

# Self-Constitution: Action, Identity, and Integrity

## Lecture Four

### Expulsion from the Garden: the Transition to Humanity

Christine M. Korsgaard

4.1.1 When an agent acts, even a non-human agent, the agent may succeed or fail, and to that extent the very idea of action is a normative one: it is subject to a standard of efficacy. In my last lecture I argued that actions are subject to a standard of efficacy only because they also exhibit at least a primitive kind of autonomy. Mere causality is not subject to a standard of success and failure, but if an agent determines himself to cause something, and yet does not, then the agent has failed. So only movements that can be seen as the products of self-determination are subject to a standard of efficacy. Actions are movements that are determined by the acting animal itself in the sense that they are determined by the animal's form. When an animal's form determines its movements, its movements are determined or directed by its own nature, and in that sense the animal is autonomous. In general, then efficacy and autonomy are the essential properties of agents. Since the hypothetical and categorical imperatives – from now on I'll just call them the Kantian imperatives – since the Kantian imperatives tell us to be efficacious and autonomous, the Kantian imperatives in effect tell us to be agents, and so represent constitutive standards for actions.

Oviously, however, that does not mean that non-human animals are subject to imperatives. At the end of my last lecture, I explained where the difference lies. Both animal actions and human actions involve the interaction of two factors, an incentive and a principle. The incentive is a motivationally-loaded representation of an object. It

presents the object as desirable or aversive in some specific way – as a thing to be eaten, mated with, avoided, fled, investigated, defeated or what have you. The principle determines, or we may say describes, what the animal does, or tries to do, in the face of the incentive: catch it, eat it, seduce it, get downwind of it, or whatever it might be. Incentives and principles exist in natural pairs, for its principles determine which incentives a creature is subject to as well as what it does about them. If you are a human being whose principle is to help those in need, then the neediness of another is an incentive to help. If you are a cat whose principle is to chase small scurrying creatures, then the movements of a mouse or a bug are an incentive to give chase.

As that last example is intended to suggest, a non-human animal's principles are its instincts. Its instincts are its own laws in the sense that they arise from its form, and they are the laws of its causality in the sense that they determine what it does for the sake of what. It goes downwind to escape the notice of the lion, pounces to capture the scurrying rodent, digs a burrow to build a nest in, or whatever it might be. Since its instincts are the laws of its own causality, they are in effect the animal's will, and the movements that are determined by them are its actions.

But human beings are agents in a further, deeper sense. For we do not merely determine ourselves in the sense that we act *from* the principles of our own causality – we determine ourselves in the deeper sense that we *choose* the principles of our own causality. It is up to us to decide what we will do for the sake of what. How this difference between us and the other animals emerges, and what further ramifications it has for human life, is my subject in this lecture.

4.2.1. We attribute instincts only to conscious beings, for their role is to *structure* an animal's consciousness, its conception of the world, in ways that will enable it to survive and reproduce. A non-human animal lives in a world that is in a deep way its own world, a world that is *for* that animal. I don't mean by saying this that the animal's world is favorable to its interests; often it is not. But the world as perceived by the animal is organized around its interests: it consists of the animal's food, its enemies, its potential mates, and if it is social of its fellows, its family, flock, tribe or what have you. To say that the animal is subject to incentives is just to say that its perceptions come already loaded with the practical significance of various objects for it. It confronts a world of things that are perceived directly, without calculation or conscious interpretation, as things *to be avoided*, things *to be chased*, things *to be investigated*, things *to be eaten*, things *to be fled*, things *to be cared for*, and *to be mated with* if at all possible. If you think of it, you will see that it has to be this way. Consciousness first evolved in unintelligent animals, and would have been useless if all it did was to flood their tiny minds with neutral information that needs to be processed by intelligence or reason before it is of any use. So the world comes to an animal already practically interpreted as a world of tools and obstacles, friends and enemies, of the to-be-avoided and the to-be-sought. The natural way of perceiving the world, in other words, is *teleologically*.

4.2.2 But this is not to say that non-human animals never use intelligence to find their way around the world. Here it is important for me to clarify something about the way I am using the term "instinct." I am using the term in one way more narrowly and in another more broadly than it is generally used. First, I am using it to refer to an established connection between a representation (the incentive) and a certain primitively

normative response, an automatic sense of the response as *appropriate* to or perhaps better *called for* the representation. Sometimes people use the term “instinctive” for reactions and movements that are not experienced as called for but rather are wholly automatic, like salivating. The smell of food to a hungry animal is an incentive for eating, not an incentive for salivating – it is only a cause of salivating. Still it must be admitted that the line between these two things in nature is indistinct, and this is one of the ways in which the concept of action, especially when applied to non-human animals, is indeterminate. We can form a continuum going from, say, salivating at the smell of food, to blinking when an object comes close to your eye, to ducking when something is hurled in your direction, to running when you see a predator. As we move along this continuum we move from the wholly automatic to a case where there is room in consciousness to experience the response as called-for or appropriate. It is more this latter kind of case that I have in mind when I talk about instinctive action. In that sense, my use of the term is narrow.

