Shall I not reckon among the perfections of the human understanding that it can reflect upon itself? Consider its habits as dispositions arising from past actions? Judge which way the mind inclines? And direct itself to the pursuit of what seems fittest to be done? Our mind is conscious to itself of all its own actions, and both can and often does observe what counsels produced them; it naturally sits a judge upon its own actions, thence procures to itself either tranquillity and joy, or anxiety and sorrow. In this power of the mind, and the actions thence arising consists the whole force of conscience, by which it proposes laws to itself, examines its past and regulates its future conduct. Richard Cumberland

INTRODUCTION

3.1.1

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. Voluntarists try 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. Realists try 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 so promote our self-interest and our flourishing.

But the normative question is one that arises in the heat of action. It is as agents that we must do what we are obligated to do, and it is as agents that we demand to know why. 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 towards 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.

3.1.2

In this lecture and the next I will lay out the elements of a theory of normativity. This theory derives its main inspiration from Kant, but with some modifications which I have come to think are necessary. What I say will necessarily be sketchy, and sketchily argued. In this lecture, I will argue for two points: first, that autonomy is the source of obligation, and in particular of our ability to obligate ourselves; and second, that we have moral obligations, by which I mean obligations to humanity as such. However, it will be no part of my argument – quite the contrary – to suggest either that all obligations are moral, or that obligations can never conflict, and at the end of this lecture, I will say a little about that.
In lecture 4, I will respond to some natural objections to the argument of this lecture and, in so doing, I will develop the view further. In particular, some readers will think that the argument of this lecture shows only (or at most) that an individual has obligations to his own humanity, not that of others. In answering this worry I will be led to address the question of the scope of our obligations. I will argue first, that in the same way that we can obligate ourselves, we can be obligated by other people, and second, that we have obligations both to, and with regard to, other living things.

I will have little to say about the content of any of these obligations. I believe that the view suggests, although it does not completely settle, what that content should be, but I have made no attempt to work that out here. My aim is show where obligation comes from. Exactly which obligations we have and how to negotiate among them is a topic for another day.

Finally I will address another worry. The argument of this lecture is intended to show that if we take anything to have value, then we must acknowledge that we have moral obligations. Because that conclusion is conditional, you might think that I have not answered the sceptic. At the end of the lecture 4, I will discuss this objection.

THE PROBLEM

3.2.1

The human mind is self-conscious. Some philosophers have supposed that this means that our minds are somehow internally luminous, that their contents are completely accessible to us — that we can always 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 atten-
The authority of reflection

tion 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 perceptions and desires themselves, on to our own mental activities, 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 on to 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,

---

2 As the quotation from Cumberland at the beginning of this lecture shows, the idea that a moral motive is one approved in reflection did not originate with Kant. It is carried on the surface of the relation between the words ‘consciousness’ and ‘conscience’, as well as their Greek predecessor ‘skeidesis’ [συνείδησις] all of which mean, roughly, ‘to know in common with’ and which came to have the interesting meaning ‘to know in common with oneself’ and so ‘to be able to bear witness for or against oneself’. (I draw here on Potts, *Conscience in Medieval Philosophy*, pp. 1–2). In modern moral philosophy, the idea of the reflective endorsement of motives was brought into prominence by the work of Shaftesbury (*An Inquiry Concerning Virtue or Merit*, treatise iv of *Characteristics*) who thought of the moral sense as a kind of automatic approval or disapproval of our motives. Shaftesbury in turn was drawing on Locke’s notion of an ‘idea of reflection’, one that arises from the mind’s observation of its own activity.
names for things that automatically give us reasons, 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.

Scepticism about the good and the right is not scepticism about the existence of intrinsically normative entities. It is the view that the problems which reflection sets for us are insoluble, that the questions to which it gives rise have no answers. It is the worry that nothing will count as reflective success, and so that the work of reflection will never be done. It is the fear that we cannot find what Kant called ‘the unconditioned’.

3.2.2

The problem can also be described in terms of freedom. It is because of the reflective character of the mind that we must act, as Kant put 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’.

If the bidding from outside is desire, then the 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. As Kant puts it, we must make it our maxim to act on the desire. Then although we may do what desire bids us, we do it freely.

Occasionally one meets the objection that the freedom that we discover in reflection is a delusion. Human actions are causally determined. The philosopher’s bugbear, the Scientific World View, threatens once more to deprive us of something we value. When desire calls we think we can take it or leave it, but in fact someone could have predicted exactly what we will do.

But how can this be a problem? The afternoon stretches before me, and I must decide whether to work or to play. Suppose first that you can predict which one I am going to do. That has no effect on me at all: I must still decide what to do. I am tempted to play but worried about work, and I must decide the case on its merits.

Suppose next I believe that you can predict which one I’m going to do. You’ve done it often enough before. What then? I am tempted by play but worried about work, and I must decide the case on its merits.

The worry seems to be that if we were sure we were determined or knew how we were determined then either we could not act or we would not act, or else we would act differently. But why is this supposed to happen? Having discovered that my conduct is predictable, will I now sit quietly in my chair, waiting to see what I will do? Then I will not do anything but sit quietly in my chair. And that had better be what you predicted, or you will have been wrong. But in any case why should I do that, if I think that I ought to be working? Well, suppose that you tell me what you predict I am going to do. If you predict that I am going to work, and I think that I should work, then there is no problem. Or do I now have to do it less freely? If you predict that I am going to play, and I think that I should work, I am glad to have been forewarned. For if I am about to do what I think I have good reason not to do, then a moment of weakness or self-deception must be in the offing, and now I can take precautions against it. And then perhaps I will work after all.

If you are going to tell me what you predict I will do, then your prediction must take into account the effect on me of knowing your prediction, because otherwise it will probably be wrong. Of course it can happen, in a specific kind of case, that knowing the sort of thing I am usually determined to do diminishes my freedom. If I see that I often give in to temptation, I might become discouraged, and fight against it even less hard. But there is no reason to think that this kind of discouragement would be the general result of understanding ourselves better. Or if there is, it must come from some pessimistic philosophy of human nature, not from the Scientific World View. If predictions can warn us when our self-control is about to fail, then they are far more likely to increase that self-control than to diminish it. Determinism is no threat to freedom.

Now it will be objected that this is not what philosophers mean when they claim that determinism is a threat to freedom. They aren’t talking about a practical problem — that knowledge could somehow take away our freedom — but about a theoretical one —
that knowledge would show us we weren’t free after all. But how is it supposed to do that? By showing that we could not have done otherwise?

That might show that we aren’t responsible. But it is a different question whether determinism is a threat to responsibility. Freedom is the capacity to do otherwise, not the capacity to have done otherwise. No one has _that_ capacity, because you cannot change the past. That sounds like a joke but I mean it. The freedom discovered in reflection is not a theoretical property which can also be seen by scientists considering the agent’s deliberations third-personally and from outside. It is from within the deliberative perspective that we see our desires as providing suggestions which we may take or leave. You will say that this means that our freedom is not ‘real’ only if you have defined the ‘real’ as what can be identified by scientists looking at things third-personally and from outside.

