

# Self-Constitution: Action, Identity and Integrity

## Lecture Two

### Practical Reason and the Unity of the Will

Christine M. Korsgaard

2.1.1 In my last lecture I explained the general thesis for which I am arguing in these lectures: that action is self-constitution. By this I mean that we human beings constitute our own personal or practical identities – and at the same time our own agency - through action itself. We *make* ourselves the authors of our actions, by the way that we act. As I said last time, this apparently paradoxical thesis depends on the ideas that action requires agency, and agency requires unity. An action is a movement attributable to an agent considered as an integrated whole, not merely a movement of a part of an agent. Since some ways of acting unify their agents better than others, some actions are more thoroughly actions than others. The actions that are most genuinely actions, I will argue, are the ones that accord with the principles of practical reason. These standards are therefore constitutive standards of actions, standards that we must be at least trying to follow if we are to count as acting at all. And this is what explains their normativity, or the way that they bind us.

In this lecture, I will make a first stab at defending this thesis – or anyway, at making it less abstract - by showing how the principles of practical reason unify and so constitute the will. I will also contrast this account of the grounds of their normativity to some other more familiar accounts of practical principles motivate and bind us. Strictly speaking, I believe that the conclusion I reach in this lecture is not short of the conclusion I hope to reach in these lectures in general, but I don't think that will be obvious. In the

lectures that follow, I will be developing richer conceptions of action, agency, and identity that I hope make it clearer what self-constitution involves.

2.2.1 First I need to say a word about which principles I am talking about when I talk about the principles of practical reason. In the philosophical tradition, three kinds of practical principles have been proposed as rational principles. First, there is the principle of instrumental reason, the principle that instructs us to take the means to our ends. Kant identifies the instrumental principle as one kind of hypothetical imperative, as he calls it a technical imperative. It says that if we will an end, we must also will the means to that end. Kant's derivation of this principle, which I will be discussing later in this lecture, identifies the means to an end with the *causes* of the end. But nowadays the hypothetical imperative is widely taken to extend to ways of realizing ends that are not in any technical sense *causes* of those ends, for instance to what is sometimes called "constitutive" means. Say that my end is outdoor exercise; here is an opportunity to go hiking, which is outdoor exercise; therefore I have reason to take this opportunity, not exactly as a means to my end, but as a way of realizing it. When we take this line of thought to extremes, it appears that any case in which your action is guided by the application of a concept to an actual particular object or event in the world is governed by the hypothetical imperative. Compare, for example: I need a hammer; *this* is a hammer; therefore I shall take *this*, not exactly as a means to my end but as a way of realizing it. Some of Aristotle's examples of practical syllogisms are exactly like this. Consider for example this one from *On the Movement of Animals*: "I want to drink, says appetite; this is drink, says sense or imagination or thought: straightaway I drink."(701a33-34). Or take the notorious "dry

food” syllogism of *Nicomachean Ethics* VII, in which Aristotle toys with the idea that weakness of will occurs in a man who believes that “Dry food is good for any man” when he reasons that “I am a man” and “such and such food is dry” but then fails to exercise the knowledge that “this food is such and such.” In this way the hypothetical imperative may be extended to cover *any* case of action that is self-conscious, in the sense that the agent is guided by a conception of the state of affairs he is trying to realize in the world. In this guise, the hypothetical imperative appears to be the general normative principle of practical application, the principle by which we are guided whenever we consciously attempt to realize our ends in the world in any way. It is the hypothetical imperative in this most general form, the principle of practical application, that I will be discussing in this lecture, and that I claim is constitutive of volition.

2.2.2 Moving to the other extreme, moral principles have often been identified as principles of practical reason. Here it is important to distinguish between views according to which moral conduct is in fact rational, as judged by some other principle of reason, and views according to which moral principles are themselves principles of reason. Neo-Hobbesian arguments purporting to show that morality is in our self-interest are examples of the first type of view. Morality is deemed rational judged by the standard of self-interest, which is supposedly a rational standard. That is not the sort of thing I am talking about here: I am talking about the view that moral principles are direct dictates of reason themselves.