But in another way I am using the term “instinct” more broadly than some people do. Some people contrast the idea of “instinctive response” with a “learned response.” The term instinct as I am using it involves no such contrast. And this is what brings me to the issue of intelligence.

An intelligent animal is characterized by its ability to learn from its experiences. It is able to extend its repertoire of practically significant representations (or even just cues, for a machine may be intelligent) beyond those with which instinct (or the inventor) originally supplied it. So intelligence is a capacity to forge new connections, to increase your stock of automatically appropriate responses. Intelligence so understood is not something contrary to instinct, but rather something that increases its range and ramifies

the view of the world that it presents. After the puppy's encounter with the porcupine and the beehive, the porcupine and the beehive get added to the category of things-to-be-avoided and are now perceived that way automatically. The human being you live with becomes a pack member and yowling in the kitchen becomes a way of getting food. Not the way you would have done it in nature but you can learn that it works. Hume's description of the process by which human beings and the other animals automatically come to make new causal inferences is a pretty good description of the workings of intelligence. Hume was right to deny that reason needs to come into the story at that level, though he was wrong to deny that it needs to do so later on.

Instrumental thinking of the kind found in the higher animals is a further stretch of intelligence. By instrumental thinking, I mean figuring out that something could work as a tool or an obstacle not by accidental experience but by noticing its properties. The intelligent animal learns about the practical significance of objects – although it wasn't always obvious, this object is a food, or a predator, or a source of pain. The animal capable of *instrumental* intelligence discovers the practical significance of the *properties* of objects: the heavy object can be used to smash coconuts or the sharp one to pierce the shell. For the intelligent animal, the world of tools and obstacles presented by instinct is elaborated and changed in ways that contribute to the animal's flexibility in dealing with changes in its environment and its success in getting around. But the world it perceives is still a world teleologically structured around its own interests by its instinctive responses.

4.2.3 And this is true for us as well. Our *natural* way of perceiving and conceiving the world is teleologically. Perceptual salience in everyday life is determined by our interests and commitments, even once those begin to change in the ways that are characteristic of

the human. And this remains true even when we become *very* intelligent. As basic a fact as the fact that we divide the world into objects, each associated with a cluster of causal powers, is a way in which our perceived world is a world that is for-us. For example: why do we say that the stone broke the window? Surely it was not just the stone, but the passing car that dislodged the stone, and the driver of the car, and whatever caused him to choose that route. And surely it was not just this, but also the constitution and consistency of the atmosphere through which the stone flew, the drag of gravity upon it, the brittleness of the glass, the molecular structure that made it brittle. Was it really anything less than nature itself that broke the window? So why do we say that the stone broke the window? That is easy – because one can *use* a stone to break a window. Our world is a world of means and ends: it is divided into objects we associate with clusters of causal powers because it is a world for intelligent action, a world for us. The scientific conception of the world, the view of the world as a system of neutral laws and forces whose impact on our own fates and interests is largely accidental, requires the *detachment* of perception from the rich normative significance that naturally inhabits it. Some philosophers seem to suppose that we start from some crude but still purely theoretical conception of the world, and work out the practical significances of things, using intelligence and reason. This is backwards: the truth is that detachment of perception from practical significance, the detachment that makes the scientific conception of the world possible, is an extraordinary conceptual achievement. It is an achievement made possible, and also necessary, not by intelligence but by the interaction of intelligence with another power, namely reason, the thing that makes us us.

4.2.4 The distinguishing feature of human beings, what makes us different from the other animals, is that we are self-conscious. I am not talking now about having a self-conception, recognizing yourself as one of the objects in the public world, maybe even that fellow you see in the mirror. Some of the other animals as we know have that sort of self-conception. I am talking about the fact that we are conscious, not only of the outside world, but also of our own inner states – and so among other things of the workings of incentives within us. This, as I have argued elsewhere, sets us a problem that the other animals do not have. For our capacity to turn out attention on to our own mental states and activities is also a capacity to distance ourselves from them, and to call them into question. I perceive some situation as dangerous, and find myself with an instinctive impulse, an incentive, to run. But when I bring that impulse itself into view, when I reflect on it, then I can call it into question. Shall I run? Does this situation really give me a *reason* to run? And now I have to decide.

The first result of the development of self-consciousness is liberation from the control of instinct. Instincts still operate within us, in the sense that they are the sources of many of our incentives – in fact, arguably, though by various routes, of all of them. But instincts no longer *determine* how we respond to those incentives, what we do in the face of them. They *propose* responses, but we may or may not act in the way they propose. Self-consciousness opens up a space between the incentive and the response, a space of what I call reflective distance. It is within the space of reflective distance that the question whether our incentives give us reasons arises. In order to answer that question, we need principles, which determine what we are to count as reasons. Our rational principles replace our instincts – they will tell us what is an appropriate response to what, what

makes what worth doing, what the situation calls for. And so it is in the space of reflective distance, in the internal world created by self-consciousness, that reason is born.

