

Thoughtful Brutes

Jonathan Bennett

APA Eastern Division Presidential Address, in *APA Proceedings* 62 (1988), pp. 197–210.

1. Introduction

I am interested in what main differences there are between *Homo sapiens* and other known terrestrial species, or (for short) between man and beast. We have a sense that we differ vastly from all the rest in some respect that is mental rather than grossly physical, but we are not agreed on what respect it is. This is my topic today.

I shall bring in some work done in recent years by ethologists and animal psychologists. It is relevant less because it provides news about the beasts than because it spurs our thinking about what we already think or intuit about the beasts and how they differ from us. That is *really* my topic: although we don't agree in what we *say* about what mainly differentiates us from the other animals, I think we have the same *picture* of the difference, the same *sense* of what it is; and I want to know whether that shared picture, or intuition, can be parlayed into an agreed description. One proper study of humankind is ourselves; that includes our thoughts, which include our thoughts about what makes us special.

I shall go on talking as though the question were: What is the difference? But really it is: What difference do we already think is there? Or, anyway: What difference can be

known to obtain between us and them, on the basis of what we already know about us and about them?

I am interested in differences *of kind* rather than *of degree*. If the final story is merely that we are more intelligent than the beasts are, that means that our common picture of how we relate to them is false. Mortimer Adler, in his book *The Difference of Man and the Difference it Makes* argues for a fundamental intellectual difference of kind between humans and other animals. Having identified me in that book as an ally, on the strength of my book *Rationality*, he wrote to me about the matter. During the correspondence that followed, I wrote that while I agreed with his conclusion, I didn't hold it with such passion as he did, perhaps because I was sure that the difference between kind and degree is itself a difference of degree. Adler took me to have implied that there are no differences of kind, at which he protested strongly. But I implied no such thing. The difference between tall men and short men is one of degree, but there are tall men. I have a real topic; there are differences of kind, and it is a real question what if any of them obtain between humans and other animals.

My project is part of a larger one that some of us have been intermittently engaged in for some years, namely: trying

to help cognitive ethologists to interpret their findings. They are the people who study how nonhuman animals respond to their environments, in the hope of figuring out what is going on in the animals' minds. This is respectable though soft science, and some of its practitioners—wanting to be rigorous and disciplined—flatteringly look to philosophers for help. We haven't yet provided much, but we are trying. Help is certainly needed.

Someone might be interested in my project for moral reasons. Most of us treat one another somewhat differently from how we treat nonhuman animals; if we have a clear conscience about that, it must be based on some beliefs about how the other animals differ from ourselves. We think that they experience suffering, and most of us regard that as *relevant* to how they should be treated. So we need to be able to accompany the fact of their suffering with an 'On the other hand. . . ' that will make them weigh less than we do in our moral scales. A better understanding of what differences there are might bear on our morality of conduct towards other species—confirming it just as it stands, or refining it, or subverting it.

It would be subverted if the inquiry robbed us of our strongest reasons for behaving worse towards nonhuman animals than towards one another. This is a real possibility. If you are a meat-eater, and have a conscience, you engage in cognitive ethology at your own risk.

Another risk is that we shall characterize the morally relevant differences between humans and other animals in such a way that some defective humans fall on what we regard as the wrong side of the line. This was a matter of anxious concern to Leibniz in his great discussion of the topic in the *New Essays on Human Understanding*, but he had a way of keeping the worry under control. He held that our rational soul is what centrally marks us off from

the beasts, and that any offspring of humans—however behaviorally incompetent—may *have* a rational soul and suffer merely from some blockage to its being publically exercised. Even where this doesn't seem very plausible, Leibniz thought, it is better to play safe—better to baptise and care for something which really lacks a rational soul than to slaughter someone who really has one. But if the crucial differences are not 'hidden' or 'inner' as Leibniz thought them to be, that move won't work. Then some other way must be found for blocking the conclusion that offspring of humans who don't intellectually surpass the chimpanzees deserve only such moral respect as chimpanzees do. The best blocker I can think of is the view that those defective humans are entitled to special treatment from us because they are *ours*, they are *of us*, they are members of our clan. On this view, our special moral relation to them doesn't depend on their having some general feature that qualifies them; rather, it depends on their being related to us in a special non-moral way, namely by belonging to the same species. This is like the view that friendship creates moral relationships: what I owe to my friend depends less on what he is like than upon the sheer fact that he *is* my friend—the sheer fact of a shared past of a certain kind.

The status of nonhuman animals as moral patients has generated a whole literature; but I have to leave it now, because my present concerns are elsewhere.