Moral philosophers have proposed several different kinds of moral principles as principles of practical reason. First, intuitionistic philosophers and dogmatic rationalists

have proposed that reason is the source of certain substantive principles dictating the performance of certain act types – keep your promises, tell the truth, and so forth. Or more modestly, they may suppose that certain substantive characterizations of actions – for instance, that it would be kind, or help someone in need – count as reasons for those actions. Second, some consequentialists have proposed that a principle directing that we maximize the good is a principle of reason. I will have more to say about maximizing principles in a moment. Finally, of course, Kant proposed his formal principle, the categorical imperative, which tells us to act only on maxims we can will as universal laws, as a principle of reason. It is this principle, the categorical imperative, that I will argue is constitutive of volition.

2.2.3 Now let me mention, only to lay aside for now, the issue of what I will call the middle principle. Kant identified what he thought was a second sort of hypothetical imperative, a principle of prudence, which instructs us to take the means, in a broad sense of means, to our own happiness. And many philosophers have agreed with Kant about this, believing that there is some sort of rational principle of self-interest, instructing us to promote or to maximize our own good, the satisfaction of our desires or whatever it might be. They have also agreed with Kant in categorizing this principle along with the instrumental principle as a hypothetical imperative, occasionally even treating it as a mere extension or application of the instrumental principle itself, that is, characterizing self-interested reasoning as “instrumental.”

I believe that this way of thinking of the principle of prudence or self-interest, as an application or extension of the instrumental principle, is completely confused. The

instrumental principle tells us only to take the means to any given end; it tells us nothing about what our ends should be. It therefore cannot possibly tell us that we ought to pursue a maximum of satisfaction, or any other form of overall good. In fact it does not even say that we ought to prefer the satisfaction of a conjunction of any two of our ends, even when there is no conflict, to the satisfaction of any one of them. No doubt there is some rational requirement of that sort – a point I shall come back to – but it is not the principle of instrumental reason. Nor of course does the instrumental principle say that we ought to prefer a maximum of satisfaction or pleasure or good to any particular satisfaction or pleasure or good when there *is* a case of conflict.

This is particularly obvious when the supposed maximum is composed of quite disparate objects. Suppose one can argue that it is irrational to prefer a penny to a dollar (or I guess here I should say a pound), on the grounds that a pound is more of the same thing. It certainly wouldn't *follow* from that sort argument that it is irrational to prefer a torrid though ultimately heartbreaking love affair to a lifetime of amicable matrimony or a whole bottle of champagne to a hangover-free Sunday. For the argument that one must prefer the pound to the penny will work only, or at most, on the assumption that one values both of them for the sake of their exchange value alone, while only the crudest hedonist thinks one values everything for the sake of some common experience or result.

But in fact the problem is much greater than that makes it seem, for there is a problem even for the crudest hedonist. Even the argument for preferring the penny to the pound can't be made on the basis of instrumental reason. If I want to buy a piece of penny candy, the instrumental principle judges a penny and a pound to be *equally good* means so far as *that* end is concerned. It is only on the assumption that I ought to pursue or stand

ready to pursue more than one of my ends and also that my other ends might cost money that I am *required* to prefer the pound to the penny. And as I have just pointed out, the instrumental principle does not require us to prefer the achievement of a conjunction of our ends to the achievement of any one of them. An argument in favor of pursuing one's greater good and also of always preferring it to one's lesser good, can't follow from the instrumental principle, because it must be in place *before* the instrumental principle tells us to prefer the greater to the less. And of course if we leave crude hedonism and its easy commensurability behind, any argument that is to show that one ought to prefer the hangover-free Sunday to the bottle of champagne or the years of happy matrimony to the torrid love affair on grounds of greater good will have to do so in terms of some quite specific conception of the good. The maximization of satisfaction over time, as well as any other form of overall good, is a *substantive* end, not a formal one, and any principle directing us to promote and prefer it would be a substantive principle. It would not be a hypothetical imperative, since it dictates the adoption of a particular end, and it would certainly not be arrived at by an application of instrumental reason.

