

Moral Animals

That short but imperious word *ought*: Human Nature and the Right

Christine M. Korsgaard

Harvard University

I fully subscribe to the judgment of those writers who maintain that of all the differences between man and the lower animals, the moral sense or conscience is by far the most important. This sense, as Macintosh remarks, “has a rightful supremacy over every other principle of human action;” it is summed up in that short but imperious word *ought*, so full of high significance.

- Charles Darwin, *The Descent of Man*¹

I. Human and Animal Differences

Human beings are animals: homo sapiens, members of the phylum chordata, class mammalia, order primates, family, hominids. According to current scientific opinion, we evolved about 200,000 years ago in Africa from ancestors we share with the other great apes. But a long-cherished view of our species is that something sets us apart from the other animals - something, moreover, that is not merely a matter of degree like intelligence, but a matter of kind.

Once upon a time, for many, and still for some, this view was associated with another cherished idea - that we human beings are made in the image of God, and that the natural world and all its contents, including the other animals, were made for our use, or perhaps simply to set the scene for a drama of which we are the

¹ Ibid., p.70.

protagonists. The properties that set us apart from the other animals were supposed to be the same as those that we share with God. But even if we leave these theological motives aside, it seems tempting to believe that the human way of life is qualitatively different from that of the other animals, and that this difference must be explained by something that is distinctive about human beings. But where exactly does the difference lie? It's easy enough to point to some differences between our lives and those of the other animals. Our lives are elaborately structured by social and political institutions and cultural traditions. Our fate is both eased and complicated by complex technologies. Our experience is graced and enhanced by the creation and enjoyment of literature, music, the arts, and displays of athletic excellence. We engage in scientific research and philosophical speculation. We are theoretical creatures, who devise metaphysical and theological theories, or conceptions of the inevitable march of history, to give direction to our actions and meaning to our sufferings and our fates. We strive to live in accordance with moral values and with other sorts of values and ideals. And all of these features of human existence are facilitated by the complex communication system of human language. But do these obvious features of human life really set us decisively apart from the other animals?

Some philosophers and scientists have answered that they do not. Darwin, in his book on *The Descent of Man*, reviews the various attributes that people have claimed are exclusively human, seeking out their original forms and analogues in the lives and actions of the other animals. At the end of his review he remarks:

There can be no doubt that the difference between the mind of the lowest man and that of the highest animal is immense...Nevertheless, the difference in mind between man and the higher animals, great as it is, is certainly one of degree and not of kind.²

Darwin's lead has been followed both by scientists eager to confirm the theory of evolution and by philosophers eager to establish the moral standing of the other animals. Perhaps the other animals do not have institutions, but social animals have dominance hierarchies that organize and structure their lives, and their lives have a recognizably political dimension as a result. We now know that we are not the only animals who use tools or even the only animals who design tools or even the only animals who sometimes save the tools they have found or designed for future use. The natural communication systems of the other animals have been found more complex and versatile than scientists once believed, and some apes, birds, and dolphins have been taught the rudiments of human language. Scientists have discovered that learning plays a greater role in the lives of non-human animals than they had once suspected, and because of the role of learning, different populations of the same species of animal have been found to exhibit different "cultural" traditions with respect to matters such as which things they eat and how they procure and prepare them. Darwin himself believed that our aesthetic sensibilities were prefigured in the responses of female birds to the elaborate and colorful sexual

² *The Descent of Man*, pp. 104-105. (Princeton)

displays of their males. He also proposed that perhaps some “distant approach” to religious feeling could be discerned in the absolute reverence with which dogs and other captive animals regard the human beings on whom they so completely depend.³ And some scientists and philosophers believe that they have found the roots of morality in the sympathy, altruism, and cooperation that are sometimes exhibited by the other social animals.

It is with this last question – whether morality is a distinctively human attribute – that I will primarily be concerned in this lecture. Darwin, in the chapter I have mentioned, was among those who first pointed out the continuity between kinds of actions prompted by the social instincts and feelings of the other animals, and the kinds of actions that we credit to our own moral nature. Yet Darwin also said:

I fully subscribe to the judgment of those writers who maintain that of all the differences between man and the lower animals, the moral sense or conscience is by far the most important. This sense, as Macintosh remarks, “has a rightful supremacy over every other principle of human action;” it is summed up in that short but imperious word *ought*, so full of high significance.⁴

³ *The Descent of Man*, pp. 63-68 in the Princeton edition.

⁴ *Ibid.*, p.70.

A moral being, Darwin remarks later, is one who is capable of approving or disapproving his own past and future actions, and, he adds “we have no reason to suppose that any of the lower animals has this capacity.”⁵

So Darwin claims that the difference between human and animal minds is one of degree and not kind, and yet that morality is a unique human attribute. But I don’t think that Darwin is being inconsistent. The trouble is that the question whether human beings are in some qualitative way different from the other animals is somewhat ill-formed. What exactly makes a difference “qualitative” in this context? What Darwin meant when *he* asked it was whether we are different in some way that could not possibly be explained by the theory of natural selection, and it was *that*, primarily, that he was eager to deny. This emerges when he says:

If it be maintained that certain powers, such as self-consciousness, abstraction, &c. are peculiar to man, it may well be that these are the incidental results of other highly advanced intellectual faculties....⁶

Darwin had no objection to the idea of distinctively human attributes, provided they could be explained in terms of the interactions between other attributes whose presence in us could be explained through natural selection – attributes we share with the other animals, although we may have them to a higher degree. And so his

⁵ Ibid., pp. 88-89.