2. Descartes's thoughtless brutes

Descartes offered a hard, sharp answer to my question. Beasts cannot think, he said, which for him meant that they are not the subjects of any mental states or processes at all. Why not? Because beasts provide no evidence of thought, since everything they do can be explained physicalistically, without bringing in the hypothesis of mind.

Suppose we have built a perfect *physical* replica of a mentally endowed human being, Descartes says. If the replica's behavior is not interfered with from outside the physical realm (that is, by God or by a mind) it will replicate only some of the behavior of its original, namely the part that is explicable purely through physics. Here is some of what it could not do:

It could never use words, or put together other signs, as we do in order to declare our thoughts to others. For we can certainly conceive of a machine so constructed that it utters words, and even utters words that correspond to bodily actions causing a change in its organs (e.g. if you touch it in one spot it asks what you want of it, if you touch it in another it cries out that you are hurting it, and so on). But it is not conceivable that such a machine should produce different arrangements of words so as to give an appropriately meaningful answer to whatever is in its presence, as the dullest of men can do. (Descartes, *Discourse on the Method* 5, AT 6.56f.)

On the other hand, Descartes held, a perfect physical replica of a sheep would behave in every way like a sheep.

That is a risky basis for a line between humans and other animals. Descartes's contemporary Arnauld warned him that he would have trouble convincing people that 'it can come about, without the assistance of any soul, that the light reflected from the body of a wolf onto the nerves of a sheep should move the minute fibres of the optic nerves, and that on reaching the brain this motion should spread the animal spirits throughout the nerves in the manner necessary to precipitate the sheep's flight'. (Antoine Arnauld, Fourth Objections to the *Meditations*, AT 7.205.)

Descartes certainly wasn't entitled to be sure about the sheep, but nor was it reasonable for his contemporaries to

be sure that he was wrong. Intuitions of incredibility were worthless—as Spinoza said a few years later—given how little was known of anatomy and physiology. We can more easily see what Spinoza saw, helped as we are by microscopic knowledge of the brain's complexity, and by a shift from a mechanical to a chemical understanding of neural processes.

Those developments, which make the sheep side of Descartes's contrast easier to believe, make the human side more dubious—increasing the likelihood that all human behavior, even speech, can be explained in physical terms. If we are to rest anything much on a conviction of a large difference of kind between ourselves and the rest, it needs a solid foundation than Descartes offers. Any materialist will say 'Amen to that!', but today's dualists should agree to it also. Even if there is a distinct realm of mentalistic facts, it is not credible that they plug gaps in physical causal chains.

Still, Descartes must be granted this much: We attribute mentality to others on the basis of their behavior, and no other basis looks even remotely plausible. In particular cases we might infer something about x's state of mind from data concerning x's neural state, but we would only connect a neural item with a mental item if each was associated with a relevant kind of behavior. Or again, facts about x's environment might convince us about x's mental state—surrounded by burning trees, x must be *terrified*—but we can link environment with mentality only through links that they both have to behavior.

So how do we connect behavioral premises with mentalistic conclusions? If in attributing beliefs and desires we are not trying to fill a causal gap, as Descartes thought, what are we doing? I am not in love with any of the answers that have been given to that question, and I have none of my own. I raised the question only so as to set it aside openly rather than on the sly.

3. Guinean pessimism

Although I don't know what, ultimately, is going on when we infer mind from behavior, I am pretty sure that *something* is going on. That is, we have some reasonably disciplined ways of getting from behavior to mind, even if in practice we often talk with licentious indiscipline, out of carelessness, eagerness to publish results in cognitive ethology, or loyalty to domestic pets.

Some philosophers disagree. They hold that we cannot respectably attribute contentful thoughts to animals on the strength of nonlinguistic behavior. So they come out where Descartes does, though not for his reason. They hold that we have no good *standards* governing our moves to conclusions about thoughts from premises about nonlinguistic behavior. Whatever the metaphysical underlay is of those moves, they say, the moves themselves are a mess.

Quine, for instance, says that when we attribute beliefs to other animals we are speaking in a 'dramatic idiom', imagining ourselves in the animal's shoes, so to speak, and saying on its behalf what we imagine *we* would think or be prone to say if *we* were barking at a cat up a tree or lunging at a toreador with our horns.¹ When we say of a lower animal that it thinks *that P* for some P, Quine holds this to be just a loose, impressionistic *façon de parler*, with no structure worth examining.