While I think the views I've mentioned about the middle principle are confused, there is a truth behind them. As I said before, it seems overwhelming plausible to believe that there is a rational requirement that we should prefer the achievement of a conjunction of our ends to the achievement of any one of them. And once that idea is on the table, we also seem to need some method of balancing our various ends against one another when they cannot be practically combined. That is why we need a "middle principle," something between the instrumental principle and the categorical imperative, since neither of them will do this job. But there are many possible ways we could balance our various

ends against one another in order to choose among them. Given this fact it is perhaps unsurprising that there is more disagreement among philosophers about the correct formation of this middle principle than about any other part of the theory of practical reason. Some philosophers think it requires us to maximize the sum total of our satisfactions or pleasures over the course of our whole lives; others, that it requires us merely to give some weight, possibly discounted, to the ends and reasons we will have in the future as well as the ones we have now. Derek Parfit's "present aim" theory requires only that we try to satisfy our "present" desires, projects, and aims to as great an extent as possible. And it is important to notice that the idea that we should have some way to balance our ends against one another is in itself independent of the idea that when we balance it is with an eye to our personal greater good.

For the rest of this lecture I am going to leave the complicated issue of the middle principle aside, after making just one point. I have suggested that maximization principles are substantive and not formal. Indeed this is why Kant's imperative of prudence fits so badly with the rest of his theory, since his other principles are formal ones. I have no formulation to offer for the middle principle, but it seems clear to me that if we can formulate a middle principle that is formal rather than substantive, then it should be easy for me to make my case for it. What I mean is that it seems rather *obvious* that a formal principle for balancing our various ends and reasons must be a principle for unifying our agency, since that is so exactly why we need it: so that we are not always tripping over ourselves when we pursue our various projects, so that our agency is not incoherent. Indeed, the argument for crude hedonism is actually often *based on* this formal consideration. There must be some sort of common coin that makes our ends

commensurable, says the hedonist, for how else could we ever choose among them? Here the need to balance precedes the theory of the good, instead of the possibility of balancing following from it. Perhaps in general a thought about the unity of our agency rests behind the conviction that there must *be* some single or unified human good that we can always see ourselves as striving for in our actions. My thesis is less obvious when applied to the hypothetical and categorical imperatives, to which I now turn.

2.3.1 I will begin by arguing against two other accounts of the normativity of practical principles, grounded in the empiricist and rationalist traditions respectively, and starting in each case by looking at the hypothetical imperative.

Empiricists have standardly assumed that hypothetical imperatives either are automatically normative or do not need to be normative because they are automatically motivating. This view has emerged not so much in direct arguments for the hypothetical imperative as in arguments against the reason-giving force of supposedly more ambitious rational principles. Hume's famous argument to the effect that reason is the slave of the passions is the locus classicus for this sort of thing, and Bernard Williams's argument that internalism favors a view similar to Hume's is another example. Indeed I think we may see Williams's view as generalizing Hume's argument to fit the broader version of the hypothetical imperative I mentioned earlier, the hypothetical imperative as the principle of practical application in general.

According to Hume, the role of reasoning is to ascertain the relations between things. The only relation he thinks could conceivably be directly relevant to action is the causal relation. And that relation can have motivational impact only if we have a desire to

obtain or avoid one of the two objects thus related. As Hume seems to picture it, the relations ascertained by reason function hydraulically, providing conduits by which motivational force passes from the desire for the end to the idea of taking the means, thus making the idea of taking the means desirable. It does not follow that there is any rational requirement to take the means, nor I think does Hume suppose that it does. On Hume's view we are reliably motivated in this way, and he says explicitly that it is only in cases of theoretical error, either about the nature of the object or about the means, that we ever do anything that might be called irrational. Our desire for an object may be based on a mistake, as when we wrongly believe that something would be pleasant; or we may be wrong about what would constitute the means to achieving it. Even then, Hume says, it is strictly speaking the false belief, not the action or the desire that prompts it, that is contrary to reason. Hume asserts quite explicitly that when mistakes of this kind are out of the case, we are *always* in fact motivated to take the means to our ends. We are never instrumentally irrational.

But this apparently optimistic view of human rationality threatens to dissolve into tautology. The problem arises when we ask what makes something someone's end. Suppose someone claims to desire a certain object. We inform him that taking a certain action is the adequate and sufficient means to the achievement of that object, yet he fails to form the desire to do that action. Then we are entitled to conclude that he does not desire the object, or does not desire it enough to inspire him to take those means. That being so, the object is not his *end*, and that being so, he has not failed to act on any instrumental reason that he has. If all we mean by your "end" is that which you in fact pursue, it is *conceptually* impossible for you to fail to take the means to your end. If you fail to pursue

something, then it isn't your end, and then you don't act irrationally in failing to pursue it. But then the force of saying you acted on an instrumental "reason" becomes unclear. Your desire for the end plays a role in explaining why you took the action, but there is no requirement of taking the means to your end that has any normative force for you, and so no reason *on which* you acted.

