Hume's epistemology (5): reason and belief in animals (*EHU* section 9)

Do non-human animals (henceforth just 'animals') have minds? Do they feel, perceive, think, reason and so on? And if they do possess some or all of these capacities, are the ways in which they feel, perceive, think, or reason, similar to the ways in which we humans feel, think, perceive, or reason? In *EHU* 9 Hume engages (albeit briefly) with these sorts of questions.

Historical accounts

Descartes on animals

Before we investigate what Hume says about animals, it will be useful to see what one of his most illustrious predecessors had to say on the issue. Recall that, according to Descartes, or, at least, according to a certain standard interpretation of Descartes, the universe was divided into two sorts of thing. On the one hand there was the material world (including the human body and animals). The material world was a giant machine, unfolding relentlessly according to the laws of brute physical causation. On the other there was mind (consciousness), separate from, but interacting causally with, the material world. Thus we get the traditional Cartesian dualism of machines and minds, or, as Anthony Kenny puts it, of 'clockwork and consciousness'. But now since, for Descartes, animals are 'mere' machines, that means that animals don't have minds. They don't have conscious feelings, sensations, or thoughts, and they certainly don't engage in rational deliberation. (Actually, there are some tricky interpretative issues here that are currently being fought out on the playing fields of Descartes scholarship.) Hume's view however is very different from Descartes'.

Hume on animals

Hume states that his naturalistic science of human nature will 'acquire additional authority, if we find, that the same theory is requisite to explain the same phenomena in all other animals' (*EHU* 9:1). So he thinks that his theory of human nature will receive extra support if we find that the same style of explanation which is available for patterns of behaviour in animals, is available for those patterns of behaviour in humans that appear to be the same. He then argues that humans and animals share certain core psychological states and mechanisms.

First, Hume suggests that animals learn from experience, and thus engage in cause-effect reasoning (inductive inference). You will recall that, according to Hume, in matter-of-fact reasoning to so-far-unobserved cases, we always infer from causes to effects, or from effects to causes. Such inductive inferences involve an extrapolation from observed events of a particular kind to so-far-

unobserved events *of the same kind* (notice Hume's comments on similarity and analogy at the start of *EHU* 9). The psychological mechanism responsible for induction in humans is *custom* or *habit*, which is a *natural instinct* or *mechanical tendency* (*EHU* 5:22). When we have experienced a *constant conjunction* (regular and uniform succession) between two events, say A being followed by B, custom leads us to form an expectation that when an instance of event-type A occurs, it will be followed by an instance of event-type B.

Hume claims that animals perform this kind of experience-based cause-effect reasoning too. Older animals are smarter than younger animals (*EHU* 9:2), and the behavioural tendencies of animals can be altered by 'the effects of discipline and education' (*EHU* 9:3), in ways which indicate that animals are able to learn from experience to expect 'from the present object the same consequences, which it has always found in its observation to result from similar objects' (*EHU* 9:4). These observations are followed by a line of reasoning which is intended to establish that, in the case of animals specifically, the capacity for inductive inference cannot be due to the capacity for reason (in the sense of demonstrative reasoning). If inductive inference were due to reason (which it isn't!), the arguments involved would surely be too 'abstruse for the observation of such imperfect understandings' (*EHU* 9:5). Hume also thinks that they would be too abstruse for human infants. Moreover they would be too abstruse for adult humans (including both the 'vulgar' and the philosophers) during their ordinary, non-reflective lives. In other words, induction is a widespread, ubiquitous phenomenon in ordinary animal and human life, so it can't be a product of reason, which a sophisticated and rarely exercised cognitive capacity. Hume's final flurry is to suggest that reason is in fact too unreliable for nature to have entrusted it with the important task of performing inductive inferences. The psychological mechanism actually at work is, of course, custom, which 'engages animals, from every object, that strikes their senses, to infer its usual attendant, and carries their imagination from the appearance of the one, to conceive of the other, in that particular manner, which we denominate *belief*' (*EHU* 9:5) So experimental reasoning (cause-effect reasoning, induction) is a *natural instinct,* an innate, unconsciously operating, 'mechanical power' (*EHU* 9.6) possessed by animals and humans alike.

Here, then, are some of Hume's conclusions concerning animals: animals engage in inductive inference; animals have beliefs (in the same sense that we do); animals don't have the faculty of reason (demonstrative reasoning); animals experience fear, pleasure, and other passions (NB: this claim comes out in some of Hume's descriptions of animal behaviour; see, e.g., *EHU 9.2*); we share many psychological capacities with animals.