Some have followed Quine in extending this to the attribution of thoughts to humans. Speech provides evidence of what the speaker thinks only if we know what she means; and Quine holds that there is no such thing as exactly what a person means by an utterance. This leads him to conclude that the notion of contentful thought has no role in a clean

description of any part (even the human part) of the universe. But I want to keep out of the argument about that. Any Quinean will say that to attribute thoughts on the basis of speech is more respectable than to attribute them on the basis of other behavior. When we infer states of mind from speech, even if our inferences are somewhat shaky and their conclusions are not rock-hard news about reality, we aren't engaging in the loose, unprincipled, libertine conduct that is involved in saying what a dog wants or what a chimpanzee believes. In short, say Quine and his followers, there is a fairly decent basis for crediting people with thoughts, and no fairly decent basis for attributing thoughts to animals that don't have languages.

If this is right, it could be the source for the common view that we have readily and superficially available to us a single deep wide difference in mental capacity between us and them. The person in the street does believe that some beasts think; but if really we have no sound basis for attributing thoughts in the absence of language, this might be something that plain folk unconsciously recognize, and their intuitive picture of a chasm between humans and other animals might come from that unconscious recognition. Adapting a phrase from Descartes, it may be that ordinary people represent the beasts to themselves as unthinking without judging them to be unthinking.

But I don't believe any of this. Starting in my book *Rationality* in 1964, and continuing in *Linguistic Behaviour* twelve years later,² I have argued that our moves from nonlinguistic behavior to belief need not be as sloppy as the Quineans imply them to be. My attempt to lay bare the central conceptual structure was incomplete, and partly

¹ W. V. Quine, *Word and Object* (M.I.T. Press: Cambridge, Mass., 1960), pp. 218f.)

² Jonathan Bennett, *Rationality: an Essay Towards and Analysis* (Routledge and Kegan Paul: London, 1964); *Linguistic Behaviour* (Cambridge University Press, 1976).

wrong, but I still think it pointed the way to something feasible. Despite recent sceptical literature tending the other way, I unashamedly conjecture that many animals of other species have beliefs and wants, and that this is not anthropomorphism—not a ‘dramatic’ way of pretending to be a mouse or a gorilla—but rather a sober though perhaps not basic statement of fact.

4. Language: expressive power

Most of us agree with Descartes that the chasm between us and the other animals has *something* to do with language. In this address I shall give some reasons for thinking that that is right. I shall look at various suggestions that have been made for how men and women differ from the beasts, and shall contend that each seems to depend on facts about our languages.

To get a fix on what might be special about our languages, we need to contrast them with the signaling systems of lower animals, the natural repertoires of sounds and movements whereby they inform one another. What do our languages have that is not shared by those signaling systems?

Well, firstly, our languages have larger vocabularies than theirs. Vervet monkeys are renowned for their three kinds of warning cry, responding to three kinds of predator and leading to three kinds of behavior by the hearers. That the vervets are famous for this richness in their signaling system provides some measure of the gap between them and us so far as vocabulary is concerned.

Secondly (and this is more important), the so-called vocabulary in nonhuman signaling systems consists of whole utterances (analogous to our sentences) rather than of separately meaningful parts of utterances (analogous to our words). Nothing in those signaling systems resembles the syntactic devices through which our languages assemble meaningful

elements in many ways, creating a practically limitless range of possible utterances. That is how ‘the dullest of men’ does what Descartes thought no physical mechanism could do—namely ‘produce different arrangements of words so as to give an appropriately meaningful answer to whatever is in his presence’.

Let me qualify a little what I have just said. The dances of honey-bees signal, among other things, the direction from the hive to the source of food. Here is my favorite empirical fact: the angle between the axis of the dance and the direction of gravity is the same as the angle between the line from the hive to the food and the line from the hive to the sun. This signaling of direction is a quite fine-grained affair, giving honey bees a large repertoire of dances specifying many directions in which food may be found.

But this semantic machinery is analog, not digital as ours is. The wider the angle between the vertical and the axis of the dance, the wider the signaled angle between the line to the food and the line to the sun. Thus, of any three signals, the one whose meaning is intermediate between the meanings of the other two will itself be physically intermediate between the other two. There are many other examples of this—for instance all the species that have warning signals where the loudness or length or shrillness of the signal correlates with the severity of the danger. In all of these, the place that a signal occupies on a certain physical continuum determines the place that its meaning occupies on a certain continuum of meaning.

This is like the difference for us between the meaning of ‘Ouch!’ and that of ‘OUCH!’, not like the difference between ‘That hurts a little’ and ‘That hurts a lot’. The latter difference involves the digital kind of semantic structure that gives our languages nearly all their expressive riches; and *that* is what nonhuman signaling systems all lack.