The point here is the same as the point I made in lecture 1 against the argument that reasons are not real because we do not need them for giving scientific explanations of what people think and do. That is not, in the first instance, what we need them for, but that does not show that they are not real. We need them because our reflective nature gives us a choice about what to do. We may need to appeal to the existence of reasons in the course of an explanation of why human beings experience choice in the way that we do, and in particular, of why it seems to us that there are reasons. But that explanation will not take the form ‘it seems to us that there are reasons because there really are reasons’. Instead, it will be just the sort of explanation which I am constructing here: reasons exist because we need them, and we need them because of the structure of reflective consciousness, and so on.

In the same way, we do not need the concept of ‘freedom’ in the first instance because it is required for giving scientific explanations of what people do, but rather to describe the condition in which we find ourselves when we reflect on what to do. But that doesn’t mean that I am claiming that our experience of our freedom is scientifically inexplicable. I am claiming that it is to be explained in terms

---

1 Actually, I don’t think it does. See my ‘Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations’.
of the structure of reflective consciousness, not as the (possibly delusory) perception of a theoretical or metaphysical property of the self.

The Scientific World View is a description of the world which serves the purposes of explanation and prediction. When its concepts are applied correctly it tells us things that are true. But it is not a substitute for human life. And nothing in human life is more real than the fact we must make our decisions and choices 'under the idea of freedom'. When desire bids, we can indeed take it or leave it. And that is the source of the problem.

3.2.3

'Reason' 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 run into 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 arguments against realism await us. Does the desire or its object inherit 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. What brings such a course of reflection to a successful end?

Kant, as I mentioned, described this problem in terms of freedom. He defines a free will as a rational causality which 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 according 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 . . .' Alternatively, we may say that since the will is

---

5 Kant himself says that 'People who are accustomed merely to explanations by natural sciences' refuse to acknowledge the existence of freedom and its imperatives because 'they are stirred by the proud claims of speculative reason, which makes its power so strongly felt in other fields, to band together in a general call to arms, as it were, to defend the omnipotence of theoretical reason.' Kant, The Metaphysics of Morals, p. 378; in Gregor's translation, pp. 183-184.

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 make the law 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 making one law rather than another?

Well, here is Kant’s answer. The categorical imperative, as represented by the Formula of Universal Law, tells us to act only on a maxim which 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, as represented by the Formula of Universal Law. The categorical imperative merely tells us to choose a law. Its only constraint on our choice is that it has the form of a law. And nothing determines what the 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.7

3.2.4

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

7 This is a reading of the argument Kant gives in Foundations of the Metaphysics of Morals, pp. 446–448; in Beck’s translation, pp. 64–67; and in Critique of Practical Reason under the heading ‘Problem II, p. 29; in Beck’s translation, pp. 28–29. It is defended in greater detail in my ‘Morality as Freedom’.
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 which I 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 the argument I just gave 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 a reason, and her conduct will be that of a wanton.\(^8\) If the law ranges over the 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.

Because of this, it has sometimes been claimed that the categorical imperative is an empty formalism. And this has in turn been conflated with another claim, that the moral law is an empty formalism. Now that second claim is false.\(^9\) Kant thought that we could test whether a maxim could serve as a law for the Kingdom of Ends by seeing whether there is any contradiction in willing it as a law which all rational beings could agree to act on together. I do not think this test gives us the whole content of morality, but it is a mistake to think that it does not give us any content at all, for there are certainly some maxims which are ruled out by it. And even if the test does not completely determine what the laws of the

---

8 I have a reason for saying that her behaviour 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 should be obvious.

9 I argue for this in 'Kant's Formula of Universal Law'. There however I do not distinguish the categorical imperative from the moral law, and my arguments claim to show that the categorical imperative has content when actually they show only that the moral law has content.
Kingdom of Ends would be, the moral law still could have content. For it tells us that our maxims must qualify as laws for the Kingdom of Ends, and that is a substantive command as long as we have some way of determining what those laws would be. And there are other proposals on the table about how to do that: John Rawls’s to name only one.

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

3.3.1

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 has a reflective structure use the term too, but for a different reason. The reflective structure of the mind is a source of ‘self-consciousness’ because it forces us to have a conception of ourselves. As Kant argued, this is a fact about what it is like to be reflectively conscious and it does not prove the existence of a metaphysical 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 deliberate. When you deliberate, it is as if there were something over and above all of your desires, something which is you, and which 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 way of choosing is to be, in St Paul’s famous phrase, a law to yourself.¹⁰

¹⁰ Romans 2:14. This paragraph is lifted with modifications from my ‘Personal Identity and the Unity of Agency: a Kantian Response to Parfit’, 111. I believe there are resources in this line of thought for dealing with the problem of personal identity, and some of them are explored in that paper.
The authority of reflection

An agent might think of herself as a Citizen of the Kingdom of Ends. Or she might think of herself as someone’s friend or lover, or 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 determine whether it is the law of the Kingdom of Ends, or the law of some smaller group, or the law of egoism, or the law of the wanton that will be the law that she is to herself.

The conception of one’s identity in question here is not a theoretical 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 group, a member of a certain profession, someone’s lover or 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 behaviour 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 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 to no longer be who you are. That is, it is to no longer be able to think of yourself under the description under which you value yourself and find your life to be worth living and your actions to be worth undertaking. It 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 could just as well 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.

3.3.2

Actually, all obligation is unconditional in the sense that I have just described. An obligation always takes the form of a reaction against a threat of a loss of identity. But there are two important complications, and both spring from the complexity of human identity. One is that some parts of our identity are easily shed, and, where they come into conflict with more fundamental parts of our identity, they should be shed. The cases I have in mind are standard: a good soldier obeys orders, but a good human being doesn’t massacre the innocent. The other complication, more troublesome, is that you can stop being yourself for a bit and still get back home, and in cases where a small violation combines with a large temptation, this has a destabilizing effect on the obligation. You may know that if you always did this sort of thing your identity would disintegrate, like that of Plato’s tyrant in Republic ix, but you also know that you can do it just this once without any such result. Kant points out that when we violate the laws of the Kingdom of Ends we must be making exceptions of ourselves, because we
cannot coherently will their universal violation. In one sense, a commitment to your own identity – that is, to your integrity – is supposed to solve that problem. But as we have just seen, the problem reiterates within the commitment to your own integrity. The problem here does not come from the fragility of identity, but rather from its stability. It can take a few knocks, and we know it. The agent I am talking about now violates the law that she is to herself, making an exception of the moment or the case, which she knows she can get away with.