⁶ Ibid., p. 105

main effort, when he turned his attention to “the moral sense or conscience” was to give an explanation of that sense that would prove that:

...any animal whatever, endowed with well-marked social instincts, would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well-developed, or nearly as well-developed, as in man.⁷

The moral sense, Darwin believed, was a distinctively human attribute, but it arose from the interactions between two other evolved human properties in which we differ from the other animals only in degree – intelligence and the social instincts.

The details of Darwin’s proposed evolutionary explanation of the moral sense will not concern me in this lecture. But there are two aspects of his view that I want to take notice of here, for I think that both of them are correct.⁸ One is the view I have just described - that if we do identify certain attributes as distinctively human, it should still be possible to explain their presence in us in evolutionary, or more broadly speaking, naturalistic terms. Darwin looks for this as an empirical conclusion that will support the theory of human evolution, but since I take human evolution as a given, I will treat it as a methodological desideratum instead. The other is the

⁷ Ibid. pp. 71-72

⁸ Darwin’s theory turns on a distinction between two kinds of instincts. Appetitive instincts are sharp, urgent, and occasional, demanding immediate satisfaction. Social instincts are milder but more constant. Appetites are urgent in a way that is disproportionate to the sense of satisfaction that gratifying them produces, so that we often feel, after satisfying them, that it was not quite worth it. Suppose a female animal leaves her offspring unattended in order to satisfy the urge to feed or mate, and the offspring is killed. Once her intellect is advanced enough that she can remember what she does, it will seem to her that it was not worth it, and she will regret it. The ‘ought’ is born from the sense of regret.

substantive view that what most clearly distinguishes human beings from the other animals is what Darwin, like the philosophers of his day, called the moral sense: the fact that only human beings are under the dominion of the governing force designated by Darwin's "short but imperious word *ought*." Why this should be so, and what some of its consequences are, will be my topic in this lecture.

II. Levels of Intentionality

In my last lecture I described the metaphysical difference that according to Aristotle sets animals apart from other living things. An animal, according to Aristotle, is characterized by agency in a very basic sense: locomotion guided by perception. An animal forms a representation of her environment by means of perception, and that representation, together with her instinctive evaluative attitudes, enables her to find what she needs and to avoid predators and dangers so that she can survive and reproduce. In this sense, an animal is essentially an organism that achieves her own well-functioning through the power of action.

Now just as an animal is a special kind of organism, an agent, a human being is a special kind of animal: according to Aristotle, a rational animal, and so, a rational agent. This is, of course, another of the standard views about what distinguishes human beings from the other animals – we have the power of reason - although like the other views, it is also often denied. Darwin, to use him as my exemplar again, remarks almost dismissively that:

Few persons any longer dispute that animals possess some power of reasoning. Animals may constantly be seen to pause, deliberate, and resolve. And it is a significant fact that the more the habits of any particular animal are studied by a naturalist, the more he attributes to reason and the less to unlearned instincts.⁹

But Darwin here contrasts the effects of reason to those of unlearned instincts, and I think that we can question whether that is the right contrast, or the only contrast, that we can draw when we think about the possible guides to action. Reason, on one way of understanding the term, amounts to more than the capacity to learn and to guide one's actions accordingly, and in this sense, reason is a distinctively human power. Or anyway, so I will argue – and in this I will be following the work of Immanuel Kant.

Moral standards are standards that govern action, so if it is true that only human actions have a moral character, and it is true that only human actions are governed by reason, it seems plausible that these two properties should be associated. That is, there must be something quite distinctive about human action – something that makes our actions different from those actions of the other animals – which in turn explains why human actions, and those alone, are both rational and subject to moral governance. It is this idea – that there is something distinctive about the character of human action - that I want to pursue here.

⁹ Descent, p. 46

The idea of acting for a reason is often identified with the idea of acting intentionally, or on purpose. The agent's intention or purpose supposedly provides the reason for what he does.¹⁰ So one possible way to explain what is distinctive about human action, would be to claim that only human beings do things intentionally or on purpose. Now I think that this hypothesis is pretty plainly wrong, but I do think that by examining the concept of intentional action, we can uncover what is distinctive about the human way of acting. So, I want to take a closer look at the concept of acting intentionally or on purpose. And this will also enable me to pay off a philosophical debt I have left unpaid up until now. For I have been freely characterizing non-human animals as agents, but there are philosophers and scientists who would deny that non-human animals, or at least the more primitive non-human animals, are agents at all. According to these philosophers and scientists – we might call them the skeptics about animal agency – animals, or at least many animals, do not really act, but instead simply *react* in mechanical ways to certain stimuli – either instinctively, or as a result of conditioning. These ideas are predicated upon a theory that places intentional actions done for a consciously-held purpose or a reason on one side, and mechanical responses to stimuli on the other, with little or nothing in between. In my view, what those who hold these views fail to recognize is that the concept of intentional action does not mark off a single phenomenon, but a number of things that can be ranged on a scale.

¹⁰ See my "Acting for a Reason"

At the bottom of the scale, there is the idea of intentionally or functionally describable movement. The concept of intention in this form applies to any object whatever that has some sort of functional organization, including not only human beings and animals but also plants and machines. In fact, according to the metaphysics of Aristotle, as I explained last time, it applies to almost anything we can identify as an *object* at all and that moves. Within the economy of a functionally organized object, certain movements can be described as having certain purposes. The heart beats in order to pump the blood, the alarm rings in order to wake you up, the plant's leaves reach out towards the sun in order to collect its rays. So we say, and there is no implication that the purposes served by these movements are before the minds of the objects that move, or even necessarily before the minds of someone who created those objects. Attributing purposes to these movements just reflects the fact that we view the object as functionally organized.