5. The psychological underlay

We often speak *intentionally*, whereas the natural communicative behavior of other animals appears to be fairly hard-wired. I am assuming that the only mark of behavior's being intentional is its being flexible, finely and rapidly adaptable to relevant changes in the circumstances; and it is not clear that the signaling systems of nonhuman animals in the wild are ever like that. Perhaps some of them are responsive over time to the animal's experience of what works in achieving its goals, but not enough to provide honest work for the concepts of intention and individual purpose.

Ethologists at Rockefeller University, led by Peter Marler, have been studying alarm calls, and food calls, by cockerels— young roosters. They have found that when a cockerel sees an apparent predator, it is most likely to utter an alarm call if there is a female of its own species in the next cage, less likely to if its neighbor is a male conspecific, and almost certain not to if the next cage is empty or contains a bird of some other species. There is a different but equally definite pattern in its uttering of food calls.

Marler offers this study as a modest start on the question of whether the cockerel's calls are intentional. It is indeed a start; for it answers 'No' to the question: 'When the cockerel utters an alarm call, is this just a hard-wired reflex response to seeing a predator?' But the start is modest indeed. I asked Marler whether the cockerel, when it sees the seeming predator, can at that moment see what is in the next cage; and he said Yes. So the alarm calls of the cockerel might be hard-wired responses to a double visual stimulus—the appearance of a predator linked with the appearance of another bird of its own species. I want to know what happens when the bird has only indirect evidence for what is in the

next cage: can its experience alter what it takes as evidence for the presence of a female, say, and does that in turn alter the probability of its uttering alarm calls? Marler's team are starting to look into that; my guess is that there will turn out to be only a slight complexity in the hard wiring, not the degree of soft wiring—of individual flexibility—that indicates intention.

6. Teaching languages to chimpanzees

So there are the two big obvious differences between our languages and nonhuman signaling systems. We can put subsentential units together to form sentences that have significant structure; and our linguistic behavior is often highly intentional, ad hoc, deliberate, chosen. The natural signaling systems of other animals—the ones they use in the wild—have neither that kind of structure nor, it seems, that kind of psychological underlay.

With that in the background, let us look at the attempts that have been made over the past two decades to teach languages to chimpanzees. These attempts have varied a lot in how well they were conceived and how sharply observed and scrupulously documented the data were. The pioneer effort involved teaching a chimpanzee the American Sign Language; but performances with a gestural language are hard to document reliably, and when a few years ago some published reports of American Sign Language 'conversations' with a chimpanzee were carefully checked against videotapes of those same episodes, the fit was poor.

Much better was the work of David Premack and his associates, who invented a language in which the 'words' were specially colored and shaped plaques and uttering a sentence consisted in putting some of these, in the right order, onto a board.¹ Premack invented puzzle situations

¹ See, for example, David Premack and Ann James Premack, *The Mind of an Ape* (Norton: New York, 1983), ch. 1.

where chimpanzees were invited, in effect, to construct indicative sentences that would be true of some bit of the world, and to act in ways that were enjoined by imperative sentences. Eventually, the chimpanzees learned *principles* of sentence construction, enabling them to understand sentences that they had not previously encountered—producing new indicatives in the right situations, and acting in the right way in obedience to new imperatives.

Let us compare and contrast the systems taught to Premack's chimpanzees with the languages of humans.

Well, the chimpanzees didn't develop theirs for themselves, but that doesn't matter for my topic. If we think that language makes us special among animals on this planet, *that* belief doesn't involve anything about the origin of language. It wouldn't evaporate, for example, if we discovered that our languages descended from one that was taught to Adam by God or by a snake on a Wednesday afternoon in 4004 BC.

A more pertinent question is this: Is the chimpanzees' use of their artificial languages intentional? Well, it certainly isn't hard-wired, as is shown by the fact that they learn it. But is it flexible and adaptable enough to count as intentional, or at least to count as very intentional? So far as I can see, it isn't and *couldn't* be so. The trouble is that what the chimpanzees are taught is a *system*: their grasp of it is to be shown in their ability *systematically* to match sentences onto the world either by uttering true indicatives or by obeying imperatives. The very regularity of this behavior unfits it to display that adaptive flexibility that is the only clear mark of intention.