This is why it is best if we love our values as well as having them. But lest you think that I am about to make the same mistake of which I have accused Hume, let me admit that I think this argument establishes an authentic limit to the depth of obligation. Obligation is always unconditional, but it is only when it concerns really important matters that it is deep. Of course, since we can see that the shallowness of obligation could give rise to problems, we must commit ourselves to a kind of second-order integrity, a commitment to not letting these problems get out of hand. We cannot make an exception ‘just this once’ every time, or we will lose our identities after all. But the problem will reiterate within that commitment, and so on up the line.

That, by the way, is why even people with the most excellent characters can occasionally knowingly do wrong.

3:3:3

To get back to the point. 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 so far 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

---

12 I mean the objection at the end of lecture 2. Hume forgot that knowing that our hatred of injustice was based on general rules would have a destabilizing effect on the obligation always to be just.
which 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, 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 which 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. The realist objection – that we need to explain why we must obey that legislator – has been answered, for this is a legislator whose authority is beyond question and does not need to be established. It is the authority of your own mind and will. So Pufendorf and Hobbes were right. It is not the bare fact that it would be a good idea to perform a certain action

13 What I am saying here is that the categorical imperative is the general principle of normativity in the practical sphere. In ‘Reason and Politics in the Kantian Enterprise’, Onora O’Neill argues that the categorical imperative is the supreme principle of reason in general, which in my language means it is the supreme principle of normativity in general. It will become apparent in the course of this lecture and the next that I agree with that, although of course the idea is not completely defended here.

14 In lecture 4, 4.3.8, I present a further account of these moral emotions and how they are related to autonomy.

15 The distinction between the thinking self and the acting self is very close to Kant’s distinction between Wille (will) and Willkürlich (choice). See The Metaphysics of Morals, pp. 213–214; in Gregor’s translation, pp. 41–43.

16 In The Metaphysics of Morals, Kant says that all duties must be grounded in duties to the self, and yet that duties to the self are only intelligible if there are two aspects to the self. He calls them ‘homo noumenon’ and ‘homo phenomenon’ (pp. 417–418; in Gregor’s translation, pp. 214–215). Notice the strange alternation of one and two here: duties must arise within one, rather than between two, and yet for them to arise that one must be two. The idea of the reflective character of human consciousness, together with the thesis that obligation springs from autonomy, explains why it has to be this way.
that obligates 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.

3.3.4

With that in mind, let me return to the example I used in lecture 1 to illustrate the voluntarist conception of the motive of duty: the example of a student who takes course because it is a required. In lecture 1 I said that acting on the motive of duty as Pufendorf and Hobbes understood it seems appropriate in this kind of case. Although the student might appreciate the reasons why it is a good idea that the course should be required, it would be a little odd to say that that is his motive, since he has a decisive reason for taking the course whether he understands those reasons or not. I had in mind a story like this: you are visiting some other department, not your own, and fall into conversation with a graduate student. You discover that he is taking a course in some highly advanced form of calculus, and you ask him why. With great earnestness, he begins to lay out an elaborate set of reasons. ‘Philosophers since the time of Plato’, he says, ‘have taken mathematics to be the model for knowledge: elegant, certain, perfect, beautiful, and utterly a priori. But you can’t really understand either the power of the model or its limits if you have an outsider’s view of mathematics. You must really get in there and do mathematics if you are to fully appreciate all this . . . ’ And just when you are about to be really impressed by this young man’s commitment and seriousness, another student comes along smiling and says ‘and anyway, calculus is required in our department’.

In that story, the first student seems like a phony. Since he has that motive for taking the course, all the rest seems a little irrelevant. But now I am saying that when we are autonomous, we bind ourselves to do what it seems to us to be a good idea to do. So isn’t the first student, after all, more autonomous than the student who takes the course merely because it’s required? And isn’t the first student’s action therefore more authentically an action from duty?

If he weren’t required to take the course, and he took it for the reasons he gives you, then in one sense he would be more autonomous than the student who takes it merely because it is
required. He would be guided by his own mind, not that of another. But if he is required to take it, the reasons he gives should not be his motive. This may seem odd, since in a sense they are better reasons. But even if he understands them, they are excluded by his practical identity. Because his practical identity, in this case, is being a student. And this has two implications. First, to the extent that you identify yourself as a student, you do act autonomously in taking a course that is required. And second, it is an essential part of the idea of being a student that you place the right to make some of the decisions about what you will study in the hands of your teachers. And that means that when one of those decisions is in question, you are not free to act on your own private reasons any more, no matter how good those reasons are in themselves.17

This is not just because there is an inherent element of subordination in the position of a student. For exactly similar reasons, a good citizen cannot pay her taxes because she thinks the government needs the money. She can vote for taxes for that reason. But once the vote is over, she must pay her taxes because it is the law. And that is again because citizenship is a form of practical identity, with the same two implications. To be a citizen is to make a certain set of decisions in company with the other citizens — to participate in a general will. In so far as you are a citizen, you do act autonomously in obeying the law. And for exactly that reason, in so far as you are a citizen, you aren’t free to act on your own private reasons any more.

Some will be tempted to say that the student who understands the reasons why a course is required, and who therefore would take it even if it weren’t required, is somehow more autonomous than the student who takes the course just because it is required. If a student understands why the course is required, his taking it is endorsed both from the point of view of his identity as a student and from the point of view of his identity as a rational being with a mind of his own. So he seems to be more autonomous. But we shouldn’t be too quick to jump to the conclusion that this is the way things work in general. The student’s autonomy may be augmented in this case,

17 By ‘private reasons’ here I mean reasons arrived at by thinking through the problem yourself. In lecture 4 I deny that ‘private reasons’ in another sense, reasons that have normative force only for one person, exist. That’s not what I mean here.
because his understanding of the reasons for the requirement also helps him to make sense to himself of his being a student. It helps him to endorse his identity as a student, for it gives him confidence in his teachers’ judgment. But other cases are different. The reason for participating in a general will, and so for endorsing one’s identity as a citizen, is that we share the world with others who are free, not that we have confidence in their judgment. A citizen who acts on a vote that has gone the way she thinks it should may in one sense be more wholehearted than one who must submit to a vote that has not gone her way. But a citizen in whom the general will triumphs gracefully over the private will exhibits a very special kind of autonomy, which is certainly not a lesser form. Autonomy is commanding yourself to do what you think it would be a good idea to do, but that in turn depends on who you think you are. That’s what I’ve been saying all along.

3:3:5

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’s 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 arrangement. That is, it is the arrangement of the matter or of the parts which 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 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.¹⁸

¹⁸ These views are found throughout Aristotle’s writings, but centrally discussed in books viii–ix of Metaphysics and in On the Soul.
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: 19
1 I will keep my weapon, because I want it for myself.
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 one and three are good: maxim two is bad. What makes them so? Not the actions, for maxims two and three have the same actions; not the purposes, for maxims one and two 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 acting on that impulse as a law. So the test is a test of endorsement.