In the case of living things, especially animals, including the so-called "lower" animals, some of these purposive or intentional movements are guided by the animal's perception. A fish swims upwards towards a surface disturbance that may mean an insect; a cockroach runs under cover in response to your movements as you try to swat him with the newspaper; a spider crawls towards the vibrating struggles of the moth that is caught in the middle of her web. Here we begin to be tempted to use the language of *action* – to say that the fish and the spider are seeking food, and the cockroach safety – and it is clear enough why: when an animal's movements are

guided by her perceptions, they are under the control of her mind, and when they are under the control of her mind, we are tempted to say that they are under the animal's own control. And this, after all, is what makes the difference between an action and a mere movement – that an action can be attributed to the agent, that it is done under the agent's own control.¹¹ At this level, should we say that the animal acts intentionally, or on purpose? It depends how you understand the question. The animal is directing her movements, in the sense that her movements are flexibly responsive to perceived features of her environment – that is, they are governed by her mind. And her movements are intentional movements – *the movements* have a purpose. In that sense the animal acts with a purpose, but at this stage there is no need to say that this purpose is somehow before the animal's mind.

The contrasting view would withhold the characterization of the action as intentional unless the animal was guided by, say, a conscious desire to implement the purpose. But the argument here cannot be merely about the conditions under which we may correctly apply the words “intentional action.” My point is just that there is such a thing, or so we believe, as mind-guided intentional movements that do not involve the conscious entertainment of purposes. The movements of insects are probably like this. Yet this kind of movement is in a certain way continuous with

¹¹ This is a bit mysterious, actually. Think of a mosquito. Why should a movement under the control of his mind be any more attributable to *him* than a movement attributable to one of his other organs? We are certainly not more tempted to hold him responsible for it! We are autonomous in virtue of properties of our minds, so we identify ourselves with our minds when we take ourselves to be the authors of our actions. Therefore we take lower animals to be the authors of their actions when their movements are attributable to their minds. But of course it helps that mind-directed movements exhibit the kind of flexibility of response that we take to differentiate action from mechanical motion.

more explicitly purposive action. For it is important to acknowledge that when we try to look at the situation from the animal's own point of view, when we ask ourselves what exactly it is that the animal perceives that determines her movements, it is almost irresistible to describe it purposively. Why does the spider go towards the moth caught in her web unless there is some sense in which the spider perceives the moth as food and therefore some sense in which she is trying to get food? Or to put the point another way, if a spider responds to a certain perceptual stimulus by attempting to eat, why should we be shy about saying that she sees or smells the thing as food, and that her intention is to eat it? We need not take this to mean that she entertains thoughts about food, or that she reasons that it must be food and that therefore she should go and get it. What we should suppose instead is that, precisely because she lacks the intellectual power to interpret what she sees, her instincts structure the perceptual world for her in such a way that no act of interpretation is needed. She sees (or smells or hears) the things she encounters in her environment as food, as mates, as threats, and so attractive or aversive accordingly. These evaluative perceptions determine her movements directly. How else could perception, at this primitive stage of intellectual development, do an animal any good? If the only way to benefit from perception was to think about what you perceive, perception would only have evolved in intelligent beings, if it could have evolved at all.

On the other hand, once we are dealing with an intelligent animal, who can think about what she perceives, there is no reason *not* to suppose that she

consciously recognizes a certain object *as food* and that in that sense her purpose is to get something to eat. I don't mean that she is aware, at this stage, of having a purpose – I mean that she thinks of the object in a way that has purposiveness built into it. To get from what we suppose the spider does to what we suppose an intelligent animal does – roughly, from seeing something as attractive in a certain way that causes you to eat, to seeing it *as food* - you need only add a certain sharpening of the cognitive powers, perhaps an ability to categorize things according to purposive types, like *food*, or *predator*. If it weren't already obvious enough from the general character of their actions, we know from the language-trained animals that they at least can do this. And some other animal communication systems also involve purposive categories – different alarm calls for different types of predators, for example. So I see no reason why we should not suppose that there is a gradual continuum between whatever is going on when a spider's perceptions direct it towards the moth caught in its web and straightforward cognitive awareness of something as *food*, or *danger*, or a mate. And when such cognitive awareness is in place – once you have the purposive category - presumably the possibility of learning from experience about *how* to get food and avoid danger is greatly enhanced. You can always learn from experience by conditioning, of course, but when you are aware of your purpose as an object of a certain type, you can presumably also begin to learn from experience by thinking and remembering.

But even if there is a gradual continuum, it seems right to say that an animal that has a conception of his object as a certain purposive type is also exerting a greater degree of conscious control over his own movements than, say, the spider, and is therefore in a deeper sense an agent. Even if he entertains no thoughts about himself, there is a sense in which such an animal *knows what he is doing*, and we don't think that insects and other primitive animal do that. One thing that reveals this is that at this stage we feel that there is room for disagreement about what the proper intentional description of a particular action is, for it is at this stage that we become committed to *keying* the intentional description of the action to what is going on from the agent's own point of view.¹² We are less inclined to do this at the earlier stage: when we do describe the spider, in intentional terms, as "trying to get food," we don't care whether that's what the spider thinks she's doing. When talking about a cognitively primitive animal like a spider, in fact, it is natural for our intentional descriptions of her movements and our biological explanations of them to run together in this way. In fact scientific writers are inclined to speak rather freely about such animals trying to make sure their genes make it into future generations, and things of that kind. Of course no one thinks animals think about their genes. It is not really that there is a reason to talk that way, but rather that there is no reason not to. But once we suppose an animal has reached the stage where he can consciously entertain thoughts about the object he is pursuing, we are committed to giving an

¹² Freudian slips pose a problem for the claim I just made.

intentional description of the action that captures something about the way it seems to the animal himself. It's because at this level we key intentional description to the animal's own perspective, that at this level it makes sense to ask, as scientists sometimes do, what the animal is *really* trying to do: whether his conduct is altruistic or he is just seeking allies, perhaps. So the entertainment of a conscious thoughts about the object of your action represents a deeper way in which an action may be said to be "intentional" than we find in the case of the intentional movements guided by a more primitive mind like that of a spider.