Behavior that does indicate intentionality will be of sorts that would be reported in *anecdotes*—accounts of

individual bits of behavior that fall into no stimulus-response pattern. Ethologists used to be trained to think that they should report hard data in the form of behavioral regularities; anecdotes are charming—they were told—but not instructive, and can only distract us from science. But that was wrong; it virtually guaranteed them against finding good evidence for flexible, individually adaptable intentionality. These days they know better. A recent issue of *Behavioral and Brain Sciences* carried an article about deceptive behavior of primates; this was intended as to contribute to the study of one kind of intention; and the data base was assembled by writing to ethologists all over the world asking them for *anecdotes*.¹ Anyway, there is plenty of anecdotal evidence that chimpanzees, captive and wild, have intentions in much that they do; but their wild uses of signaling systems, and their tamed uses of humanly devised languages, do not and apparently *cannot* offer anything much of that kind.

Our languages, on the other hand, although they are systematic, are so lavishly endowed with words and structures that our uses of them can be flexible, opportunistic, non-routine, to a virtually unlimited extent. Anyone who thinks that the average person's linguistic behavior might be just a *routine* hasn't been listening, even to our simplest speech. When Ferdinand said to Miranda 'Here's my hand', it was not the learning of a routine that enabled her to reply 'And mine, with my heart in it'. In contrast with that, the impoverished quasi-linguistic systems that are taught to chimpanzees provide little possibility for anything beyond the learning of stimulus-response routines.

¹ A. Whiten and R. W. Byrne, 'Tactical Deception in Primates', *Behavioral and Brain Sciences* 11 (1988).

7. Present and particular

Poverty in structure and vocabulary, as well as restricting the chances for chimpanzees to manifest a complex psychological underlay for their linguistic behavior, also limits the kinds of thoughts they can express in their languages. They never say anything complicated, and their range of themes is small.

One limitation of theme deserves special comment: Their imperatives always call for *immediate* action, and their indicatives always describe the *immediate* environment. Any thoughts they express are particular, not general, and concern the present or future, not the past.

In *Rationality* I contrasted the intellectual grasp of what is present and particular with the ability to think about what is past and to have general thoughts, and argued that that contrast defines the intellectual line between humans and other animals. One strand in that whole argument still seems to me to be worth considering.

It is the contention that to express beliefs about the past, or to express general beliefs, one needs a language of the right kind: roughly speaking, language with a past-tense operator (in the one case) and with something like quantifiers (in the other). Of course an animal's present behavior, and even its present thoughts, can be *affected by* the past: the dog hides from the gypsy because he looks like someone who kicked it yesterday. But that fact about the past does not have to enter into the content of the dog's beliefs—the dog doesn't have to think *Someone like this kicked me yesterday*—for the past event to influence its present behavior and even its present belief that this man is dangerous. Again, its present behavior and thoughts may *instantiate* general truths: the dog always hides from people who resemble past assailants. But that is a far cry from its having the universal thought that all such people are dangerous. Being affected by the past and behaving in accordance with general

procedures—neither of these takes an animal out of the prison of what is present and particular.

It is an open question whether evidence of having thoughts about the past, or general thoughts, *could* be provided in any way except through language. I am still somewhat inclined to think that it couldn't. Some people think it could, because a languageless animal might make gestures or grimaces that struck us as reminiscent—that seemed to say 'There you go again!' or 'If I've kicked you once, I've kicked you a thousand times!' or the like. I am turning my back on interpretations of that kind, based as they are on intuitive impressions of what an animal seems to be thinking. What one says along those lines depends far too much on how one feels about non-human animals, and not enough on any objective facts.

My hunch is that only through language can one show that one has thoughts that are not about what is present and particular. Not every language will do the job, of course. In *Rationality* I imagined a creature whom I called The Describer, who has a language that doesn't take him to what is past or to anything general: he is quick and efficient in describing what confronts him, but that's as far as his linguistic capacity goes. For him, language is just something with which he hits back at his environment. (I could have expanded the account by letting him obey imperatives, so that language would also be something that he hits back at with other behavior.) I invited the reader to agree that intuitively this 'Describer' seems to belong on the nonhuman side of the great divide—that he is more like an intricate thermometer or Geiger counter than like one of us. The Describer, it now seems to me, was a precursor of the language-using chimpanzees. Insofar as their language-use is directed purely at the immediately given (for indicatives) and the immediately do-able (for imperatives), it does not

release them from the present and particular. Human languages give us that release, transforming our condition.

8. Thoughts about thoughts

The 'languages' that have been taught to chimpanzees are not merely confined to the immediate present environment, but also stay on that environment's surface. In those languages, the topic is never what others are thinking.

Let us stay with this for a while. It may be *possible* but it is not *easy* to say what would count as nonlinguistic behavioral evidence that an animal has thoughts about thoughts. Some ethologists and psychologists credit their subjects with thoughts about thoughts by simply letting such attributions *drift* into their interpretations unexamined. Others do better, explicitly raising the question and looking for a basis for answering it. But no persuasive evidence has yet been found.