3.3.6

I’ve just claimed that realism is true after all. Realists believe that ethics is grounded in intrinsically normative entities, and a good

19 Plato, Republic, 1, 331c., p. 580.
maxim, I’ve just claimed, is exactly that – an ‘entity’ whose intrinsic properties, or internal structure, renders it normative. I want to make two points about how this form of realism is related to the more familiar views I discussed in lecture 1.

The first point concerns these questions: in virtue of what does a thing have intrinsic value or normativity, and how do we know that it does? Here we find a distinction between ancient and modern approaches to the question. Modern philosophers have tended to hold that if you can say why something is valuable, that ipso facto shows that the thing is extrinsically valuable. If I say that a hammer is good for pounding nails I am assigning it a merely instrumental and so an extrinsic value: the hammer gets its value from some further purpose that it serves. If I say that fine weather is good because today we have planned a picnic, or even just because it gives us pleasure, I do not make the weather a mere instrument, but the value still seems derivative from something outside the weather itself – namely, human purposes, interests, and capacities for enjoyment. If we extend the lesson of these cases, we may come to think that if you can say why a thing is valuable, then it does not have its value in itself. And this metaphysical view leads to an epistemological one, namely, that intrinsic values must be known by intuition. For if we cannot give a reasoned account of why something is valuable, then we cannot arrive at the knowledge that the thing is valuable by working out the reasons why it is so. So we must just ‘see’ that the thing is intrinsically good.

That Plato thought otherwise is suggested by the way he proceeds in the Republic. In Republic 11, Glaucon and Adeimantus challenge Socrates to show that justice is intrinsically good and injustice bad by showing ‘what each of them is in itself, by its own inherent force, when it is within the soul of the possessor . . .’, that is, what value there is in being just apart from any outward consequences it might have.20 Socrates of course replies by showing that justice is a form of the soul – that is, an arrangement of its parts – that makes its possessor both happy and master of himself.

Those steeped in the modern way of looking at things sometimes suppose that Plato is making a mistake here. If we give reasons why

20 Plato, Republic 11, 366e, p. 613.
justice is good, then it is only extrinsically good—good because it has these consequences, happiness and self-mastery, for the person who has it. Inward consequences may be less superficial than outward ones, and more essentially related to justice itself, but they are consequences all the same. But there is a different way to understand what is going on here. First, Plato wants to show that justice is a virtue, and a virtue makes the thing which has it good. So it is the just soul, not justice itself, which Plato aims to show is intrinsically good. And he thinks that for a thing to have intrinsic value is for it to have an internal structure that makes it good. That’s what he tries to show about the just soul in the rest of the Republic: that its internal structure makes it good. If we approach the matter this way then, as Plato thinks, we can say why a thing is good, even when its value is intrinsic.

Now it may be objected that this is not a rival conception of intrinsic value, but simply a different conception, namely the conception of virtue. For to say that something has an internal structure that makes it good must be to say that it has an internal structure that makes it good at being what it is. It is to make a claim about the thing being good at its function (its [εργον]), about its having the virtues that are proper to it. At least this is what Plato seems to mean, for Plato has Socrates argue that living and acting are the functions of the soul, and justice makes it good at those, good at living and acting. In that sense, we could say that justice gives the soul intrinsic value. But in exactly the same sense, we could say that since cutting is the function of a knife, a sharp blade gives a knife intrinsic value. But that’s just a misleading way of talking: when we say that something has intrinsic value, we do not mean merely that it has the virtues of its kind, for its kind may be of no value at all. And Plato clearly means to argue more than merely that justice is a virtue, for Socrates already did that in Republic 1, before Glaucon and Adeimantus utter their challenge. Plato also means to show that it is good to have justice and the other virtues. His argument is meant to show that a just soul is good to have for its own sake in virtue of its internal properties.

I’ll come back to Plato; I now want to approach the question from another angle. Elsewhere I have argued that it is important not to

---

21 See for instance Prichard, in ‘Duty and Interest’. 
confuse two distinctions in goodness: the distinction between final and instrumental value on the one hand, and the distinction between intrinsic and extrinsic value on the other. The distinction between final and instrumental value concerns our reasons for valuing something: whether we value it for its own sake or for the sake of some other end which it serves. The distinction between intrinsic and extrinsic value concerns the source of its value: whether it has its value in itself or gets its value from some outside source. Both final value and intrinsic value may seem to be in a certain way ultimate, or foundational. Which kind of value, or of normativity more generally, does the realist really need for his argument? That is, which kind of value brings a course of reflection about how an action might be completely justified to a satisfactory end?

On the one hand, it seems like it has to be a final good, or, if you will allow the expression, a final right: an end sought or an action undertaken for its own sake alone. For if the object is sought or the action undertaken for the sake of something else, then we do have a further question to ask: what about this other thing? Is it in turn really good, right, necessary or whatever? Yet on the other hand, it seems like it has to be an intrinsic good, or an intrinsic right, for an essentially similar reason: if the normativity comes from some other source, we can then raise a question about that source. This, as we saw in lecture 1, is the thought that drives realism in the first place. And we might think that the realist needs an intrinsic value for another reason as well. At least if we are to get anything like morality out of this line of thought, that is, if we are going to get categorical duties out of it, the value in terms of which we justify action must be independent of people’s particular desires and interests. And final goods are not, in that way, necessarily independent: what you value for its own sake at least sometimes depends on particular things about you, your own desires and interests.

The answer is that the intrinsically normative entity that serves the purposes of realism, the entity that brings a regress of justification to a satisfactory end, must combine these two conceptions. It must be something that is final, good or right for its own sake, in virtue of its intrinsic properties, its intrinsic structure. And we don’t need to dis-

---

22 Korsgaard, 'Two Distinctions in Goodness'.
cover such values by intuition, if we can show that a thing’s intrinsic properties make it a final good.23 Now this is what Plato tries to show about the just soul in the Republic: that its intrinsic properties make it a final good, a thing worth having for its own sake. In the same way, a maxim is an entity whose intrinsic properties make it a final reason for action, a final ‘right’. Something which has the form of a law, that is, which is a law by virtue of its internal structure, is intrinsically suited to answer the question why the action it dictates is necessary. In this sense, a good maxim is exactly the sort of entity which the realist argument requires.

The second clarification in a way follows from the first, and concerns the criticisms I levelled against substantive realism in lecture 1. Values are not discovered by intuition to be ‘out there’ in the world. Good maxims are intrinsically normative entities, but they are also the products of our own legislative wills. In that sense, values are created by human beings. Of course we discover that the maxim is fit to be a law; but the maxim isn’t a law until we will it, and in that sense create the resulting value. This is what Pufendorf means when he says that moral entities are produced by imposition, and that imposition in turn is a matter of making laws.24 The form of realism I am endorsing here is procedural rather than substantive realism: values are constructed by a procedure, the procedure of making laws for ourselves.