But some philosophers do not believe that this is the deepest level of intentionality. At the level of intentionality I have just been describing, the animal is aware of his object under a purposive description, and perhaps he even thinks about how to pursue it. If he is an intelligent animal, he may begin to exhibit the problem-solving ability that I think impressed Darwin, and that has led some philosophers and scientists to attribute reason – especially instrumental reason – to the other animals. But a non-human animal does not choose *to* pursue a certain object. Rather, his purposes are given to him by his instincts and his affective states: his emotions and his instinctual or learned desires. This is true, I believe, even when an animal in a sense chooses between the pursuit of two different objects. Say for instance a subordinate male chimpanzee wants to mate with a certain female but a dominant male is approaching and the subordinate chimpanzee also wants to avoid a fight. (I apologize for this somewhat raffish example, but this is the sort of choice that

the practical lives of animals consists in.) In this case, I claim, the choice is made for him by the strength of his affective states. Having been thrashed by the dominant male before, he fears the dominant male more strongly than he desires to mate. The end that the chimp decides to pursue is determined for him by his desires and emotions, where those in turn are the product of instinct and learning working together. His contribution, so to speak, is only to direct his actions towards the purpose that his desires and instincts have set for him. Later, I will explain why I am so confident that this must be so.

Kant believed that a still deeper level of intentionality is possible than this. Besides asking ourselves how to go about getting an end that is set for us by our desires and instincts, we can ask ourselves whether we *should* pursue a certain end at all – whether it is *worth* pursuing. And we can decide whether to adopt the end or not based on that assessment. We choose not only the means to ends set by our nature, but also the ends themselves. What makes this possible, I believe, is the form that human self-consciousness takes.

III. The Role of Self-Consciousness

It is sometimes said that human beings are the only animals who are self-conscious. Animals are aware of the world but not of themselves.¹³ But actually the issue is more complicated than that, for self-consciousness like other biological

¹³ Among other things by me in Sources. Whoops.

attributes comes in degrees and takes many different forms. One form of self-consciousness is revealed by the famous mirror test used in animal studies. In the mirror test, a scientist paints a red spot on an animal's body and then puts her in front of a mirror. Given certain experimental controls, if the animal eventually reaches for the spot on her body and tries to rub it off, or look away from the mirror towards the relevant location on her body, that is taken as evidence that the animal recognizes herself in the mirror, and is curious about what has happened to her body. Apes, dolphins, and elephants have passed the mirror test, in some cases moving on to using the mirror to examine parts of their bodies that they can't normally see, apparently with great interest. Other animals never recognize themselves, and some instead keep offering to fight with the image in the mirror, or to engage in some other sort of social behavior with it. An animal that passes the mirror test seems to recognize the animal in the mirror as "me" and therefore, it is thought, must have a concept of "me." I will come back to the mirror test later on.

But I think it can be argued that animals who can't pass the mirror test have rudimentary forms of self-consciousness. You have self-consciousness if you know that one of the things in your world is *you*. And a tiger who stands downwind of her intended prey is not merely aware of her prey - she is also locating *herself* with respect to her prey in physical space, and that suggests a rudimentary form of self-consciousness. A social animal who makes gestures of submission when a more dominant animal enters the scene is locating himself in social space, and that too

suggests a form of self-consciousness. Knowing how you are related to others involves something more than simply knowing about them.

Parallel to these abilities would be a capacity to locate yourself in mental space, to locate yourself with respect to your own experiences, thoughts, emotions, beliefs, and desires. This is what we more commonly think of as self-consciousness, a reflective awareness of our own awareness, so to speak. Do the other animals have this ability to locate themselves in subjective, mental space? Scientists have sometimes taken the mirror test to establish this kind of self-consciousness, but it is a little bit difficult to articulate exactly why. The animal grasps the relation between the image in the mirror and her own body. But in so doing, she seems to show that she grasps the relationship between herself and her own body. But what exactly does that mean? She grasps the relation between two things, a certain physical body and – well, what? – we can say “and herself” – but what exactly is the “herself” that she identifies with that body? Perhaps the idea is that what she identifies as *herself* is the self that is the subject of her own experiences, of which she must then have some awareness. That is, she must be aware not just of pain but that *she* feels pain, or not just of the smell of food but that *she* smells the food. And it is that “she,” the subject of those experiences, that she correctly identifies with the body she sees in the mirror. Some such idea must be behind the thought that the mirror test reveals an inner self-consciousness.

Interestingly, however, even if this is right, it does not yet seem to show that the animal must be aware of herself as the subject of her *attitudes* – her beliefs, emotions, and desires. And this suggests a further division within this form of self-consciousness. An animal might be aware of her experiences and of herself as the subject of those experiences, and yet her attitudes might be invisible to her, because they are a lens *through* which she sees the world, rather than being parts of the world that she sees.¹⁴ In fact this possibility has been implicit in my description of the intentionality of animal actions. Earlier I suggested that simple animal agents must directly perceive things as aversive or attractive in ways that evoke certain responses. Animals with more conceptual equipment might perceive things as, say, *food* or *danger*. I suggested that perception must work this way, loading the correct response into the perception, so to speak, because the capacity for perception could hardly be useful to creatures of primitive intellect if it did not supply automatic cues about how to respond to what is perceived. You don't need to know of yourself that you want food in order to respond to it correctly: you only need to perceive it as food.