The point I am making turns on the distinction between making a theoretical error and being guilty of practical irrationality – that is, violating a principle of practical reason. When a person's action is based on a mistake, the person does the wrong thing, objectively speaking, but that does not show that the person is practically irrational. A person who adds a little dry vermouth and some olives to glass of vinegar, believing it to be a glass of gin, is not doing anything irrational, for by her own lights the action makes perfectly good sense. There is nothing amiss with her motivation, nothing wrong with her will: it is only her factual judgment that needs correcting. But if people cannot ever be guilty of violating the instrumental principle by their own lights, then it is not a rational principle. It is simply a description of the inevitable effect that a certain kind of judgment has on the human will: prove to a human being that a certain act will promote a sufficiently desired object, and that will cause her to desire to take that act.

On the empiricist view, then, rational principles turn out to be essentially *descriptions* of the effects that certain judgments or ideas have on the will. But there is a problem with this conception of rationality, which is that it cannot support the normative use of "ought." Some philosophers – I think Williams may be among them – believe that the possibility of theoretical error or ignorance is sufficient to support the normative "ought." But I think that this is mistaken. For according to the empiricist view, if I say to

you “you really ought to see a dentist about that tooth” all that I can mean – *all* – is that if you came to understand that a visit to the dentist is essential to the achievement of your end of avoiding toothache, you would in fact be motivated to go to the dentist. And this is not a consideration *on which* you can act. The “ought” judgment here is not really a recommendation, but rather a sort of hypothetical prediction. And it is *not* that I predict you would be motivated to go to the dentist if you understood that going to the dentist would promote your end of avoiding toothache because you would then see that you have a *reason* to go. It is not *that*, for on the empiricist view the claim that you have a reason to go *just amounts to* the claim that if you made the means-end judgment in question you would be motivated to go. So it turns out that what looks like the normative “ought” is really just a version of the “ought” of expectation. On this view, saying of someone on the brink of toothache that he ought to go to the dentist is exactly like saying of someone who is late that he ought to be home by now. Given human nature, we would have predicted that the person on the brink of toothache would be motivated to go to the dentist; just as given the distance, we would have predicted that the person who left the office an hour ago would be home by now. If these predictions turn out to be false we know that something has gone wrong – perhaps theoretical error or ignorance. But what has gone wrong can no more properly be described as a failure of practical reason in the first case than in the second.

The inadequacy of the view is clear from this fact: there may be many principles that accurately describe the way human beings are characteristically motivated, or the effects that certain judgments have on the will. And this conception of rationality leaves us with no way of distinguishing which ones are principles of reason and which ones are

not. We *can* reliably predict that people will be motivated to take the means to their ends, or to act on the hypothetical imperative more generally, mistakes and ignorance being out of the case. But we also can reliably predict that that people will buy objects that come in packages with erotic pictures on them. And no doubt the transfer of libido from the package to the object occurs in the same way that, on Hume's account, desire is transferred from the end to the means. We are not therefore tempted to think that buying seductively packaged objects is rationally required of us.

As I will be arguing in the lectures that follow, the reason why we are confident that taking the means to your end is rationally required while buying seductively packaged objects is not is that taking the means to your end is essential to agency, while buying seductively packaged objects is not. That difference is not adequately captured by the theory that rational principles essentially describe the effects of judgments or ideas on the will. Hume's view depends on an inadequate conception of action. For Hume an action is nothing more than a movement caused by a judgment or idea. That, I will be arguing, is not what it is.

2.4.1 The rationalist view seems to take a step in the right direction, for on a rationalist view action is not merely caused by a judgment, but rather guided by it. The traditional or dogmatic rationalist is a realist about moral principles. The view is one of the most common and regularly recurring positions in the tradition of modern moral philosophy, and versions of it have been held by Samuel Clarke, John Balguy, and Richard Price in the 18<sup>th</sup> century, William Whewell in the 19<sup>th</sup> century, and W. D. Ross, H. A. Prichard, and more recently Derek Parfit in the twentieth and twenty-first. Here, however, I must enter

a caveat: realists of this stamp ordinarily endorse what I earlier characterized as substantive rather than formal principles of practical reason – the view is that certain act-types or actions just are inherently right or wrong. In more recent incarnations, the view is that certain irreducibly normative reasons for action simply exist. The combination of this view of normativity with a formal principle of reason is unusual, but as I have argued elsewhere, there is evidence that Kant himself somewhat confusedly held such a view in his earlier writings. I won't rehearse those arguments here. In any case it is certainly not uncommon for people to interpret Kant as holding this sort of view. Sidgwick, I think, reads Kant in this way.