4.2.5 Let me be clearer about one of the distinctions I am employing here, the distinction between reason and intelligence. Intelligence, as I said before, is a power of forging new connections, through thought and learning. It elaborates an animal's representation of the world in practically useful ways, making the animal more flexible and adaptive to change. Instrumental intelligence may even enable an animal to make choices of a sort, as between two different hiding places or tools. But what makes either of the two options worth choosing is still determined by instinct: instinct settles it that the predator is *to be avoided* or the nut *to be cracked*. Intelligence still looks outward, at the world. It is only when we become self-conscious, when we look inward, that we are faced with normative problems, and must decide what is worth doing for the sake of what. It is reason, not mere intelligence, which places us in the realm of the normative.

4.3.1 In his essay, "Conjectures on the Beginning of Human History," Kant explains one of the first results of the development of reason, namely the enormous proliferation in the possible objects of desire that is characteristic of the human. Intelligence can find new means, but when we enter the realm of reason, we begin to find new ends.

In this essay, Kant uses *Genesis* as the basis for a fanciful account of the historical development of the power of free choice. The first object of free choice in history is the apple in the story of Adam and Eve and Kant explains how it came about. Initially, Kant explains, the primordial human animal must have been guided by instinct to its natural food, just as the other animals are. We need not suppose, Kant says, that there was some

special instinct that has now been lost. Rather, the sense of smell carried with it an ability, “still in evidence today,” as Kant says, to pick out our natural food. But then, reason, operating through a power that Kant calls comparison, begins to change things. Kant explains – I’m quoting now:

... Reason soon made its presence felt and sought to extend [the human being’s] knowledge of foodstuffs beyond the bounds of instinct; it did so by comparing her usual diet with anything which a sense other than that to which her instinct was tied ... represented as similar (III.6). ... The initial incentive to abandon natural impulse may have been quite trivial. ... Thus, it may have been only a fruit which, because it looked similar to other agreeable fruits which she had previously tasted, encouraged her to make the experiment. There may also have been the example of an[other] animal to which such food was naturally congenial...[And] this was enough to give reason the initial inducement to quibble with the voice of nature (III.1), and despite the latter’s objection to make the first experiment in free choice...

I’ve taken some liberties with the translation of the pronouns here, for in Kant’s story it is obviously Eve who makes the *first* free rational choice. How does it happen? Eve is naturally drawn by the sense of smell, say, to pears. They smell like food to her, and when she is hungry her instinct directs her to eat them. But then she becomes self-conscious and so she notices the working of this incentive within her. One day when she is hungry she finds herself thinking of pears. “I have an inclination to eat some pears,” she says to herself. (Later in this lecture I’ll explain exactly how that step happens, how she comes to think “I have an inclination to eat some pears.”) There before her is the tree, dripping with juicy green apples. “Those apples look like pears” she says to herself,

making the first comparison. And there is her playmate, the serpent, munching away on an apple. “He’s like me, in some ways, and he eats apples” she says to herself, making the second comparison. And then it happens. “You know, I could eat one of those apples to satisfy my hunger,” she says to herself. That’s the first maxim, of the first free action, and there our story begins. Kant comments:

No matter how trivial the harm it did may have been, it was nevertheless enough to open her eyes. She discovered in herself an ability to choose her own way of life without being tied to any single one like the other animals. But the momentary gratification which this realization ... afforded ... was inevitably followed ... by anxiety and fear as to how she should employ her newly discovered ability, given that she did not yet know the hidden properties or remote effects of anything. She stood, as it were, on the edge of an abyss. For whereas instinct had hitherto directed her towards [specific] objects of ... desire, an infinite range of objects now opened up, and she did not yet know how to choose between them. Yet now that she had tasted this state of freedom, it was impossible for her to return to a state of servitude under the rule of instinct.” (56/224)

In the essay, Kant goes on to trace further steps by which the powers of reason are developed. Reason not only directs the human being to new objects of desire among the things she finds around her, but leads to the development of altogether *new kinds* of objects of desire. Kant anticipates Freud, for example, in suggesting that the senses of beauty and romantic love are the products of the suppression of our sexual instincts (CBHH 8:112-113), a suppression made possible by reason’s power to say no to an incentive. In this way distinctively human desires and interests come about.

But our liberation from the government of instinct is also our expulsion from the Garden, our banishment from a world that is teleologically ordered by our instincts and presented as such by our incentives, a world in which we nearly always already know what to do. It is this essential homelessness that leads to the misogyny Kant famously describes in the opening section of the *Groundwork*, the hatred of reason and nostalgia for instinct which he says is felt most sharply by cultivated people who try to find their happiness through reason. For now we must use intelligence and reason, both to reconstruct a picture of the world that we can find our way around in, and to decide where to go and what to do in that world. That is to say, we must resort to *science* to reconstruct a usable conception of the world, and to *ethics* to determine how to live our lives.