But of course more intelligent animals might also be aware of their own attitudes. Some of the language-trained animals seem able to express the idea “I want” – Koko the gorilla and Alex the African gray parrot, two famous language-

¹⁴ It's easier to understand what I mean here when you are thinking about practical, evaluative attitudes. It sounds odd to think of beliefs as a lens through which we see the world. But they are, in the sense that an animal could be moved by one belief to take up another without having any awareness of making an inference. Unlike a person, a non-human animal can think “X” without commitment to “I believe X” or “X is true,” because he has no commitments of that sort.

trained animals, can both do this - so perhaps they have the ability to think about their own mental states.¹⁵ But of course it is also possible that they have just learned that such utterances will produce the desired effect, by a sort of conditioning. Some scientists have also pointed to cases of deception to suggest that some animals are aware of the beliefs of others, and therefore, presumably, of their own. The evidence on these questions is, I think, inconclusive.

Human beings are aware of our attitudes.¹⁶ We know of ourselves that we want certain things, fear certain things, love certain things, and so on. But we are also aware of something else – we are aware of the potential influence of our attitudes on what we decide to do. What I have in mind is this: a non-human animal may be conscious of the object of his fear or desire, and he may be conscious of it as *fearful* or *desirable*, and so as something to be avoided or to be sought. When he acts accordingly, we even say that that is the *reason* for his action: that he was afraid, or that he saw it as a threat. What we mean is that his attitude caused him to do what he did. But a human being is conscious *that* she fears or desires the object, and, in addition, *that* she is inclined to act in a certain way as a result. She does not just think about the object that she fears or even about its fearfulness but about her fears and desires themselves, and their influence in urging her towards action. In evolutionary

¹⁵ See the Alex Studies, pp. 197-208, for Irene Pepperberg's account of teaching Alex to use "wants" and her own conclusions about what exactly he learned when he learned it. Koko has a sign for "wants."

¹⁶ Strictly speaking, I do not think the grounds of our actions count as *reasons* for them until we have endorsed them in reflection.

terms, being conscious of the operation of your attitudes on your inclinations to act is surely just a step away from being conscious of your attitudes themselves. Yet I believe that the effect of this little step is momentous. For once you are aware that you are being moved in a certain way, you are in a position to form an attitude towards the fact that you are being moved in that way. And in fact, there is a way in which you *must* form an attitude towards the way you are inclined to act, because you now must decide whether to go along with that inclination or not. As I have put it elsewhere, you now have a certain reflective distance from the impulse that is motivating you, and you are in a position to ask yourself “but *should* I be moved in that way? Wanting that end inclines me to do that act, but does it really give me a reason to do that act?” You are now in a position to raise a *normative* question, a question about whether the action you find yourself inclined to perform is justified.

I believe that, in general, this form of self-consciousness – consciousness of the grounds of our beliefs and actions – is the source of reason, a capacity that I think is distinct from intelligence. Intelligence is the ability to learn about the world, to learn from experience, to make new connections of cause and effect, and put that knowledge to work in pursuing your ends. Reason by contrast looks inward, and focuses on the connections between mental states and activities: whether our actions are justified by our motives or our inferences are justified by our beliefs. I think we could say things about the beliefs of intelligent non-human animals that parallel everything I have said about their actions. Non-human animals may have beliefs and

may arrive at those beliefs under the influence of evidence, but it is a further step to be the sort of animal that can ask oneself whether the evidence really justifies the belief, and can adjust one's conclusions accordingly.¹⁷ The structure of human self-consciousness is what makes human beliefs and actions the subject of rational norms.

So, when we are aware of the possible influence of our attitudes on our actions, we can take control of that influence: we can – and must - ask whether we should go along with it or not. But how do we answer these normative questions – how do we decide whether we should go along with an inclination to act in a certain way? In other words, where do we find the norms? Obviously, I can't answer that question in general, but I am going to try to sketch an account of the way it works in the case of the norm that is my topic today: the norm of right action.

IV. The Origin of the Right and Human Intentionality

A consequence of the nature of human self-consciousness is that when we decide on an action, we are making a different kind of decision than an animal does – the object of our decision is different. An animal, finding himself strongly drawn to a certain end, is prompted to do an action which he knows instinctively or has learned will bring about that end. But the object of a human choice includes both the act and the end: what we choose is “to do this for the sake of that.” A tiger is hungry so she chooses to hunt. “Hunting” is the object of her choice. You are hungry so you

¹⁷ I pursue this argument in *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.

choose to go to a restaurant. “Going to the restaurant in order to get something to eat” is the object of your choice. And this makes an important difference: since you think you can choose for or against “going to the restaurant in order to get something to eat” you think you can choose for or against getting something to eat. You don’t think of yourself as governed, or determined, by your desire to eat, the way another animal is. In that sense, you think of yourself as free.