Hume famously accused this sort of view of what we nowadays call externalism. That is, he argued that it is impossible to explain how rational principles of the type in question are supposed to motivate us. His own contemporaries, philosophers like Clarke and Price, saw no difficulty here: they simply claimed that the perception of rightness necessarily motivates. That dispute, I think, may be a standoff. But if we think about normativity, rather than motivation, then we will find that there is something in Hume's complaint.

The realist supposes that there are eternal normative verities of some sort – facts about which act-types or actions are right, or facts about what counts as a reason for what. How do we act on these verities? Apparently, by applying them in particular instances. We apply our knowledge that an action is right by choosing it. It sounds natural enough. But notice that this sort of account could not possibly explain the normativity of the hypothetical imperative. We can see this by thinking about how it would have to work. The agent would have to recognize it, as some sort of eternal normative verity, that it is

good, or required, or that there is a reason, to take the means to his ends. Let's say, just to make a choice, that it is rationally required. How does he act on this recognition? How does he apply it to the case at hand? The problem is that the extended version of the hypothetical imperative is itself, as I mentioned earlier, the principle of application. But we cannot explain how we are motivated to act on the hypothetical imperative, much less how we are bound by it, by appealing to the hypothetical imperative itself. We would have to say that an agent's end is to do what is rationally required, and that he sees that actions conforming to the hypothetical imperative are rationally required. Taking the means to his ends is therefore *itself* a means to his end of performing rationally required actions, and he chooses to conform to the hypothetical imperative under the influence of – the hypothetical imperative itself. The point is that the hypothetical imperative cannot be a normative truth that we *apply* in practice, because it is the principle *in accordance with which* we act when we apply truths in practice.

Even if this were not an incoherent way of thinking about what happens when we act on the hypothetical imperative, it would be an incoherent way of thinking about its normativity. For on the picture I have just sketched, the hypothetical imperative can bind us only by way of a prior commitment to doing what is rationally required. So the very phenomenon we are trying to explain – rational requiredness – must be, so to speak, front-loaded into the picture. And for this reason, the realist picture in fact works no better for moral principles than it does for the hypothetical imperative. For even if we know what makes an action morally required, so long as that is just a piece of knowledge, that knowledge has to be applied in action by way of the hypothetical imperative. But the hypothetical imperative cannot itself bind us to do what is rationally required. There

would have to be some further obligation to take the performance of rationally required actions as our end. And this further obligation could not just be another piece of knowledge, for the same problem would then arise once again. What obligates us to apply it?

The problem here rests in thinking of the principles that define the obligatory and the forbidden as standards we *apply* when we are deliberating about what to do. Normative standards for things other than action do work in this way. Having decided, say, that you are going to buy a car, you then ask yourself what makes for a good one, or perhaps what makes for a good one for you. But these standards exert their normativity, if they do so at all, through action itself – they guide our choices. But the normative standards for actions themselves cannot work the same way. To see the problem, consider the fact that in the case of things other than action, you don't absolutely *have to* apply the normative standards. Usually you have the same reason for choosing a good X (or a good X for you) that you have for choosing an X. But it is at least imaginable that you might just *pick* an X without reference to its goodness, like someone who is asked to pick a number from one to ten or randomly picking a cookie from a passing tray. So if you are *obligated* to pick the best X or a good X, we need some further argument for that. But no further argument could be given why we are obligated to choose obligatory actions. So long as we think of the principles of rational action as something we may or may not *apply* when we deliberate about what to do, then we cannot really be obligated to perform rationally required actions. For either we are obligated to apply the standard for rationally required action or we are not. If we say we are not, there is no normativity, whereas if we say that we are, because the standard is known to hold, the whole argument starts all over

again – why must we apply this piece of knowledge? This is the truth in Hume’s complaint about the rationalists. Knowledge is external to the will, and external standards cannot obligate the will.