4.4.1 But self-consciousness introduces another, more immediate difference into the lives of human animals. It introduces what, following Plato, I will call the parts of the soul. Let me explain what I mean.

In one sense, as I have already emphasized, the soul has parts even without self-consciousness. Non-human animals experience incentives, and have instincts that tell them how to deal with those incentives, just as people have inclinations, and reason that tells us how to deal with those inclinations. But in another sense the non-human animal soul does not have parts, for what counts for a non-human animal as an incentive is settled by its instincts, and its instincts immediately tell it how to deal with the incentive: as I said last time, in an animal soul the work of incentive and the work of instinct are not phenomenologically distinct. So in a non-human animal, the whole psychic system is closed and tightly knit. And that means that there's a sense in which self-consciousness *produces* the parts of the soul.

4.4.2 The first and most obvious way in which it does this we have already looked at. Self-consciousness is the source of reason. When we become conscious of the workings of an incentive, the incentive is experienced not as a demand but as a proposal, something we need to make a decision about. Cut loose from the control of instinct, we must formulate principles that will tell us how to deal with the incentives we experience. And the experience of decision or choice, the work of these principles, is a separate experience from that of the incentive itself.

One effect of this division is to make the distinction between action and mere reaction sharper and crisper in humans than it is in the other animals, although it is important to admit that it even in our case it is not always sharp and crisp. The fact that human actions involve a distinct moment of choice sharpens up the distinction. Earlier I said that we could see a kind of continuum from salivating at the smell of food, to blinking when an object comes close to your eye, to ducking when something is hurled in your direction, to running when you see a predator. In the human case there are sharper divisions. We normally cannot choose whether to salivate or not, so usually that counts as pure reaction. We normally can choose whether to run. The reflective distance between incentive and principle sharpens up the difference between action and reaction, making the latter case clear. But we are talking about nature, so there are still unclear cases. Ducking, for instance, is still somewhere in the middle – often it occurs without a decision like a reaction, yet it is a reaction that can be resisted if you have some powerful reason to resist it. If I say “you must hold absolutely still, or the children will die” and you duck, then it will have been an action.

4.4.3 Self-consciousness creates the need for the principles of reason, which are then more firmly separated from the associated incentives than their instinctual predecessors were. But self-consciousness also transforms the other side of the equation – it transforms incentives into what Kant calls *inclinations*. To see this how this works we must look a little more closely at the relationship between incentives and inclinations.

Strictly speaking incentives are, as I said at the beginning of this lecture, features of the represented object that make it, from some point of view, attractive or aversive. When we say that a person has an inclination for something, what we mean that he is responding to the incentive that makes that thing attractive. For instance if dancing is pleasant, that means there is a natural incentive to dance, and if a person's awareness of that incentive is drawing him towards dancing, then we say that he has an inclination for dancing.

Now there are two important points to notice about this kind of case. One is a standard admonition. We should not say, in this kind of case, that pleasure is really the object of the inclination, rather than dancing. Dancing is the object of the inclination; to say it is pleasant is just to say that there is a natural incentive for dancing. In fact it is not even quite right to say that pleasure is the incentive: the incentive has no name, but is simply whatever it is about dancing that, given the psychological and physical characteristics of human beings, draws us to dancing. To say that dancing is pleasant is not to say exactly what the incentive is: pleasure is just a kind of dummy word that indicates that there is one. Given our nature as physical and animal beings, we need to move our muscles in certain controlled ways and we experience that movement as welcome. The arousal of our erotic capacities may also come into it, and that is welcome too. You can explain the incentive, sometimes, but at bottom, it is not something articulable. We just

gesture at the fact that there is an incentive, a positive attraction, a suitability of the object to our nature, when we say that the object is pleasant.

Having said that, let's nevertheless call the incentive, for short, "pleasure." Then the point of saying that pleasure is the person's incentive is not to say that he likes the pleasure rather than the dancing, but rather that the pleasure, the sheer joy of the movement, is what he likes about the dancing. Or we may even say that pleasure is what makes him like it, for there is nothing inconsistent with the concept of action in the fact that incentives operate on us causally. All we need for the concept of action is that neither incentives nor the inclinations to which they give rise cause the actions directly, without the intervention of principles.