Now ordinarily, when we want to do something, we suppose that the thing we want to do would be good for us.¹⁸ That is natural, for in terms of evolution, that is what desires are for: they guide us and the other animals towards the achievement of our good, our own well-functioning. But since we think of ourselves as free, we don’t think that we *have* to do something just because it conduces to our good. And this brings me to a question I left undiscussed in my last lecture. Many people think that the idea that something is good is a normative idea. That is, they think that if we say something is good we are saying there is reason to promote it, or bring it about, or not to hinder it, or to respect it, or whatever. We use the term that way sometimes, but not always, and I was not using the term that way in my lecture on Monday. For one thing, as I argued last time, every thing that is good is good relationally, good for someone or other. But it does not follow automatically from the fact that something

¹⁸The route I take in the text is the Formula of Humanity route. But there is also a route from the idea that we choose an “act for the sake of an end” to the Formula of Universal Law, via the idea that the Formula of Universal Law is a constitutive principle of action – or rather, human action.

is good for someone or something that it is good in what we might call the normative sense, or good absolutely.¹⁹ So when we do decide to do something because we want to do it, because we think it would be good for us to do it, we are according ourselves a certain standing. We are claiming that what is *good for us*, our final good, is good in the normative sense, good absolutely. We are claiming that our own good is the source of reasons, that we are the source of reasons. We are claiming the status of what Kant called ends-in-themselves.

Furthermore, when you have the thought that you are the source of reasons, that your good is the normative good, you have to be identifying yourself in some particular way. Just as the chimp who looks in the mirror and thinks “that’s me” has to mean something by “me” so you have to mean something by “me” when you think “my good is a thing worth pursuing.” Kant supposed that what you mean by “me” is “this human being, the human being who I am.” I believe that he thought that because it is in our capacity as human beings that we think these thoughts – only a human being can ask and answer the question whether there is a reason to promote his own good. In my next lecture, I am going to raise some questions about Kant’s view on that point, but for now I will stick with it. When you think that your own good, *as the good of a human being*, is good in the normative sense, you commit

¹⁹ There are actually two reasons for this. The final good of an assassin, considered as such, is to have a steady aim and a ruthless heart, to be assigned unwary victims and to live in a city where the police are corrupt. Is there a reason to promote it? In the end, I think my theory commits me to the view that the only reason an assassin’s final good isn’t good is because it’s bad (inconsistent with the value of humanity). A different case, more relevant here, is that of a final good that’s just indifferent, a matter of no concern. One might think this of the final good of an artifact or perhaps even a plant, if such things do have final goods. Anyway, that is more the contrast I’m looking for here.

yourself to the idea that the good of human beings in general is good in the normative sense. You commit yourself to the value of human beings.²⁰ For, as I have said before, what it is to value a sensible being is to value its good. And in making that commitment, in deciding that human beings are ends-in-themselves, you bring forth the standard of the right. The standard of the right demands that we treat all human beings in accordance with their value.

Now, I haven't said anything about what the content of this standard is, and won't be doing so in these lectures. But just to avert possible misunderstandings, I want to emphasize here that good for human beings – which I will be discussing next time – is a complex thing that importantly includes our own autonomy. And I also want to emphasize that even when we decide that the good of human beings is good in the normative sense, it remains a relative value in the sense that it cannot be added across the boundaries between persons. So the resulting ethic will not be a utilitarian one, but will involve respect for rights and freedom. The argument I've just given is a redescription of Kant's argument for his formula of humanity, the principle that tells us to treat every human being as an end in itself. Working out the details of how we do that is a task for another time. For our purposes here, the important thing is that once you have the standard of the right, you have a way of answering the question whether you should do this action for the sake of this end. You can ask whether it is

²⁰ This comes close to claiming that the Formula of Humanity is a constitutive standard of action. It falls a little short because as far as I can see we don't have to act on our desires, even when morality allows it. We just do.

consistent with the standard of the right – whether it is justified. And that means human actions exhibit a deeper level of intentionality than animal actions. Because it means that even if you do judge the action you propose to yourself to be justified and act upon it, you are acting not merely from your desire but from your *judgment* that the action is justified: the end is indeed the kind of thing that provides a good reason for doing that sort of act. And in acting on a judgment about what counts as a good reason for what, you are acting on principle.

Why do I say this represents a deeper level of intentionality? In the first place, an agent who is capable of this form of assessment is capable of rejecting an action along with its purpose, not because there is something else she wants (or fears) even more – like the chimp who fears a beating more than he wants to mate - but simply because she judges that doing that sort of act for that sort of purpose is *wrong*. As Kant put it, you are motivated by pure practical reason, for you can be motivated directly by the judgment that you ought to do something. In a famous passage in the *Critique of Practical Reason*, Kant argued that we are *capable* of setting aside even our most urgent natural desires – the desire to preserve our own lives and to secure the welfare of our loved ones – in order to avoid performing a wrong action. Kant gives the example of a man who is ordered by his king, on pain of death, to bear false witness against an innocent person whom the king wants to get rid of. Would you impugn the honor of an innocent person in order to save your life? While no one can say for sure how he would act in such a situation, Kant argues, each of us must admit

to himself that he is capable of doing the right thing.²¹ We can be motivated to rise above even our strongest desires, like the desire for life, in order to do the right thing.

I want to emphasize one thing that is a little complicated about the argument I've just given. Earlier I said that we regard ourselves as free to reject an action we propose to ourselves – that is, we regard ourselves as free to reject a possible “this act for the sake of this end” when we consider whether it is justified. I did not mean to imply by that we have free will in some extravagant sense. All I meant was that that whole formula – “this act for the sake of this end” is the object about which we are choosing, and we can choose for or against it. But to the extent that we really are free in a way the other animals are not – and I don't think it is an extravagant sense - it is partly *because* we have the standard.²² The standard gives us a motive for rejecting the purposes our nature proposes to us. But if we are capable of setting aside our ends when we cannot pursue them by any decent means, then there is also a sense in which when we *do* decide to pursue an end, we can be seen as having *adopted* that end. Our ends may be suggested to us by our desires and emotions, but they are not determined for us by our affective states, for if we had judged it wrong to pursue them, we could have laid them aside. Since we choose not only the means to our ends

²¹ *The Critique of Practical Reason* (1788), translated by Mary Gregor. Cambridge: Cambridge University Press, 1997, p. 27.