2.4.2 The argument I just gave is essentially the same as a now-familiar argument, to the effect that the standards of logic cannot enter into reasoning as premises. Suppose George does not reason in accordance with modus ponens. He does not see how you get from “If A then B” and “A” to the conclusion that “B.” As is often pointed out, it does not help to add modus ponens as a premise, that is to say, to add “if A and B and also A then B.” For you still need to reason in accordance with modus ponens in order to get any conclusion from these premises, and this is what George does not do. The argument I’ve just given against the rationalist account of normativity amounts to an argument that the principles of practical reason cannot obligate us to act if they enter practical deliberation as premises. Adding practical principles as premises cannot bind us to act, just as adding logical principles as premises cannot bind us to draw the conclusion. And this case is related to the case I am trying to make here in another important way. For it is important to notice that if George lacks logic, his mind will be a disunified jumble of unrelated atomistic beliefs, unable to function as a mind at all. It will be a *mere heap* of premises. And that is where the normativity comes in. What obligates him to believe B in these circumstances is not merely his beliefs that if A then B and also A, but rather modus ponens itself. And it is not his *belief* in modus ponens that obligates him to believe B, for that, as the argument we have just looked at shows, is irrelevant. What obligates him to

believe B is that if he does not reason in accordance with modus ponens, he will not have a mind at all.

The principles of practical reason, if they are to be normative, must be principles of the logic of practical deliberation. And this brings us at last to Kant.

2.5.1 Kant's derivation of the hypothetical imperative is simple. The hypothetical imperative says that if you will an end, you must will the means to that end. This principle, Kant tells us, is analytic. For, he says:

Whoever wills the end, wills (so far as reason has decisive influence on his actions) also the means that are indispensably necessary to his actions and that lie in his power. This proposition, as far as willing is concerned, is analytic. For in willing an object as my effect there is already thought the causality of myself as an acting cause, i.e., the use of means.

What makes willing different from merely desiring or wishing or thinking-it-would-be-nice-if is that the person who wills an end determines himself to bring the end about, that is, to cause it. And to determine yourself to be the cause of an end is to determine yourself to set off a chain of causes that will lead to the achievement of the end. Thus the person who wills an end *constitutes himself* as the cause of the end.

The account goes back to Kant's original definition of rationality. A rational being, according to Kant, is distinguished from everything else in nature by the fact that it acts not merely in accordance with laws, but in accordance with its own representation or conception of a law. If I pitch my pen into the air, it will act in accordance with the law of gravity: it will eventually fall back down to the earth. But it doesn't say to itself when it

reaches its acme “I guess I’d better go back down now.” But this is the sort of thing we do when we act. We don’t choose to obey the law of gravity, of course. But in certain circumstances, we might choose to go back down. For instance, maybe I am climbing a mountain, and if I don’t go back down now I won’t make it home before sunset. So I say to myself: I guess I’d better go back down now, in order to make it home before sunset. That thought – I will go back down now, in order to make it home before sunset – is my maxim and it is also my conception of a law. And it is because my maxim determines what I do that my movements count as willed – that is to say – as an action. In other words, it is because I determine myself to go back down that my movement is attributable to me. Rationality is a power of self-determination.

This is a general point, not just a point about practical reason. Go back to the case of logic. Perhaps you don’t arrive at all your beliefs through reasoning, but when you do, it’s an act of self-determination, in the sense that the activity of your own mind is part of what produces the belief in you. Suppose you believe two premises, and a certain conclusion follows. You won’t automatically believe that conclusion, because you might not notice the connection between them. But if you do notice the connection, and put the premises together in the way suggested by the connection, then you *do* something: you *draw* the conclusion. In drawing the conclusion – or, as we say, in making up your mind, in *constituting* your mind – you determine yourself to believe it. The principle of modus ponens describes what you do when you draw the conclusion, but it is also a normative principle. In the same way, the hypothetical imperative describes what you do when you will an action: you determine yourself to be a cause, the cause of something. But it is also a normative principle. It is a constitutive standard for the will.