3.3.7

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

23 In ‘Two Distinctions in Goodness’ I argue that part of the problem with G. E. Moore’s idea that we recognize intrinsic values by the intuitions we have when we view certain ‘organicunities’ in isolation is that it conceals the fact that an organic unity has value in virtue of its structure, of the internal relations of its parts (pp. 193–195). Moore did insist that for a thing to have value its parts have to be combined in just the right way – that is the whole point of the doctrine of organicunities – and this suggests that he shares Plato’s sense – and Kant’s – that a thing has intrinsic value in virtue of its internal structure. But he did not think that we could say anything about how the structure gives the thing intrinsic value: we just have to recognize, by intuition, that it does so. This is where I think he goes wrong. Now I am making a similar point against Nagel and about reasons. We do not recognize reasons by intuition but by examining their internal structure – that is, of course, the internal structure of the maxims of acting on them.

24 See lecture 1, 1.3.1; and Pufendorf, The Law of Nature of Nations, in Schneewind 1, p. 171.
thing as obligation. The reflective structure of human consciousness sets us a problem. Reflective distance from our impulses makes it both possible and necessary to decide which ones we will act on: it 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 which identifies us with the source of those reasons. In this way, it makes us laws to ourselves. When an impulse – say a desire – 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. What I have established so far is that obligation in general is a reality of human life. That we oblige ourselves is simply a fact about human nature, and our maxims can be seen as intrinsically normative entities. But there is still a deep element of relativism in the system. For whether a maxim can serve as a law still depends upon the way that we think of our identities. And as I’ve said already, different laws hold for wantons, egoists, lovers, and Citizens of the Kingdom of Ends. In order to establish that there are particular ways in which we must think of our identities, and so that there are moral obligations, we will need another step.

**Moral Obligation**

3.4.1

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, or, if you prefer, refers in a formal way to the solution of that problem. The problem is what we might call the distribution problem: people join together in a cooperative scheme because it will be better for all of them, but
they must decide how its benefits and burdens are to be distributed. A *conception* of justice is a principle that is proposed as a solution to the distribution problem. How are we to distribute the benefits and burdens of cooperative living? ‘So that aggregate happiness is maximized’ is the utilitarian conception of justice. ‘So that things are as good as possible for the least advantaged, in so far as that is consistent with the freedom of all’ is Rawls’s. The concept names the problem, the conception proposes a solution. The normative force of the conception is established in this way. If you recognize the problem to be yours, and the solution to be the best one, then the solution is binding upon you.\(^25\)

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 at, and care about in our lives. ‘Right’ names the more specific problem of which actions we may perform.\(^26\) The ‘thinness’ of these terms, to use Bernard Williams’s language, comes from the fact that they are, so far, only concepts, names for whatever it is that solves the problems in question. We need *conceptions* of the right and the good before we know what to do.

---

\(^25\) At least until a better solution is proposed. See Rawls, *A Theory of Justice*, section 9.

\(^26\) The distinction between the right and the good is a delicate one which it is a little difficult to articulate clearly. One way to put it would be to say that ‘rightness’ refers to the way action relates us to the people with whom we interact, whereas goodness, at least as applied to action, refers to the way in which it relates us to our goals and the things we care about. It would not follow from the fact that an action was good, in the sense that it related us correctly to what we aim for and care about, that it was right: its rightness has to do also with its acceptability to those with whom we interact. But this way of describing the difference might be misleading in two ways. First, it might make it sound as if right and wrong refer only to actions which relate us to others and not to those which concern only ourselves. That is not what I mean: we can wrong ourselves, but this is because we can interact with ourselves. I know that this sounds paradoxical. But look: someone who becomes addicted to a drug is not just failing to do what will best serve his future interests. He is hurting *himself*. He is making himself weaker, less free, and less competent, and his future self will be in a sense cornered by what he is doing now. So he is not treating himself with respect; he is using himself as a mere means. Its effects on his interests makes the addiction *bad*; its effects on himself, and the self-disrespect that imposing those effects expresses, makes it *wrong*. The second way in which this formulation of the distinction might be misleading is that the way I’ve put it might make it sound as if being rightly related to other people is not among the things we aim for and care about. Of course it is, and for this reason right actions are normally also good.
3.4.2

How do we get from concepts to conceptions? As suggested above, what mediates is a conception of practical identity. This conception both embodies the problem and serves as an aid in finding the solution. For example, in Rawls's argument, we move from concept to conception by taking up the standpoint of the pure liberal citizen, who has only the attributes shared by all the citizens of a well-ordered liberal state: a willingness to abide by whatever principles of cooperation may be chosen in the original position, and her own conception of the good. We ask what laws such a citizen has reason to adopt. And in so far as we regard ourselves as such citizens, those are laws which we have reason to accept. In Kant's argument, we move from concept to conception by taking up the standpoint of a legislative Citizen in the Kingdom of Ends, and asking what laws that kind of citizen has reason to adopt. Again, in so far as we regard ourselves as Citizens of the Kingdom of Ends, those laws are ones we have reason to accept. Citizen of the Kingdom of Ends is a conception of practical identity which leads in turn to a conception of the right.

3.4.3

If this is correct, then Williams is wrong to say that reflection is not inherent in, or already implied by, what he calls 'thick ethical concepts'. Thick ethical concepts stand to thin ones as conceptions to concepts. Since they are normative, they are essentially reflective, and that means they embody a view about what is right or good.

And there is another implication. Williams concluded that our ethical concepts, unlike the ones we employ in the physical sciences, need not be shared with members of other cultures. But our thin ethical concepts, although not necessarily our thick ones, will be shared, even with the alien scientific investigators that his argument invokes. For the fact that they are scientific investigators

---

27 Rawls, A Theory of Justice, section 2, p. 8; and section 4, p. 19; and 'Kantian Constructivism in Moral Theory', 524ff.
28 See lecture 2, 2.3.6.
29 See lecture 2, 2.3.2.
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 they will have views about what is right and what is good, and their language will have terms in which these views are expressed. So we will be able to translate our own terms into their language, and to talk to them about the right and the good. And if we can come at least to see their conceptions as solutions to the normative problems that they face, there will even be a kind of convergence.

But neither the fact that we will share thin ethical concepts with the aliens nor the way in which reflection is inherent in thick ones suggests that we are converging on an *external* world of objectively real values. Value is grounded in rational nature – in particular in the structure of reflective consciousness – and it is projected on to the world. So the reflection in question is practical and not theoretical: it is reflection about what to do, not reflection about what is to be found in the normative part of the world.