The second and more complicated point is that inclinations themselves are in a sense the *products* of self-consciousness. And now I'm going to tell you, as I promised earlier, how Eve came to say to herself "I have an inclination to eat some pears." When we are talking about the actions of a non-human animal, we may *attribute* an inclination to it - we may say that the cat *wants* some food, for instance. But this is really completely third-personal and attributive, in the sense that it is just a way of describing the fact that the cat is responding to an incentive, viz., hunger -- or more precisely, since an incentive is a feature of the represented object, to the fact that food would satisfy its hunger. We don't need to say that the incentive causes an inclination, and the cat then acts from the inclination, as if the inclination were some *additional* and separate mental state. The cat does not *entertain* a desire for some cat chow: saying it has the inclination to eat is just a way of saying it experiences the incentive to eat. In fact in one way it is better not to talk about the cat's inclination, for when we mention the cat's reasons, or indeed our own, we do not mention the inclination, but rather the incentive. We say that cat seeks food

“because it is hungry” and you dance “because it is pleasant” - not “because you want to.” Saying that the cat has the inclination is just a way of *naming* the fact that the cat is responding to the incentive.

But there is a sense in which self-consciousness moves this attributive operation indoors, and does make our inclinations into mental items. For when we are conscious of the fact that an incentive is working upon us, our self-consciousness of our state does reify it into a kind of mental item. And that’s when we say of ourselves that we “have an inclination,” something we now have to decide whether to satisfy or not. So self-consciousness is the source of inclinations as well as of reason. Self-consciousness produces the parts of the soul.

4.4.4 This is a good place to explain my stance on a point of current discussion. A moment ago I said that when we tell others the reasons for our actions, we mention the incentive, not the inclination: you say you dance “for the sheer joy of dancing,” say, not that you dance “because you want to.” But if it is the inclination you decide whether to act on or not, does that mean that it is the inclination that provides the reason? And is the reason then after all “because you want to”?

Some philosophers suppose that the fact that we mention the incentive rather than the inclination favors a realist account of value. According to these philosophers, desires and inclinations are simply responses to the good-making properties of objects, and it is only the good-making properties of objects that we need to talk about when we talk about our reasons, not the desires and inclinations themselves. When we say why we pursue various objects, we mention the attractive features of those objects, rather than our own

psychological states, and so – these philosophers conclude -- it must be in virtue of the attractive features of the objects that the objects have value.

As a Kantian I disagree with the realist picture in two ways. The picture of a desire or inclination as a response to the good-making features of its object is too cerebral, making our own desires and inclinations essentially epistemic, and too unlike the incentives of our fellow creatures. In Kant's view the features of the objects we desire that we mention when we explain why we value those objects would not give those objects value were it not for the way in which those features are related to human physiology and psychology. At the basis of every desire or inclination, no matter how articulately we can defend it, is a basic suitability-to-us that is a matter of nature and not of reason. Value is relational and what it is related to is our nature. The fact that when we run out of other ways to articulate what attracts us to an object we can just say that it is "pleasant" tends to obscure this fact, for "pleasant" seems to name a good-making property of an object. But as I said a moment ago, it does not – "pleasure" is just a dummy word that indicates that some positive form of this basic suitability-to-us is in place.

The second difference follows. As a Kantian I believe that it is our own choices that ultimately confer value on objects, even though our choices are responsive to certain features of those objects. In choosing objects, in conferring value on things that answer to our nature in welcome ways, an agent is affirming her own value. She takes what matters to her to matter absolutely and so to be worthy of her choice. But even if the agent herself believes this Kantian theory, it doesn't follow that she must think of herself as choosing objects simply because she wants or likes them. She can still talk to herself, and to others, about what she likes about them, and why. So even though there *is* a sense, on my account, in which we choose things "because we want them", a sense in which the

inclination provides the reason, it doesn't follow that when someone asks you why you chose something, that "I wanted it" is the right answer. In fact the Kantian theory of value puts us in a position to offer an alternative account of the fact that we mention the incentive rather than the inclination when we want to convey our reasons to another. In the sense I am using the term here, an inclination just is the operation of a natural incentive upon you, as viewed through the reifying eye of self-consciousness. In that quite general sense all of our non-moral values depend on inclinations. So when you are explaining your values to another person, it is quite uninformative to mention the fact that you have an inclination for the object as the basis of the value. He knows that; or if he doesn't know it, what he doesn't know is a thesis in value theory, and that isn't what he wants to know. He wants to know why you value *this*, what you find attractive about it, not what the structure of human valuing consists in. To put it another way, he wants to know *which* inclination you are having, what is drawing you to the object. And you specify that by describing, as far as you can, the incentive.

[4.4.4a Realists, I said a moment ago, take the fact that when we talk about our reasons we talk about the attractive features of objects, rather than about our psychological states, to favor a realist conception of value. But some anti-realist philosophers, such as Blackburn, take the very same facts to show that the Kantian account of deliberation involves the "fundamental mistake" of thinking that deliberation looks inward to the self rather than outward to the world. According to Blackburn, the inner act of choosing to satisfy the inclination is just an unnecessary mental duplicate of the outer act of choosing the object. To suppose such a duplicate is necessary is to make a mistake analogous to that of sense-data theorists who think perceptions are mental copies of the objects we perceive.

