

LECTURE 3

The authority of reflection

Christine Korsgaard

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, and 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 Cumberland¹

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

¹ Cumberland, *Treatise of the Laws of Nature*, 1672, in Schneewind 1, pp. 146–147.

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-

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.² 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 'syncidesis' [συνεπίσις] 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’.³ **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.

³ Kant, *Foundations of the Metaphysics of Morals*, p. 448; in Beck’s translation, p. 66.

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.⁴ 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.

freedom as
constructed

freedom as real
only in the
'deliberative
perspective'

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

⁴ 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'.⁵ 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 . . .'⁶ 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.

⁶ Kant, *Foundations of the Metaphysics of Morals*, p. 446; in Beck's translation, p. 65.

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.

Frankfurt's
account

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.