For example, Robert Seyfarth and Dorothy Cheney in the course of their fine work on vervet monkeys, have looked into the psychology of hearers, studying how vervets respond to the utterances of a liar. The monkeys recognize the individual voices of their companions; what Seyfarth and Cheney did was to make recordings of alarm calls by one monkey, and to play them through hidden loud-speakers at a time when that monkey was out of sight and there was in fact no danger. After a few such false calls, the others stopped responding in the normal way to alarm calls of that kind from that particular animal (whether uttered live or played through the speaker). This is offered as *prima facie* evidence that the vervets hear one another's alarm calls as intended, or as having minds—sometimes lying minds—behind them, which is to credit vervet monkeys with thoughts about thoughts.

But really the evidence shows only that the vervets' responses to alarm calls can be altered in the light of experience. They instantiate a principle which informs the behavior of almost all animals, however lowly—the principle 'If it doesn't work, drop it'. There is nothing here to give pause to the most hardened Skinnerian behaviorist.

Premack and his associates tackled the question in a more experimental way, with chimpanzees. Their paper 'Does the Chimpanzee have a Theory of Mind?' gives their reasons for answering Yes to that question; but their data don't support their answer very well.¹ The crux of the experiments consisted in training the chimpanzees to study videotapes of a human engaged in some task where something goes wrong; the chimpanzee is then shown pictures of possible continuations, only one of which shows the human fixing whatever went wrong; the chimpanzee's task is to select that one from all the pictures; and to the extent that it succeeds, that is evidence—says Premack—that it has a true belief about what the human *wanted* to do; and that is a thought about a thought.

This reasoning is vulnerable. The data provide some evidence that the chimpanzee makes predictions about how the person in the video will behave, and about how he will behave if. . . : 'If the rope doesn't break, he will pull up the bucket. If the rope does break, he will tie the ends together.' That is the kind of thing that is needed for evidence of beliefs about the person's thoughts: I can't *show* through my behavior what I think you think except by showing how I think how you will behave, or how you will behave if. . . But merely having predictive and conditional opinions about your behavior is not necessarily enough for having opinions about your state of mind. Quite generally, if the subject has a rich

¹ D. Premack and G. Woodruff, 'Does the Chimpanzee have a Theory of Mind?', *Behavioral and Brain Sciences* 1 (1978), pp. 515–26.

and complex profusion of opinions about behavior, which can be reduced to a manageable order only by being represented not as hundreds of beliefs about the object's behavior but as a few beliefs about what the object thinks and wants, *then* there is evidence that the subject has thoughts about the object's thoughts; but not otherwise. Premack's chimpanzees reveal no such profusion, and we are thus under no pressure to credit them with thoughts about thoughts as distinct from thoughts about behavioral dispositions.

9. Gricean thoughts

Some of us hold that thoughts about thoughts, rather than being a conceptual luxury, are needed for language properly so-called. When you and I speak to one another we can and sometimes *do* speak under the guidance of thoughts about what the other person will think that we want to get across. Those are thoughts about thoughts about thoughts. I shall call them 'Gricean', in honor of Paul Grice's theory that such thoughts are of the essence of meaning, and thus of language. I spent half of my book *Linguistic Behaviour* defending this theory, but even if it is wrong no-one could deny that a large, conspicuous, and valuable part of the human use of language consists in adaptations that may be made by a speaker in the light of what he thinks that the hearer will think that he wants. In short, we often speak with Gricean intentions.

Is there anything like this in the linguistic behavior of chimpanzees? Some of the chimpanzee-language people (not in the Premack group) are given to saying things like: When Sherman made the sign for a rake, he wanted Austin to *think that he (Sherman) wanted the rake*. This credits the chimpanzee Sherman with a desire to produce a belief about a desire. But the behavioral evidence doesn't stretch further than: When Sherman made the sign for a rake, he wanted

Austin to *bring the rake*.

There is a question as to how the animals *could possibly* give evidence that they were using their language with Gricean intentions. How *could* there be evidence that the speaker intended anything fancier than just to get the hearer to do such-and-such?

Well, how do *we* give evidence of this? I tried in my book to describe non-linguistic behavior that would show that the agent had behaved with a Gricean intention. But to succeed I needed a fancy story of a kind that could not loom large in everyday life. It seems to me that what entitles each of us to be *confident* that his or her fellows have Gricean intentions firmly in their repertoire consists in things we can *say*. In Dickens's *Our Mutual Friend*, for example, a wife says to her husband: 'I believe, dear John, that you believe that I believe that we have as much money as we require.' This reports a belief about a belief about a belief; that has the same abstract structure as a Gricean intention, which is an intention to produce a belief about an intention. We have no trouble understanding the remark, and could easily be satisfied that the speaker meant by it what it conventionally means. Furthermore, once the remark had been made it is not so hard to think up nonlinguistic things she might do that would *confirm* that she did indeed believe that he believed that she believed that they had enough money.