But here it is important to keep in mind that the question of deliberation is neither whether a certain object or state of affairs would be good nor whether to satisfy a certain inclination but whether to do a certain action: to do an act for the sake of an end. No sense can be attached to the question that decision looks inward or outward. Actions are not inward or outward; they are, like utterances, intelligible objects in the public world whose nature it is to very bridge the gap between inner and outer. They are acts of thinking performed by the body, mental states in the flesh: there is no need to imagine something inner behind them that causes them. There *is* an error of the sort Blackburn describes. It is made by philosophers who deny Aristotle's dictum that the conclusion of a practical syllogism is an action. These philosophers think that deliberation leads instead to the formation of an intention, a mental object which in turn causes an action which is in itself a mere outer movement. That amounts to *denying* that the gap between inner and outer can be bridged, and that is exactly the problem with sense data theory: it adds extra entities, endless copies, in a fruitless attempt to bridge a gap it conceives as unbridgable. Intentions so conceived are unnecessary mental duplicates of actions. It's as if you thought your sentence were caused by a proposition that preceded it, rather than embodying the proposition you are trying to express.

Why do philosophers suppose we need these mental objects, these intentions? One source of the temptation rests in the fact that, being prudent, we often think about what we are going to do in advance. So I decide, for instance, now, that I am going to Rome next summer. And since I can't act on my decision immediately, I form an intention, and I store it somewhere until I'm ready to use it.

But that's not what I do. As soon as my decision is made the action begins, and I start taking the means. For now every move I make must be guided by the decision I have

taken. I must learn Italian, or buy a suitcase, or even if all these things are done, I must be careful not to do things that would obstruct or prevent my going to Rome, like spending all my money so I can't buy a ticket, or promising a friend I will spend the summer with him in Antarctica, or whatever it might be. And when I decide to go to Rome next summer, I am deciding that the end makes all of this, all of the relevant parts of the act I take to achieve it, worth doing. So there is never any gap between decision and action. The conclusion of a practical syllogism is an action; Aristotle was right.

4.4.5 Self-consciousness opens up a space between incentive and instinct; and that space transforms incentives into inclinations and governing instincts into free reason. Self-consciousness is therefore the source of a psychic complexity not experienced by the other animals, and it transforms psychic unity from a natural state into something that has to be achieved, into a task and an activity. Once we are self-conscious the soul has parts, and then before we can act it must be unified. At the very same time, and for the same reason, practical deliberation becomes necessary, for free reason need not follow inclination. We must now decide what to do. These conditions - the need to work at being unified and the need for practical deliberation - are brought about together. And this means that the function of deliberation is not merely to determine how you will act, but also to unify you. Or rather, to put the point more correctly, those are not two different things, for your movement will not be an action unless it is attributable to you - to you as a whole or a unified being - rather than merely to something in you. And the task of deliberation is to determine what you - you as a whole or a unified being - are going to do.

The phenomenology of deliberation, especially in hard cases, bears this out. Suppose we are dealing with a case where you have to choose between two options, two

courses of actions for both of which you experience some incentive. When you choose, as I've said before, it is as if there were something over and above your incentives, something which is you, and which chooses which incentive to act on. That much just follows from the fact of self-consciousness. You must decide whether to act on the incentive, and in effect that means you must decide whether to identify your will with it or not. But it is not as if the incentive itself appears in consciousness as something that is *not* you, but just something you might pick up the way you might pick up a stone off the ground. The incentive arises from your nature and it appears in consciousness as something that wants to *be* you. For the incentives are incentives to *act*, and what we feel in a hard case of the sort I have described is that we ourselves are being pulled in both directions. That's what we say in such cases: *I feel so torn*. And I think in such a case this is literally true – it is you, your will, your agency, that is in danger of tearing. I mentioned in lecture one that Socrates says that phrases like “self-control,” “self-mastery” or “self-command” are like “tracks or clues” that virtue has left in the language (430d). So are the things we say to people when it is time for them to stop dithering and bring deliberation to an end - *Make up your mind*, or even better, *Pull yourself together*. The work of practical deliberation is reunification, reconstitution: and the function of the principles that govern deliberation - the principles of practical reason - is the unification of the self. So we arrive again at the conclusion of Lecture Two – the function of practical reason is to unify us into agents who can be the authors of our actions.

4.5.1 I'll come back to the parts of the soul, and the way that we unify them, or fail to, in the next lecture. Here I want to notice another important difference all this makes to the specific character of human action. As we have seen, liberation from the government of

instinct means that it is up to us to decide what justifies what, what counts as a reason for what, what is worth doing for the sake of what. We don't need to think of this decision, and in fact we shouldn't think of it, as a decision made prior to action: it is a decision embodied in the action. Action involves an incentive and a principle; the principle describes the agent's contribution to the action. In our case, that contribution takes the form of deciding *whether* to act as the incentive bids. And that decision may be *described* as your principle. If you choose to run in order to escape your predator, to stand your ground in order to protect your offspring, or to dance for the sheer joy of dancing then those are your principles, your conception of what is worth doing for the sake of what. Human action is in that further sense autonomous; in that sense, we choose our principles.