Of course in the case of the hypothetical imperative this conclusion seems very natural. Last time I said that constitutive standards are particularly immune to skeptical challenge. Judging by the philosophical tradition, hypothetical imperatives, unlike moral principles, seems to be immune in this way, for most philosophers do not even pause to ask what makes them normative. And certainly we do not imagine a character parallel to the amoralist, who rebels in a general way against the hypothetical imperative. Take the means to my ends! Bah Humbug! Why should I have to do that? Such a character could not be rebelling against anything less than the necessity of action itself, and as I argued in my last lecture, that's one form of rebellion that isn't open to us.

2.5.2 But Kant thought that the fact that action is self-determination, that is, determining yourself to be a cause, has a further implication. It also implies that the categorical imperative is a constitutive standard of volition. For determining yourself to be a cause is not the same as allowing some part of you, say some desire or impulse – to use Kant's own language now, some *incentive* - to operate as a cause. The person who determines her own causality must operate as a whole, as something over and above her parts, when she does so. And in order to do this, Kant believed, she must will her maxims as universal laws.

To see why, we only need to consider what happens if we try to deny it. The argument that follows is what I call “the argument against particularistic willing.” If our reasons did not have to be universal then they could be completely particular - it would be possible to have a reason that applies only to the case before you, and has no implications for any other case. Willing to act on a reason of this kind would be what I will call

“particularistic willing.” If particularistic willing is impossible, as I will argue, then it follows that willing must be universal - that is, a maxim, in order to be willed at all, must be willed as a universal law.

Now I must start with a couple of caveats to avert confusion. First of all, the question before us - the question whether particularistic willing is possible - is not whether we can will a new maxim for each new occasion. We may very well do that, for every occasion may be different in relevant ways from the ones we have previously encountered. In Kant’s theory, any difference in the situation that is actually relevant to the decision properly belongs in our maxim, and this means that our maxim may be quite specific to the situation at hand. The argument here is not supposed to show that reasons are general. It is supposed to show us that reasons are universal, and universality is quite compatible with - indeed it requires - a high degree of specificity.

The second caveat is that it will be enough for the argument if the principle that is willed be, as I will put it, provisionally universal. Let me define my terms so you will see what I mean by these caveats. There are three different ways in which we can take our principles to range over a variety of cases, and it is important to keep them distinct. We treat a principle as *general* when we think it applies to a wide range of similar cases. We treat a principle as universal, or, as I will sometimes say, *absolutely universal*, when we think it applies to absolutely every case of a certain sort, but all the cases must be exactly of that sort. We treat a principle as *provisionally universal* when we think it applies to every case of a certain sort, unless there is some good reason why not. The difference between regarding a principle as universal, and regarding it as provisionally universal, is marginal. Treating a principle as only provisionally universal amounts to making a mental

acknowledgment, to the effect that you might not have thought of everything needed to make the principle universal, and therefore might not have specified it completely. Treating principles as general, and treating them as provisionally universal, seem superficially similar, because in both cases we admit that there might be exceptions. But in fact they are deeply and essentially different, and this shows up in what happens when we encounter the exceptions. If we think of a principle as merely general, and we encounter an exception, nothing happens. The principle was only general, and we expected there to be some exceptions. But if a principle is provisionally universal, and we encounter an exceptional case, we must now go back and revise it, bringing it a little closer to the absolute universality to which provisional universality essentially aspires.

The rough causal principles with which we operate in everyday life (I am not talking now about quantum physics) are provisionally universal, and we signal this sometimes by using the phrase “everything else equal.” The principle that striking a match causes a flame holds everything else equal, where the things that have to be equal are that there is no gust of wind or splash of water or oddity in the chemical composition of the atmosphere that would interfere with the usual connection. There are background conditions for the operation of these laws, and without listing and possibly without knowing them all, we mention that they must be in place when we say “everything else equal.” Although there are certainly exceptions, these kinds of natural laws are not merely general, for whenever an exception occurs, we look for an explanation. Something must have made this case different: one of its background conditions was not met.

To see how it works in the practical case, consider a standard puzzle case for Kant’s universalizability criterion. It may seem as if wanting to be a doctor is an adequate reason

for becoming a doctor, for there's nothing wrong with being a doctor because you want to - in fact, really, it's rather admirable - and if you ask yourself if it could be a law that everyone who wants to be doctor should become one, it seems, superficially, fine. But then the objector comes along and says, but look, suppose *everyone* actually wanted to be a doctor and nobody wanted to be anything else. The whole economic system would go to pieces, and then you couldn't be a doctor, so your maxim would have contradicted itself! So does this show that it is wrong to be a doctor simply because you want to?