²² I am not sure if I mean this. It's like the *ratio essendi, ratio cognoscendi* business, but I've put it more strongly than that.

but also the ends themselves, this is intentionality at a deeper level. For we exert a deeper level of control over our movements when we choose our ends as well as the means to them than that exhibited by an animal who pursues ends that are given to her by her affective states, even if she pursues them consciously and intelligently. Morality, and this deeper level of intentionality, are capacities that go hand in hand. Kant called this kind of freedom “autonomy” because in our actions we are governed by principles that, in a sense, we have legislated ourselves. We have decided to accord value to humanity, and to govern ourselves accordingly. Morality is the human capacity for normative self-government. Human beings are free and moral beings, autonomous beings, but this is not something inexplicable and mysterious. It all arises from the form of human self-consciousness.

V. Of the Reason of Animals (to borrow a chapter title from Hume)

Now I want to return to the claims I made about the reasoning capacities of the other animals. Many people would agree that only human beings are moral animals, but they would deny that the other animals lack reason. Although animals are not governed by moral standards, it is well known that intelligent animals often solve problems about how to achieve desirable ends. The literature on animals these days is full of new discoveries about animals using and even making tools. And animals, like human beings, generally get into more trouble when they are children than they do when they are adults, suggesting that they learn to think more about

consequences and so become more prudent. Many philosophers have thought that practical reason demands that we take the means to our ends, and that, other things equal, we choose what will be better for ourselves rather than what will be worse. That is, they have thought that taking the means, and prudence, are rational requirements. Do animals have this kind of reason?

Earlier I suggested that even when a non-human animal is torn between two desirable ends, he does not count as *choosing* whichever of the two ends he eventually goes for, at least in the sense of “choosing an end” that I just explained. He cannot lay aside an end simply because he thinks they ought to. Instead the choice is made for him by the strength of his affective states. Notice that this does not mean that there is no role in non-human thought for something that we would ordinarily call reasoning – for thinking about the possible consequences of each course of action, and by that means giving direction to your affective states. It is helpful in this connection to recall the way the philosopher Thomas Hobbes thought of deliberation. Hobbes says:

When in the mind of man appetites and aversions, hopes and fears, concerning one and the same thing arise alternately, and diverse good and evil consequences of doing or omitting the thing propounded come successively into our thoughts, so that sometimes, we have an appetite to it, sometimes an aversion from it, sometimes hope to be able to do it, sometimes despair or fear to attempt it, the whole sum of desires

aversions, hopes and fears, continued till the thing be done or thought impossible, is that we call DELIBERATION.²³

In Hobbes's conception, deliberation is merely imagining what it would be like to do something and envisioning the various consequences of doing it, with the result that one's affective attitude towards doing it changes accordingly to the consequences envisioned. We can certainly imagine the chimp in my example "reasoning" in this sense: thinking of mating, and wanting to, then thinking of the thrashing he will receive from the dominant male, and not wanting to, until one of these desires comes uppermost, and he acts. And while he is doing it, as Darwin says, he would "be seen to pause, deliberate, and resolve." If, being an older and wiser chimp, he resolves to back away and give up on the mating, we might even say, in a colloquial way, that he has decided that the end is not worth pursuing. So does this chimp exercise rational prudence?

I think that the answer is no, or at least, not in the same sense as we do. And this brings out something important about these forms of practical reason as they exist in human beings. One way to bring out the difference I have in mind is to notice that just as human beings can be guilty of violating the moral principle, we can violate these other rational principles – we can be guilty of weakness of will, and fail to choose the better course. Now if you believe that there is such a thing as weakness of will, you must believe there is a difference between *deciding* that a certain end isn't

²³ Leviathan, p. 33, Hackett/Curley edition.

worth pursuing, and failing to pursue it because you are overcome by pain or fear or idleness even though you do still think it is worth pursuing. There is a difference, for instance, between the soldier who flees the battlefield having *decided* that the defense of his country isn't worth risking death or injury, and the soldier who believes wholeheartedly that he should be prepared to make the sacrifice, but who is overcome with terror when confronted with the smoke, screams, explosions and chaos of the battlefield. There is a difference, again, between giving away state secrets under the influence of torture, and *deciding* that under the circumstances of torture the secrets aren't worth keeping. It would be horribly unkind to describe someone who gave state secrets away under torture as having "decided" that they weren't worth keeping. Deciding that an end is or isn't worth pursuing is not just a matter of finding that on reflection it is either more or less attractive, because even after reflection we may be most strongly attracted towards the end we have decided is not worth pursuing, or away from the one we have decided worth pursuing.

What this shows, I believe, is that in human beings, all of these forms of reasoning involved being governed by principles. And that is a different thing than merely having your attitudes shaped by taking thought. To see this, let's stay with prudence. It's quite true that thinking about the ways a certain option will be better for me in the long run can cause me to desire to take that option. But if the rationality of prudence were *based* on that fact – if it were based on the fact that the better option seems desirable to me once I think about it – then if I didn't find that

option desirable after thinking about it, I would *not* be irrational if I failed to pursue it. The best choice, for me, is not the choice I desire most, even after taking thought, but the one I judge to be best. That is true even if it turns out that the content of that judgment refers to my desires – for instance, to what I will desire later on. If there is a rational principle of prudence, what it tells me to do is to act on that judgment. It tells me that I ought to act on my judgment about what is best in preference to what I find most attractive, not merely that I will find it attractive if I think hard enough about it. (That may be true, but that is not what the rational principle of prudence tells me.) And that means that prudence, no less than morality, is a struggle for the victory of principle over what is locally more attractive. And there is no such struggle in the life of a non-human animal.