One might put the same point this way. In lecture three I argued that an animal acts when it consciously determines itself to be the cause of a change in the world. This does not mean that the animal is conscious *of* its causality – that it forms an intention or adopts a maxim - but rather that it exercises its causality by way of its consciousness, that its movements are guided by its representations. A human being acts when she *self*-consciously determines herself to be the cause of a change in the world. Unlike the other animals, we *are* conscious of our causality, and it is therefore up to us how we exercise it. We choose the laws of our own causality. And this means, to switch now back to Aristotle's terms, that there is a sense in which we choose our own forms.

4.5.2 In lecture one I mentioned that in his book *On the Soul*, Aristotle asserts that there are three forms of life, corresponding to what he called three parts of the soul. At the bottom is a life of basic self-maintenance, a vegetative life of nutrition and reproduction, common to all plants and animals. Animals are distinguished from plants in being alive in

a further sense, given, as we have seen, by the power of action. The third form of life described by Aristotle is distinctive of human beings – the life of rational activity, the life of choice.

Each part of the soul, and each corresponding form of life, supervenes on the one below it. The addition of each new part of the soul changes the sense in which the thing lives, *both* by influencing the way the “lower” functions are carried out and by adding a new kind of activity and so a new form of life. Because it has the powers that make agency possible, an animal *lives* in a sense that a plant does not: an animal is conscious; it is active in the sense that it *does* things; it pursues what it desires and flees what it fears; in some cases it builds a home and raises a family; if it is a “higher” animal it may even know how to love and to play. But these are not just powers added, so to speak, on top of the animal’s nutritive and reproductive life: they also change the way the animal carries out the tasks of nutrition and reproduction. The animal uses its capacity for action to get its food and raise its family. But it also may do things a plant doesn’t do at all, like love and play. These things make the “life” of an animal a different sort of thing than the “life” of a plant.

A human being in turn has a “life” in a sense in which a non-human animal does not. For a non-human animal’s life is mapped out for it by its instincts; and any two members of a given species basically live the same sort of life (unless the differences are biologically fixed, as by age and gender, or by kinds as among bees). Two members of the same nonhuman animal species may indeed have very different personalities, but these do not make much difference in the way that they live. A human being has a life in a different sense from this, for a human being has, and is capable of choosing, what we sometimes call a “way of life” or following Rawls, a “conception of the good.” Where her

way of life is not completely fixed by some sort of cultural regulation, a human being decides such things as how to earn her living, how to spend her afternoons, who to have for friends, and in general, how she will live and what she will live for. She decides what is worth doing for the sake of what. And again, we find a double result. Choice changes the way we carry out the activities we share with the other animals, such as house-building, child-rearing, hunting or collecting food, playing, and sexual activity. Human beings approach these activities creatively and develop various ways of going about them among which we then choose. But we also do things the other animals don't do at all, like tell jokes and paint pictures and engage in scientific research and philosophy. Choice introduces a whole new sense of *life*, a new sense in which a person can be said to "have a life." Thus personhood is quite literally a form of life. (By the way, in case you didn't recognize it, what I just said was a rendering of Aristotle's function argument from *Nicomachean Ethics* I.7.) Since being a person, like being a living thing or an animal, is a form of life, being a person is being engaged in a specific form of the activity of self-constitution. Our chosen actions constitute us as persons.

And so with this new form of life comes a new form of identity. For each part of the soul also adds a new sense in which the creature may be said to have an identity. A plant, although alive, is basically just a substance. But an animal has a kind of identity that a plant lacks. Its consciousness gives it a point of view; it's also an agent, and it does things; it is not just a substance, it's a subject, it's a someone. If it is also fairly intelligent, you can interact with it, play with it, get annoyed at it, or adore it. You're not alone when it is with you in the room. Even if it is not very intelligent you can sympathize with it and enter into its concerns, or be hostile to it and regard it as the enemy. In other

words, it's the proper object of some of what Strawson called the reactive attitudes, or Hume called the indirect passions, because it is a someone.

Because she is alive in a further sense, then, a person has an identity in a further sense. She has an identity that is constituted by her choices. This kind of identity is in a deeper way the person's own than an animal's identity, because she is consciously involved in its construction. And it is more essentially individual than a non-human animal's, because she is free. Constructing, creating, shaping, reshaping, maintaining, improving, in all these ways constituting this kind of identity is the everyday work of practical deliberation. And it is because we are characterized by this special kind of identity that a further stretch of the reactive attitudes is appropriate to us. It because we have this kind of identity that we hold one another responsible, answerable, for what we do and what we are. It is this kind or level of identity I am talking about when I say we choose our own forms – the laws of our own causality.

