirth, Alan. *Reason and Morality.* Chicago: University of Chicago Press, 1978.

Grotius, Hugo. *De juri belli ac pacis* (*On the Law of War and Peace,* 1625). Translated by Francis W. Kelsey. Oxford: Oxford University Press, 1925. (I have quoted from J. B. Schneewind [ed.], *Moral Philosophy from Montaigne to Kant.*)

Harman, Gilbert. *The Nature of Morality: An Introduction to Ethics.* New York: Oxford University Press, 1977.

Hobbes, Thomas. *Leviathan* (1651). Edited by Richard Tuck. Cambridge: Cambridge University Press, 1991.

Hume, David. *The Dialogues concerning Natural Religion* (1779). Edited by Norman Kemp Smith. New York: Macmillan Library of Liberal Arts, 1947.

————. *Enquiry concerning Human Understanding* (1748) and *Enquiry concerning the Principles of Morals* (1751). In *David Hume: Enquiries concerning Human Understanding and concerning the Principles of Morals.* 3rd ed. Edited by L. A. Selby-Bigge and P. H. Nidditch. Oxford: Clarendon Press, 1975.

————. *A Treatise of Human Nature* (1739–1740). 2nd ed. Edited by L. A. Selby-Bigge and P. H. Nidditch. Oxford: Oxford University Press, 1978.

Hutcheson, Francis. *Illustrations on the Moral Sense* (Part II of *An Essay on the Nature and Conduct of the Passions and Affections with Illustrations on the Moral Sense,* 1728). Edited by Bernard Peach. Cambridge, Mass.: Harvard University Press, 1971.

————. *Inquiry concerning the Original of Our Ideas of Beauty and Virtue* (1725). (I have quoted from D. D. Raphael [ed.], *British Moralists 1650–1800;* except on one occasion when I have quoted a passage not in Raphael from L. A. Selby-Bigge, *The British Moralists* [Oxford: Clarendon Press, 1897; reprinted by the Library of Liberal Arts, 1964].)

Kant, Immanuel. *Grundlegung zur Metaphysik der Sitten* (*Foundations of the Metaphysics of Morals,* 1785). Translated by Lewis White Beck. New York: Macmillan Library of Liberal Arts, 1959.

————. *Kants gesammelte Schriften.* Prussian Academy ed. 28 vols. Berlin: Walter de Gruyter and Company, 1902–. (The page numbers found in the margins of most translations refer to this edition. When I have cited Kant, I have therefore referred to these page numbers. The English translations quoted or cited are listed in separate bibliographical entries for the works in question.)

————. *Kritik der praktischen Vernunft (Critique of Practical Reason,* 1788). Translated by Lewis White Beck. New York: Macmillan Library of Liberal Arts, 1956.

Korsgaard, Christine M. "Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations." In *Philosophical Perspectives 6: Ethics,* edited by James Tomberlin. Atascadero, Calif.: Ridgeview Publishing Company, 1992.

————. "Kant's Formula of Humanity." *Kant-Studien* 77 (April 1986), 183–202.

————. "Kant's Formula of Universal Law." *Pacific Philosophical Quarterly* 66 (January/April 1985)) 24–47.

————. "Morality as Freedom." In *Kant's Practical Philosophy Reconsidered,* edited by Yirmiyahu Yovel. Dordrecht: Kluwer Academic Publishers, 1989.

————. "Personal Identity and the Unity of Agency: A Kantian Response to Parfit." *Philosophy and Public Affairs* 18 (Spring, 1989), 101–32.

————. "The Reasons We Can Share: An Attack on the Distinction between Agent-Relative and Agent-Neutral Values." *Social Philosophy & Policy* 10 (January 1993)) 24–51.

Mackie, J. L. *Ethics: Inventing Right and Wrong*. New York: Penguin Books, 1977.

Mandeville, Bernard. *An Enquiry into the Origin of Honor* (1732). (I have quoted from J. B. Schneewind [ed.], *Moral Philosophy from Montaigne to Kant*.)

————. *The Fable of the Bees: or, Private Vices, Public Benefits* (1714). Edited by F. B. Kaye. Indianapolis: Liberty Classics, 1988; reprint of a 1924 edition published by Oxford University Press.

Mill, John Stuart. *Utilitarianism* (1861). Edited by George Sher. Indianapolis: Hackett, 1979.

Moore, G. E. "The Conception of Intrinsic Value." In Moore, *Philosophical Studies*. London: Kegan Paul, 1922.

————. *Principia Ethica* (1903). Cambridge: Cambridge University Press, 1971.

Nagel, Thomas. *The Possibility of Altruism*. Princeton: Princeton University Press, 1970.

————. *The View from Nowhere*. Oxford: Oxford University Press, 1986.

Plato. *The Collected Didlogues*. Edited by Edith Hamilton and Huntington Cairns. Princeton: Princeton University Press, 1961.

Price, Richard. *A Review of the Principal Questions in Morals* (1758). Edited by D. D. Raphael. Oxford: Clarendon Press, 1948. (I have also cited the selections in J. B. Schneewind [ed.], *Moral Philosophy from Montaigne to Kant,* and D. D. Raphael [ed.], *British Moralists 1650-1800.)*

Prichard, H. A. "Does Moral Philosophy Rest on a Mistake?" *(Mind* 21, 1912) and "Duty and Interest" (Oxford: Clarendon Press, 1929). Reprinted in *Moral Obligation and Duty and Interest. Essays and Lectures by H. A. Prichard.* Edited by W. D. Ross and J. O. Urmson. Oxford: Oxford University Press, 1968.

Pufendorf, Samuel. *On the Duty of Man and Citizen according to Natural Law* (1673). Edited by James Tully and translated by Michael Silverthorne. Cambridge: Cambridge University Press, 1991.

————. *On the Law of Nature and of Nations* (1672). Translated by C. H. Oldfather and W. A. Oldfather. Oxford: Oxford University Press, 1934. (I have quoted from J. B. Schneewind [ed.], *Moral Philosophy from Montaigne to Kant.*)

Railton, Peter. "Moral Realism." *Philosophical Review* 95 (April 1986), 163–207.

Raphael, D. D. (ed.) *British Moralists 1650–1800.* 2 vols. Indianapolis: Hackett Publishing Company, 1991; reprint of a 1969 edition published by Oxford University Press. (Where I have quoted from this anthology rather than original sources I have cited it as Raphael I and Raphael II.)

Rawls, John. "Kantian Constructivism in Moral Theory: The Dewey Lectures 1980." *Journal of Philosophy* 77 (September 1980).

————. *A Theory of Justice.* Cambridge, Mass.: Harvard University Press, 1971.

Ross, W. D. *The Right and the Good.* Oxford: Clarendon Press, 1930.

Sandel, Michael J. *Liberalism and the Limits of Justice.* Cambridge: Cambridge University Press, 1982.

Schneewind, J. B. (ed.). *Moral Philosophy from Montaigne to Kant.* 2 vols. Cambridge: Cambridge University Press, 1990. (Where I have quoted from this anthology rather than original sources I have cited it as Schneewind I and Schneewind II.)

Williams, Bernard. *Ethics and the Limits of Philosophy.* Cambridge, Mass,: Harvard University Press, 1985.

Wittgenstein, Ludwig. *Philosophical Investigations.* Translated by G. E. M. Anscombe. New York: Macmillan, 1953.