That example—like the countless others I could give—relies on two special features of our language. One is its psychological vocabulary: its containing terms like 'believe' and 'want' and 'hope'. The other is its *iterative* power: a moderate grasp of the form 'I believe that P' brings with it a grasp of 'I believe that you believe that P', 'I believe that you want me to believe that P', and so on.

10. The 'I' thought

Some philosophers have thought that humans, alone among terrestrial animals, are self-conscious—not meaning that they alone have conscious states, but that they alone can accompany their conscious states with the thought that 'I am in such and such a condition'—the 'I' thought. Kant wrote that if he had the mentality of a subhuman animal, he might have sensory states but

I should not be able to know that I have them, and they would therefore be for me, as a cognitive being, absolutely nothing. They might still. . . exist in me (a being unconscious of my own existence) as representations. . . , connected according to an empirical law of association, exercising influence upon feeling and desire, and so always disporting themselves with regularity, without my thereby acquiring the least cognition of anything, not even of these my own states.¹

Even if we don't agree with Kant—and I don't—about how much depends on one's ability to have the 'I' thought, we can agree that that ability is remarkable, valuable, and not obviously possessed by other animals on our planet.

Being more cautious than Kant was, I say 'not *obviously* possessed by other animals'. What grounds have we for thinking that other animals do not have the 'I'-thought? What is missing from their repertoire such that, if it were included, we *would* have reason to credit them with that thought?

Well, in recent years some psychologists have offered putative empirical evidence for a view about how widely

the concept of oneself is distributed through the animal kingdom. Oddly, their evidence does not mark off humans as special, nor humans and the great apes; rather, it separates humans and chimpanzees and orang-outangs from all the rest, including gorillas!

Here is a sketch of the facts. Place a large mirror beside a chimpanzee's cage, and within a few days the chimpanzee will start using her image in the mirror as a means of exploring parts of her body that she cannot see directly. A chimpanzee that has long shown an interest in strange marks on her arms, say, will use her mirror image to explore marks that she otherwise couldn't see on her forehead. Most chimpanzees start to respond in this way quite quickly, and so did an orang-outang, whereas animals of other species don't do so *ever*. Some monkeys have been given months, even years, in which to recognize themselves in mirrors, with no success. As the years roll by, the monkey continues either to neglect its mirror image or to treat it like a potential aggressor.

Independent studies show that some of these species can use mirror-images to locate things in real space; they can even use them to recognize and locate other animals with which they are familiar. But they seem *never* to use them to recognize themselves.

The discoverer of this phenomenon, Gordon Gallup,² explains the data by conjecturing that only chimpanzees and orang-outangs share with humans a *concept of self*. According to this hypothesis, which has been widely accepted by psychologists who have explored the matter, the monkey never confronts its own image with the thought 'That is me'

¹ From a letter to Marcus Herz, quoted in Norman Kemp Smith, *A Commentary to Kant's 'Critique of Pure Reason'* 2nd. edn. (Humanities Press: New York, 1962), pp. xlix-l.

² See, for instance, Gordon G. Gallup, Jr., 'Chimpanzees and Self-Awareness', in M. A. Roy (ed.), *Species Identity and Attachment: a Phylogenetic Evaluation* (Garland: New York, 1980), pp. 223-243.

because that involves the 'I'-thought, which monkeys are not equipped to have. Here is my way of putting the case for this interpretation. The difference to be explained is systematic, and it involves mental capacity; it doesn't concern sensory intake, so it must be a difference in what thoughts are had; the relevant thought has to do with *recognizing oneself in a mirror*, but monkeys and the rest don't have trouble with recognition or with mirrors, or even with recognition in mirrors; so it seems that the thought they cannot have is the thought of *oneself*—the 'I'-thought. That's fairly persuasive, on the face of it; but let's not go too fast.

Consider a bright female monkey that can use some mirror data, e. g. using a mirror image to recognise her mate, and to locate things in real space. Presumably she can use a mirror to locate things that are close to her, as well as things a few feet away; but then can she locate things that are brushing against her fur, pressed hard against her body, painted onto her body? One would expect the answer Yes to each of these questions, though the experimental literature does not address them. If the answer is Yes to all of them, then an animal that sits below Gallup's line can use a mirror as an aid to finding items painted onto its own body.