3.4.4

But this does not eliminate the element of relativism that Williams has sought to preserve. The mediation between concepts and

---

30 There are, I think, at least two ways in which this could happen. One is that the aliens’ psychology might be quite different from ours. Perhaps we can imagine that nature equipped them to deal with danger by some mechanism other than the emotion of fear, for example, so that they will not need courage. (Some people think that they might have no emotions at all, although this is less obviously imaginable than it seems at first sight. Some sort of affect which will direct attention in useful ways is absolutely requisite to getting around in the world at all.) A more interesting possibility is that their identities might be constructed quite differently from ours. Some of the possibilities explored in both science fiction and the personal identity literature might be true of them: they might be ‘series people’ (see Parfit, *Reasons and Persons*, pp. 289–293) or exist in clone clusters or have no genders, so that the relationship between their practical lives and their physical lives would really be very different from ours. These exotic possibilities may actually more closely resemble the cultural differences we really find among human beings. Human beings – indeed all animals – have strong psychological resemblances, but our identities are constructed in very different ways. For instance, not being able to conceive yourself except as a member of a certain family might be like being a series person or a member of a clone cluster.
conceptions comes by way of practical identity. A view of what you ought to do is a view of who you are. And human identity has been differently constituted in different social worlds. Sin, dishonour, 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 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 dishonour 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 and disorder, its common enemy’. And that is now true. But we coherently can grant that it was not always so.

3.4.5

But this is not to say to say that there is nothing to be said in favour 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 conceiving one’s identity, then the point will be that some ways of thinking of our identity are healthier and better for us than others. The basic claim here would be that it is better for us to think of ourselves, and more essentially to value ourselves, just as human beings than, say, as men or women, or as members of certain religious or ethnic groups, or as the possessors of certain talents. Or at

least it is better if these other conceptions are governed by a value one places on oneself as simply human, a member of the party of humanity.

Obviously, without the resources of psychoanalytic and sociological theory we cannot envision what this kind of argument would look like in any detail. But it is a striking fact that philosophers who promote the adoption of Enlightenment liberal ideas have often appealed to arguments of this kind. In *The Subjection of Women*, for example, Mill points out the damaging effects on men of identifying themselves in terms of gender. In *A Theory of Justice*, Rawls argues that the view of human talents as a kind of shared social resource, which he thinks would result from the just society he envisions, would make it easier for people to maintain a sense of self-worth. Both of these arguments are meant to show that societies which accord equal value to human beings as such are better for people and that this is one reason to have them.

Of course there are also different ways of thinking of what it means to be valuable as a human being, or as a member of the party of humanity. Citizen of the Kingdom of Ends, participant in a common happiness, species being, one among others who are equally real, are different conceptions of the human-being-as-such among which further sorting would have to be done.

### 3.4.6

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 which is employed in its arguments is an ‘empty self’. 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 we human beings need to constitute our practical identities, and if it is successful what it establishes

---

34 See for instance Michael Sandel, *Liberalism and the Limits of Justice*. 
The authority of reflection

is a universal fact, namely that our practical identities must be constituted in part by particular ties and commitments.\textsuperscript{35} The liberal who wants to include everyone will now argue from that fact. And the communitarian himself, having reflected and reached this conclusion, now has a conception of his own identity which is universal: he is an animal that needs to live in community.

And there is a further implication of this which 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 which 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.\textsuperscript{36}

And this further stretch of endorsement is exactly what happens. 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, which he makes because he now identifies in a certain way with us all. 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.

\textsuperscript{35} Ideas along these lines are developed in the first chapter of Scott Kim's unpublished dissertation, Morality, Identity, and Happiness: an Essay on the Kantian Moral Life. Kim works out a position on the relation between particular commitments and moral commitment which he is then able to use in an effective argument against those who criticize Kantian ethics on the grounds that it is too impartial and leaves no room for particular commitments. I have been deeply influenced by Kim's dissertation and our conversations about these issues.

\textsuperscript{36} We can see this as the kind of argument Nagel appeals to in The Possibility of Altruism (see especially chapter xi). The communitarian now has two views of himself. Subjectively, he feels essentially tied to this particular community. But when he looks at himself more objectively, he sees himself as an animal that needs to be tied to some community or other. In order to prevent dissociation between these two views of his identity, he now needs to accord normativity to the more objective view if he is going to retain it for the more subjective one.
So we may begin by accepting something like the communitarian’s point. It is necessary to have some conception of your practical identity, for without it you cannot have reasons to act. We endorse or reject our impulses by determining whether they are consistent with the ways in which we identify ourselves. Yet most of the self-conceptions which govern us are contingent. You are born into a certain family and community, perhaps even into a certain profession or craft. You find a vocation, or ally yourself with a movement. You fall in love and make friends. You are a mother of some particular children, a citizen of a particular country, an adherent of a particular religion, because of the way your life has fallen out. And you act accordingly – caring for your children because they are your children, fighting for your country because you are its citizen, refusing to fight because you are a Quaker, and so on.

Because these conceptions are contingent, one or another of them may be shed. You may cease to think of yourself as a mother or a citizen or a Quaker, or, where the facts make that impossible, the conception may cease to have practical force: you may stop caring whether you live up to the demands of a particular role. This can happen in a variety of ways: it is the stuff of drama, and perfectly familiar to us all. Conflicts that arise between identities, if sufficiently pervasive or severe, may force you to give one of them up: loyalty to your country and its cause may turn you against a pacifist religion, or the reverse. Circumstances may cause you to call the practical importance of an identity into question: falling in love with a Montague may make you think that being a Capulet does not matter after all. Rational reflection may bring you to discard a way of thinking of your practical identity as silly or jejune.\(^37\)

What is not contingent is that you must be governed by some conception of your practical identity. For unless you are committed

---

\(^{37}\) I have not mentioned giving up a practical conception of your identity (or deciding that you aren’t free to give one up) for moral reasons here. This is not because I don’t think that happens, of course, but because this argument is supposed to explain why moral identity has a special status. Until that conclusion is established, conflict between morality and other forms of identity just counts as one case of conflict between identities.
to some conception of your practical identity, you will lose your grip
on yourself as having any reason to do one thing rather than
another – and with it, your grip on yourself as having any reason to
live and act at all. But this reason for conforming to your particular
practical identities is not a reason that springs from one of those par-
ticular practical identities. It is a reason that springs from your
humanity itself, from your identity simply as a human being, a reflect-
ive animal who needs reasons to act and to live. And so it is a reason
you have only if you treat your humanity as a practical, normative,
form of identity, that is, if you value yourself as a human being.

But to value yourself just as a human being is to have moral
identity, as the Enlightenment understood it. So this puts you in
moral territory. Or at least, it does so if valuing humanity in your
own person rationally requires valuing it in the persons of others.
There’s an objection to that idea, which I will take up in the next
lecture. For now, I will assume that valuing ourselves as human
beings involves valuing others that way as well, and carries with it
moral obligations.