Aristotle thought only God, or maybe the gods, had individual forms, but I've just claimed that it follows from his view that every human being has an individual form. Let me be clear about this. Your identity as a human animal, your human form, is given to you by nature, and you share it with the species. But the form of the human is precisely the form of the animal that must create its own form. For nature sets each human being a task: self-consciousness divides his soul into parts, and he must reconstitute his agency, pull himself back together, in order to act. And that need to reconstitute yourself introduces the necessity of exercising your freedom, and the opportunity of doing so creatively. In other words, every person must make himself into a particular person. Someone who says "I want to make something of myself" is just describing the human

condition. And it is because he makes himself into a particular person, that we hold him responsible for being who he is.

4.5.3 Yes, I did just say that. Sometimes you hear philosophers say that the idea of responsibility is incoherent, because we could not be responsible for what we do unless we are responsible for what we are, and we could not be responsible for what we are unless unless we created ourselves. I think it is true that we could not rightly be held responsible unless we created ourselves, but false that that makes the idea of responsibility incoherent. Philosophers who suppose that responsibility is incoherent are thinking of it this way: our actions spring from our nature, so we could not be responsible for the specific character of our actions unless we were responsible for our nature, and we could not be responsible for our nature unless we produced ourselves. And that makes no sense. But the sense in which we must create ourselves in order to be responsible is not that we must literally bring ourselves into being. Rather, we are responsible because we have a form of identity that is *constituted* by our chosen actions. We are responsible for our actions not because they are our products but because they are us, because we are what we do.

4.5.4 I don't intend to linger over this topic, interesting as it is, but I want to avert a possible misunderstanding of what I have just said. You might think I've just committed myself to the view that everyone is to be held completely responsible for every move he makes on the grounds that, after all, it's him. But I don't in the least mean to suggest either that we can *never* excuse someone for the performance of some particular action, or that we might not in general be hesitant about judging someone who has formed his identity under particularly disadvantageous conditions. I'm not really talking, at this

point, about blaming people or punishing them or throwing them into jail. Behind those practices is something deeper and simpler, a basic condition of holding one another answerable, accountable in the most literal sense of the term, without which human relationships would not be intelligible at all. It is because our actions are expressive of principles we ourselves have chosen, principles we have adopted as the laws of our own causality, that it makes sense for us to hold one another answerable in this way: to demand one another's reasons, and to take it, as we say, personally, when we hear what they are. As I've said before, actions are intelligible objects; our actions are no more just our products than our words are. And it makes no more sense to think that we might in general cease to hold people answerable for their actions, than it does to think that we might in general cease to hold people answerable for their words.

4.6.1 So we choose our principles, and in so doing we constitute our identities as individual human agents. This doesn't mean, of course, that we choose the hypothetical and categorical imperatives themselves. The Kantian imperatives are formal principles that instruct us in *how* to formulate our maxims; autonomy and efficacy set standards for the *form* of our maxims. The content of our maxims, I believe, is drawn from natural and social resources - from what I have elsewhere called our contingent practical identities - the social, professional, religious, ethnic, political and relational identities which give rise to our everyday reasons and obligations. What do people make themselves out of? You are born into a culture, into a particular time and place, you are born into a family and a community, you have certain talents, you are assigned certain roles, you become attached to certain people whom you happen to meet - a range of identities, actual and possible, happens to be available to you. You are someone's daughter and someone's friend, a

teacher or a student, the citizen of a certain country, perhaps a member of a certain religion. Each of these roles carries with it a set of laws that defines what it means to be that person, a set of reasons and obligations definitive of the role, a conception of what is worth doing for the sake of what that characterizes that identity. Such things provide the content of our maxims; the Kantian imperatives provide their form. Unless you will some of these laws in accordance with the Kantian imperatives, you will not succeed in being anyone at all.

4.6.2 It is because for us constructing the will is in this way a *task*, that for us the standards of efficacy and autonomy take imperative form. And this shows both why and how the Kantian imperatives are normative for the human will. In Lecture One I said that a house-builder cannot completely ignore the normative standards inherent in very the idea of a house and still recognizably be building a house, because what it *means* to build a house is to be guided by those standards. In the same way, an agent who is deliberating about what to do cannot completely ignore the Kantian imperatives and still recognizably be deliberating about what to do, because what it *means* to deliberate is to be guided by those imperatives. The point is very simple. Someone who is deliberating about what to do is deliberating about *how* to exercise his own causality, what the law of his own causality is to be. The hypothetical imperative directs him insofar *what* he is exercising is causality, and the categorical imperative directs him insofar as it is to be *his own*.

Of course an agent may not attend with much care to the standards set by the Kantian imperatives, just as a shoddy builder may not attend with much care to the standards set by the very idea of a house. But this does not show that one can intelligibly reject the normative standards expressed by those imperatives. It only shows that one can

deliberate badly. The kind of practical deliberation that issues in bad action is not a different activity from the kind of practical deliberation that issues in good action. It is the same activity, badly done.

In my next lecture I will defend this conclusion, by showing how bad action is possible.