Skinner and some of his pupils got pigeons to do just that.¹ The birds were trained to peck at blue spots, and were adorned in such a way that blue spots on their own breasts could be seen by them only in mirrors; after a while, they learned to use the mirror images as evidence and to peck at the spots that were painted on themselves.

The responses to this by the Gallup camp have been unsatisfactory, but I am not yet able to do much better. Skinner's pigeons were using mirrors to locate items that coincided in real space with their own bodies; and I have been

conjecturing that any animal that can use mirror images at all can use them to do that. So, what *more* is going on when a chimpanzee uses her image in a mirror to explore paint marks on her forehead? I don't doubt that more is going on; Gallup and others have uncovered a real phenomenon here, which cries out for systematic explanation; but it is not clear what the explanation should be or even, in detail, what the phenomenon is.

It is plausible to suppose that the relevant difference between chimpanzee and pigeon is that the chimpanzee is not merely interested in paint marks that are on her, but is interested in them *because* they are on her. Her self-scrutiny in the mirror seems to express a general interest in her body—an interest that is manifested in other ways without help from mirrors. She doesn't care about other bodies in the way she cares about her own. The pigeons' use of mirrors, on the other hand, did not reflect any special attentiveness to their own bodies in particular.

Plausible as that is, I'm afraid that it won't do as it stands. Granted, the chimpanzee has a special interest in one particular animal body; among other things, an interest in marks that she finds on it. But any dog or monkey or elephant does in some way give special attention to just one animal body among all the others—for example, by taking action to avoid damage to that body, and by its being the body that the animal directly *moves*. So it can't be doubted that even quite lowly animals such as pigeons distinguish within the perceptually given world between their own bodies and the rest, and have a special cognitive relation to the former.

Perhaps what Gallup needs for his interpretation is some reason to say that the chimpanzee attends to the mark on

¹ Robert Epstein, Robert P. Lanza, and B. F. Skinner, "Self-Awareness' in the Pigeon', *Science* 212 (1980), pp. 695f.

her brow because ‘it is on *me*’ and not merely because ‘it is on *this*’, where ‘this’ is the old, familiar, specially interesting body, the one that can be moved directly and that it hurts to get damaged. I can’t see that we have any reason to say that; so I have to conclude that Gallup’s interpretation is unsupported.

That distinction between ‘it is on *me*’ and ‘it is on *this*’ brings us from the chimpanzees back to Kant. When he gave pride of place to the ‘I’-thought, he meant it to be the thought of oneself *as a subject of thoughts*; and I suggest that if the ‘I’-thought is to have any chance of marking us off from the beasts it must be the thought of oneself as something that has thoughts, enjoys experiences, and so on. It is really the ‘I think’ thought, and after all it has nothing to do with chimpanzees and mirrors.

How, then, could it possibly be manifested in non-linguistic behavior? I have no idea. It seems to me likely that it couldn’t be, and that our only way of showing that we have such thoughts is by the use of our structured, iterative, endlessly fertile language.

11. Conclusion

It is time to stop. I have been sketching reasons for two conclusions. First, it is only because our languages are rich in lexicon and structural possibilities that our speech can be seen to be strongly intentional. Second, that richness is

also required for us to be able to express thoughts about thoughts, thoughts about anything that is not present and particular, thoughts about ourselves as thinkers.

These conclusions bring me close—too close for comfort—to the widespread view that the crucial difference between humans and other animals is that we do and they don’t have languages. I would prefer to be at a greater distance, because my philosophical instincts tell me that *language* is not such a large a part of the story as it is commonly held to be. But this uncomfortable conclusion seems to force its way in whenever I try to think clearly and unsentimentally about what kinds of thoughts can be attributed on the evidence of nonlinguistic behavior.

Suppose it is right. What then? Is this just a fact about how such thoughts can be manifested? Does it leave untouched the question of whether and by whom such thoughts can be had? I don’t think so. Even the arch-dualist Descartes did not always divorce the having of thoughts from the ability to manifest them in behavior, and I don’t think that we should do so either. Saying *that* doesn’t explain what the connection *is* between having thoughts of a certain kind and being able to express them in behavior. It can hardly be doubted that there is such a connection, but I have already admitted that I don’t know—and unlike some others I *know* that I don’t know—what connection it is.¹

¹ My thinking about the issues raised in this lecture has been helped by William P. Alston, José Benardete, Gillian Bennett, Sara Bennett, Frances Howard, Thomas McKay, Brent Mundy, Emily Robertson, and Robert Van Gulick.