What it shows is that the mere desire to enter a certain profession is only a provisionally universal reason for doing so. There's a background condition for the rightness of being a doctor because you want to, which is that society has some need for people to enter this profession. In effect the case does show that it's wrong to be a doctor merely because you want to - the maxim needs revision, for it is not absolutely universal unless it mentions as part of your reason for becoming a doctor that there is a social need. Someone who decides to become a doctor in the full light of reflection also takes that into account.

That case is easy, but there's no general reason to suppose we can think of everything in advance. When we adopt a maxim as a universal law, we know that there might be cases, cases we haven't thought of, which would show us that it is not universal after all. In that sense we can allow for exceptions. But so long as the commitment to revise in the face of exceptions is in place, the maxim is not merely general. It is provisionally universal.

So particularistic willing is neither a matter of willing a new maxim for each occasion, nor is it a matter of willing a maxim that you might have to change on another

occasion. Both of those are compatible with regarding reasons as universal. Instead, particularistic willing would be a matter of willing a maxim for exactly this occasion without taking it to have any other implications of any kind for any other occasion. You will a maxim thinking that you can use it just this once and then so to speak discard it; you don't even need a reason to change your mind.

Now I'm going to argue that that sort of willing is impossible. The first step is this: When you determine yourself to be the cause of your action you must identify yourself with the principle of choice on which you act. For instance, suppose you experience a conflict of desire: you have a desire to do both A and B, and they are incompatible. You have some principle that favors A over B, so you exercise this principle, and you choose to do A. In this kind of case, you do not regard yourself as a mere passive spectator to the battle between A and B. You regard the choice as yours, as the product of your own activity, because you regard the principle of choice as expressive, or representative, of yourself – of your own causality. You must do so, for the only alternative to identifying with the principle of choice is regarding the principle of choice as some third thing in you, another force on a par with the incentives to do A and to do B, which happened to throw in its weight in favor of A, in a battle at which you were, after all, a mere passive spectator. But then you are not the cause of the action. Self-determination, then, requires identification with the principle of choice on which you act.

The second step is to see that particularistic willing makes it impossible for you to distinguish yourself, your principle of choice, from the various incentives on which you act. Kant thinks that every action involves some incentive or other, for there must always be something that prompted you to consider the action. And in order to will

particularistically, you must in each case wholly identify with the incentive of your action. That incentive would be, for the moment, your law, the law that defined your agency or your will.

It's important to see that if you had a particularistic will you would not identify with the incentive as representative of any sort of type, since if you took it as a representative of a type you would be taking it as universal. For instance, you couldn't say that you decided to act on the inclination of the moment, *because you were so inclined*. Someone who takes "I shall do the things I am inclined to do, simply because I am inclined to do them" as his maxim has adopted a universal principle, not a particular one: he has the principle of treating his inclinations *as such* as reasons. That is the law that defines his causality. A truly particularistic will must embrace the incentive in its full particularity: it, in no way that is further describable, is the law of such a will.

But this means that particularistic willing eradicates the distinction between a person and the incentives on which he acts. But then there is nothing left here that is the *person*, the agent, that is his self-determined will as distinct from the play of incentives within him. He is not one person, but a series, a *mere heap*, of unrelated impulses. There is no difference between someone who has a particularistic will and someone who has no will at all. Particularistic willing lacks a subject, a person who is the cause of these actions. So particularistic willing isn't willing at all.

If a particularistic will is impossible, then you when you will a maxim you must take it to be universal. If you do not, you are not determining yourself to action, and then you are not willing. To put the point in familiar Kantian terms, we can only attach the "I will" to our choices if we will our maxims as universal laws. The categorical imperative is a

constitutive standard for actions, according to Kant, because conformity to it is constitutive of an exercise of the *will*, of the of the determination of a person *by himself* as opposed to his determination by something within him.

2.6.1 And now I will conclude with a promise. In this lecture I have claimed that the hypothetical and categorical imperatives are constitutive of volition and so of action. Now in my first lecture I said that only someone who knows what an object essentially is, only someone who knows its form, in the Aristotelian sense, really knows its constitutive standard. The claims I am making therefore require that we know what actions are. In this lecture I have gestured vaguely at that by claiming that action is self-determination. In the next lecture, I shall present a fuller account of what that really means.