If this is right, our identity as moral beings – as people who value
themselves as human beings – stands behind our more particular
practical identities. It is because we are human that we must act in
the light of practical conceptions of our identity, and this means
that their importance is partly derived from the importance of
being human. We must conform to them not merely for the reasons
that caused us to adopt them in the first place, but because being
human requires it. You may give up one of your contingent practi-
cal roles. But so long as you remain committed to a role, and yet fail
to meet the obligations it generates, you fail yourself as a human
being, as well as failing in that role. And if you fail in all of your
roles – if you live at random, without integrity or principle, then
you will lose your grip on yourself as one who has any reason to live
and to act at all.

Most of the time, our reasons for action spring from our more
contingent and local identities. But part of the normative force of
those reasons springs from the value we place on ourselves as
human beings who need such identities. In this way all value
depends on the value of humanity; other forms of practical iden-
tity matter in part because humanity requires them. Moral identity
and the obligations it carries with it are therefore inescapable and pervasive. Not every form of practical identity is contingent or relative after all: moral identity is necessary.

3.4.8

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 relativism 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. He 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 – and for the various physiological, psychological, and social conditions which gave rise to those 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. If complete normative scepticism is to be avoided – if there is such a thing as a reason for action – then humanity, as the source of all reasons and values, must be valued for its own sake.

3.4.9

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 human consciousness has a reflective structure that sets us normative problems. It is because of this that we require reasons for action, a conception of the right and the good. To act from such a conception is in turn to have a practical conception of your identity, a

---

38 Kant, *Foundations of the Metaphysics of Morals*, pp. 427–428; in Beck’s translation, pp. 45–47. I am here summarizing the interpretation of this argument I give in ‘Kant’s Formula of Humanity’.
conception under which you value yourself and find your life to be worth living and your actions to be worth undertaking. That conception is normative for you and in certain cases it can obligate you, for if you do not allow yourself to be governed by any conception of your identity then you will have no reason to act and to live. So a human being is an animal who needs a practical conception of her own identity, a conception of who she is which is normative for her.

But you are a human being and so if you believe my argument you can now see that that 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 conceivably reject. It is simply the truth. It is because we are such animals that our practical identities are normative for us, and, once you see this, you must take this more fundamental identity, being such an animal, to be normative as well. You must value your own humanity if you are to value anything at all.

Why? Because now that you see that your need to have a normative conception of yourself comes from your human identity, you can query the importance of that identity. Your humanity requires you to conform to some of your practical identities, and you can question this requirement as you do any other. Does it really matter whether we act as our humanity requires, whether we find some ways of identifying ourselves and stand by them? But in this case you have no option but to say yes. Since you are human you must take something to be normative, that is, some conception of practical identity must be normative for you. If you had no normative conception of your identity, you could have no reasons for action, 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 value your own humanity if you are to act at all.

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

3.4.10

The argument I have just given is a transcendental argument. I might bring that out more clearly by putting it this way: rational
action exists, so we know it is possible. How is it possible? And then by the course of reflections in which we have just engaged, I show you that rational action is possible only if human beings find their own humanity to be valuable. But rational action is possible, and we are the human beings in question. Therefore we find ourselves to be valuable. Therefore, of course, we are valuable.

You might want to protest against that last step. How do we get from the fact that we find ourselves to be valuable to the conclusion that we are valuable? When we look at the argument this way, its structure seems to be like that of Mill's argument, which proved that if there were any utilitarians, they would find their morality to be normative, and invited us to think that therefore utilitarianism is normative.

But my argument, unlike Mill's, will not fail to find its target. For Mill's readers were not already utilitarians, or did not acknowledge themselves to be so, but you are already human beings, and do acknowledge yourself to be so.

And there's a good reason why the argument must take this form after all. Value, like freedom, is only directly accessible from within the standpoint of reflective consciousness. And I am now talking about it externally, for I am describing the nature of the consciousness that gives rise to the perception of value. From this external, third-person perspective, all we can say is that when we are in the first-person perspective we find ourselves to be valuable, rather than simply that we are valuable. There is nothing surprising in this. Trying to actually see the value of humanity from the third-person perspective is like trying to see the colours someone sees by cracking open his skull. From outside, all we can say is why he sees them.\[39\]

Suppose you are now tempted once more to say that this shows that value is unreal just as colour is unreal. We do not need to posit the existence of colours to give scientific explanations of why we see them. Then the answer will be the same as before. The Scientific World View is no substitute for human life. If you

\[39\] This is why Prichard, in 'Does Moral Philosophy Rest on a Mistake?' says that when we fall into doubt about whether we have obligations the remedy is to place or imagine ourselves in a situation where we are really obligated (pp. 16–17). The normative force of reasons, obligations, and values, is a force that is felt by a deliberating agent and is imperceptible from outside of the deliberative perspective.
The authority of reflection

The argument I have just given is, as I said a moment ago, a transcendental argument. What it is really intended to show is this: that if you value anything at all, or, if you acknowledge the existence of any practical reasons, then you must value your humanity as an end in itself. Or, I might put it, if you are to have any practical identity at all, you must acknowledge yourself to have moral identity – human identity conceived as a form of normative practical identity – as well. And this identity like any other carries with it obligations.

I take this argument to show that any reflective agent can be led to acknowledge that she has moral obligations. What makes morality special is that it springs from a form of identity which cannot be rejected unless we are prepared to reject practical normativity, or the existence of practical reasons, altogether – a possibility about which I will say more in the next lecture. Our other practical identities depend for their normativity on the normativity of our human identity – on our own endorsement of our human need to be governed by such identities – and cannot withstand reflective scrutiny without it. We must value ourselves as human.

But I do not take the argument to show that all obligations are moral, or that moral obligations always trump others. In fact the argument requires – and our nature requires – that we do have some more local and contingent identities, which provide us with most of our reasons to live and to act. Moral identity does not swamp other forms of identity: no one is simply a moral agent and nothing more. Bernard Williams is right when he says that if morality demanded that of us, it would be incoherent. But it

---

40 See Williams, 'Persons, Character, and Morality' and 'Moral Luck'.

Thinks colours are unreal, go and look at a painting by Bellini or Olitski, and you will change your mind. If you think reasons and values are unreal, go and make a choice, and you will change your mind.
would be wrong to conclude that therefore either moral obligation, or our other obligations, can’t be unconditional. To conclude that would not be to affirm the possibility of conflict, but rather to remove its sting. Conflicting obligations can both be unconditional; that’s just one of the ways in which human life is hard.

To clarify the point, we should distinguish between two kinds of conflict. One may have a practical identity that is in and of itself contradictory to the value of humanity – say, the identity of an assassin. Or, one may have a practical identity that is not by its nature contrary to moral value, but that leads to a conflict with it in this or that case. The first kind of identity, and the conflicts it generates, is, I think, ruled out by the course of reflection I have tried to describe. In so far as the importance of having a practical identity comes from the value of humanity, it does not make sense to identify oneself in ways that are inconsistent with the value of humanity. But the second kind of conflict cannot be ruled out in this way. Conflict between the specific demands of morality and those of some more contingent form of identity may still exist.