This point may be concealed by the fact that sometimes, when we *are* struggling to be prudent, we engage in something that looks a lot like Hobbesian deliberation – we try to dwell on the advantages of what we know to be the better option. So for instance, I may try to steal myself to going to the dentist by vividly imagining the pain of the toothache I will inevitably have if I don't. This makes it seem as if it is merely the natural attractiveness of the better end (say, of not having the toothache) that motivates me, rather than my judgment that I ought to go to the dentist. But there are two differences between this kind of exercise and genuine Hobbesian deliberation. The first is that this kind of exercise is *directed* – when I perform it, I already know which option is best – I already know what I ought to do.

The second difference, as I have already emphasized, is that the correctness of my judgment doesn't depend on the outcome of this imaginative exercise. If the imaginative exercise succeeds, I may rid myself of the temptation to be imprudent, and so solve my problem. But if the imaginative exercise fails, I will still be guilty of weakness of will if I fail to go to the dentist. That is because I am capable of choosing the better course, not because it is the most attractive to me, not even because it is more attractive on reflection, but because it is the better course. That is to say, it is because the thought that I *ought to do* what will be better for me can produce in me a motive for doing it. And this brings out something that many philosophers have failed to notice. Prudence requires the possibility that we can be motivated by pure practical reason, in exactly the same way that morality does. It is only *what* it tells us to do that is different.²⁴

I believe that the same is true even in the case of instrumental reason, although here there is a further complication: the requirement of instrumental reason cannot stand alone. I cannot be guilty of weakness of the will about the requirement to take the means to my end if my end is simply the option that is most attractive to me, for if an option does not attract me enough to motivate me to take the means to it, then *ipso facto* it is not the most attractive option for me. No matter how much I want to get in shape, if it does not motivate me to exercise, then I prefer the combination of sloth with unfitness to the combination of exercise with fitness.

²⁴Myth of Egoism

And if my end is simply whatever I prefer, then in pursuing the course of sloth and unfitness, I am pursuing my end, not failing to take the means to my end. So the principle of instrumental reason imposes no requirement on us, unless there are also principles, like those of morality or prudence, requiring us to have certain ends – ends that might not be, on a given occasion, the ones we most strongly want to pursue. Once those are in place, I can be guilty of weakness of will, if I fail to take the means to an end I have determined, for whatever reason, that I ought to pursue. And this is not because the thought that this action will secure the end has makes performing the action attractive to me – for if I find myself reluctant to take the necessary means, then of course it hasn't. Instead, it is because the *thought* that I ought to take the means to an end I regard as good *itself* serves as the motive for taking those means.²⁵

I've been drawing a distinction between two kinds of motivation. An intelligent animal might be drawn to perform an action because he finds it attractive once he sees that it will get him what he wants. And he might be drawn to perform an action, or not to, because he finds it attractive, or unattractive, once he thinks about some consequences of that action that will ensue later on – although probably not much later on. These motives mimic instrumental reasoning and prudence, but they are not the same thing. To be motivated by instrumental reason and prudence is to be motivated by the recognition of the normative force of certain rational

²⁵That was the hey presto version of "The Normativity of Instrumental Reason."

requirements or principles – the principle of taking the means to your ends, and the principle of choosing the better course – if indeed those are rational requirements – I haven't tried to argue for that conclusion here.²⁶ Since all rational action involves acting on principles, all rational action requires the same level of intentionality that morality does: the ability to be motivated to do something by the thought that you ought to do it.

To put the same point another way, if the arguments I have given here are correct, all genuine rational principles are what Kant called categorical imperatives, and our capacity to act on them therefore depends on our general capacity to act on categorical imperatives. That – to get back to an earlier question - is why I am fairly confident that non-human animals don't do more than follow their strongest impulse, even in cases where they are choosing the means to ends or choosing between two options. I don't believe that non-human animals can reason in any of these ways because I think that the principles of practical reason stand or fall together. If the other animals could be governed by reason at all, they would also be moral animals.

VI. Conclusion

Let me make it clear what I am proposing. I think there is something really new and different about human beings, and about human life – something that sets

²⁶ See the Myth of Egoism for some doubts about the rational principle of prudence.

us apart from the other animals. This new thing is the human capacity for normative self-government – our capacity to determine our beliefs and actions in accordance with normative principles. This capacity is not mysterious or non-natural, but is rather the inevitable result of a certain form of self-consciousness: consciousness of the grounds of our beliefs and actions as *grounds*. In the case of action this form of self-consciousness makes possible a deeper level of intentionality than we see exhibited in the actions of the other animals: we can choose not just to do actions that promote ends given by nature, but to-do-a-certain-action-for-the-sake-of-a-certain-end. We are therefore hold ourselves responsible for performing only actions whose ends justify us in taking whatever means we do take. Both the self-consciousness that makes this form of intentionality possible, and the form of intentionality itself, are, in one way, clearly points on a continuum with earlier forms of self-consciousness and intentionality. There is nothing to prevent us from supposing that these are properties that evolved in a natural way. But the result brought something quite new and different to the world: human beings live under the dominance of what Darwin called that short but imperious word “ought.” We are moral and rational animals.