3.5.2

One source of this second kind of conflict deserves special mention, though I cannot give it a full treatment here: personal relationships. Autonomy-based views of the sort I am advocating here are often thought to be unduly individualistic, or even to exclude deep forms of affiliation with others. In this section I want to explain why I think my view does not do this, and also why I think personal relationships can give rise to a special kind of conflict.

To do this, it will be helpful to contrast the view of personal relationships which I take to be correct with another view which is popularly assumed, although usually not fully articulated, in the philosophical literature. According to this other view, a personal relationship is an affectional tie – one loves or likes the other, where that is thought of as being some sort of an emotion. The emotion either consists in or causes two characteristic desires: you want to be around this other person, and you desire his happiness or more generally his good. This view is supposed to have three important implications. First, since having personal relationships is a matter
of having certain desires and feelings, even an egoist can have such relationships. One does not need moral character to sustain them. Second, and relatedly, personal relationships are quite different in kind from moral ones. Moral relationships are governed by reason and supposedly demand that we be relentlessly impartial; personal relationships are governed by affection and pull us towards partiality. Third, and as a consequence, there is an inherent tension between personal relationships and morality, and that in two senses. First, treating someone as a friend is quite different from treating him morally, and may in a certain way exclude treating him morally. Being motivated by a sense of duty is at odds with being motivated by affection. Second, personal relationships draw us to forms of partiality and favouritism which morality supposedly frowns on.

I think that this view is mostly nonsense. In the first place, the contrast on which it draws – the contrast between being motivated by reason and being motivated by affection – is, on my view, incoherent. To be motivated ‘by reason’ is normally to be motivated by one’s reflective endorsement of incentives and impulses, including affections, which arise in a natural way. More importantly, the account completely leaves out the essential element of willed commitment. On the Kantian account which I favour, by contrast, a personal relationship is a reciprocal commitment on the part of two people to take one another’s views, interests, and wishes into account. This kind of reciprocity leads to what Kant called ‘a unity of will’, for the two parties must, at least in the areas their relationship is concerned with, deliberate as one. Personal relationships are therefore constitutive of one’s practical identity. One is a member of the party of humanity, a Citizen of the Kingdom of Ends, but one is a member of many smaller and more local communities as well. A personal relationship is a Kingdom of Two – two who are committed to being in a special degree ends for one another.

---

41 It is only in cases of reflective rejection that the impulse to act or refrain has to ‘come from reason’. For example, when I discover that my impulse to break a burdensome promise must be reflectively rejected, that discovery itself must be the source of a new impulse, an impulse to keep the promise. This second impulse is strictly speaking what Kant called ‘respect for law’. But respect for law more generally is expressed by the standing commitment to act only on morally endorsable impulses.

42 Kant, Lectures on Ethics, p. 167.
On this view, personal relationships are structurally just like moral ones, except that they normally involve more fully realized forms of reciprocity. Friends do not merely refrain from making one another unhappy, but actively pursue each other’s interests, for example. The virtues called upon by personal relationships are the same as those called upon by moral ones: charity and respect. A real diehard egoist could not be a friend, but this is not exactly because you have to be moral to be a friend. It is imaginable that someone might stand in this relationship of shared life and deliberation with a few select persons while scorning humanity, and exercise the virtues of charity and respect only towards those few. For the reasons I have been laying out in this lecture, I think that position is reflectively unstable, but it is still possible. More importantly, however, even those who do acknowledge their obligations to humanity at large will see their obligations to particular others as having independent force. Thus personal relationships are not completely subsumed under morality, but they are not affectional ties of a wholly different kind either.43

Personal relationships, then, as a form of practical identity, are independent sources of obligation, like moral obligations in their structure but not completely subsumed under them. And the thought of oneself as a certain person’s friend or lover or parent or child can be a particularly deep form of practical identity. There is no obvious reason why your relationship to humanity at large should always matter more to you than your relationship to some particular person; no general reason why the laws of the Kingdom of Ends should have more force than the laws of a Kingdom of Two. I believe that this is why personal relationships can be the source of some particularly intractable conflicts with morality.

Conclusion

3.6.1

In this lecture I have tried to establish two points. First, the reflective structure of human consciousness gives us authority over our-

43 These views are spelled out in a little more detail in my ‘Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations’.
selves. Reflection gives us a kind of distance from our impulses which both forces us, and enables us, to make laws for ourselves, and it makes those laws normative. To make a law for yourself, however, is at the same time to give expression to a practical conception of your identity. Practical conceptions of our identity determine which of our impulses we will count as reasons. And to the extent that we cannot act against them without losing our sense that our lives are worth living and our actions are worth undertaking, they can oblige us.

Being human, we may at any point come to question the normativity of one or another of our practical identities, to ask why we must live up to them and conform to their laws. Why should it matter whether I live up to the demands imposed upon me by citizenship, or motherhood, or my profession? Most of the ways in which we identify ourselves are contingent upon our particular circumstances, or relative to the social worlds in which we live. How can we be bound by obligations which spring from conceptions of our identity which are not in themselves necessary?

This leads to the second point. The course of reflection to which this kind of question gives rise leads us to recognize the form of identity which stands behind the others: our identity as human, that is, as reflective animals who need to have practical conceptions of our identity in order to act and to live. To treat your human identity as normative, as a source of reasons and obligations, is to have what I have been calling ‘moral identity’.

In one sense, moral identity is just like any other form of practical identity. To act morally is to act a certain way simply because you are human, to act as one who values her humanity should. Among the many things that you are, you are a member of the party of humanity, or a Citizen of the Kingdom of Ends. And this identity like any other carries certain obligations. But moral identity also stands in a special relationship to our other identities. First, moral identity is what makes it necessary to have other forms of practical identity, and they derive part of their importance, and so part of their normativity, from it. They are important in part because we need them. If we do not treat our humanity as a normative identity, none of our other identities can be normative, and then we can have no reasons to act at all. Moral identity is therefore
inescapable. Second, and for that reason, moral identity exerts a kind of governing role over the other kinds. Practical conceptions of your identity which are fundamentally inconsistent with the value of humanity must be given up.

The views as I have presented it so far leaves three important worries unaddressed. First, you may think that I have shown only (or at most) that you must place a value on your own humanity, but not yet that you therefore have obligations to other human beings. Valuing your own humanity does not require valuing the humanity of others. Second, you may object that moral concern should not be limited to human beings at all: animals and other parts of nature should have moral standing. And third, you may worry that I have not really answered the sceptic, for I have several times said that we must value our humanity if we are to value anything at all, but I have not said why we must value anything at all. In the next lecture, I will develop my account further by responding to these objections.