Contemporary accounts

Davidson on animals

To see why some contemporary philosophers might be nearer to Descartes than to Hume on the issue of animal minds, we might consider an influential analysis due to the American philosopher Donald Davidson ('Rational Animals'*, Dialectica*, Vol. 36, 1982). Davidson holds that to be rational is to have propositional attitudes (beliefs, desires, hopes, fears, expectations etc.). Because Davidson is a *holist* about the propositional attitudes, he holds that for an individual to have any single propositional attitude, that individual must, as a matter of conceptual necessity, have many other related attitudes. It is crucial that many of these attitudes will be beliefs. As Davidson puts it, 'Without belief there are no other propositional attitudes, and no rationality as I have characterized it'. Some of the beliefs in the network will be general (about classes of things) and some will be specific (about individual things). Each belief (general or specific) requires other beliefs, and those other beliefs require further beliefs, and so on. Hence where we are correct in attributing one single belief, we are justified in inferring the existence of many other related beliefs. Conversely, where we cannot make sense of the required network of related beliefs, the attribution of any single belief that depends on the network is not warranted.

Davidson observes that when we turn our attention to animals, the inferential process of belief attribution soon becomes tenuous. In ordinary parlance, we all say things along the lines of 'the dog thinks that the cat went up that tree'. But the fundamental validity of the individual belief attribution ('the dog thinks that the cat went up that tree') depends on the dog having a network of related beliefs, for instance, that trees grow, or that they sometimes have leaves, or that they are living things that will die without water. Whilst there is no fixed set of beliefs that a creature must necessarily possess in order to have beliefs about trees, without many general beliefs of the sort just listed, there is no reason to attribute the creature with any beliefs about trees, or with the belief that a particular object is a tree, or with beliefs involving objects that are trees. But surely we cannot, with a straight face anyway, claim that dogs have beliefs of the sort 'trees are living things that will die without water'. In short, the inferential process of assigning a network of beliefs to the dog breaks down. Given holism about propositional attitudes, such breakdowns cast serious doubt on the original attribution of the belief that the cat went up that tree. To Davidson, these considerations suggest that the whole practice of attributing propositional attitudes to animals is really no more than a bout of anthropomorphizing sentimentality, even if the practice is, as a matter of everyday fact, predictively useful.

Why, according to Davidson, is it implausible that animals should have the background of beliefs that would enable us to justify attributions of even a single propositional attitude? In other

words, what sort of empirical evidence is relevant to deciding which creatures are bearers of beliefs? The candidate considered by Davidson is the behaviour of the creature under investigation. It seems to be a pretty straightforward implication of Davidsonian holism that the patterns of behaviour that would licence belief attribution would have to be complex. If we have no evidence that such behaviour patterns exist, then we have no good reason to attribute propositional attitudes to the creature concerned. And, according to Davidson, the only sort of behaviour that provides the necessary complexity is linguistic behaviour.

Davidson's argument for this conclusion depends on a distinction between having a belief, and having beliefs about beliefs; and one sure way to know that you have beliefs about beliefs is to be surprised. To use Davidson's example, if I believe that there is a coin in my pocket, but then discover that there is nothing in there but buttons, I am surprised. In fact, and here's the rub, having the belief that there is a coin in my pocket seems to entail the possibility that I could be surprised, that is, that there could come a time when I am aware of a contrast between what I have believed about the contents of my pocket, and what I believe then. So having the belief that there is a coin in my pocket entails that I have a belief about the correctness of that belief. The implication is that having a general stock of beliefs necessarily requires one to have beliefs about the correctness of at least some of those beliefs. Now in order to have a belief about a belief, one must have the concept of belief; so, for Davidson, it follows that in order to have beliefs at all, one must have the concept of belief.

It seems that a large part of having the concept of belief is having the concept of an objective reality which is independent of beliefs about that reality; that is, it is to have mastery of the contrast between subject and object. Thus a command of the subject-object dichotomy is a necessary concomitant of having beliefs. According to Davidson, one way of gaining mastery of the subject-object contrast is to have the concept of intersubjective truth. Linguistic behaviour suffices to demonstrate this capacity, because linguistic communication depends on the notion of a shared intersubjective world about which communicators can agree, disagree and exchange information, and about which the communicators may be mistaken; that is, linguistic communication rests on the notion of intersubjective truth. So, via the notion of intersubjective truth, language-use demonstrates command of the subject-object dichotomy, and, therefore, provides sufficient evidence for belief.

It is important to realise that, in coming to his conclusion, Davidson does not fail to recognise the subtlety of animal behaviour; he merely claims that it is not the outcome of beliefs. As he puts it: 'A creature may react with the world in complex ways without entertaining any propositions. It may discriminate among colors, tastes, sounds and shapes. It may "learn", that is, change its behavior in ways that preserve its life or increase its food intake. It may "generalize", in the sense of reacting to new stimuli as it has come to react to similar stimuli. Yet none of this, no matter

how successful by my standards, shows that the creature commands the subjective-objective contrast, as required by belief'.

The core of Davidson's argument, as I have presented it, can be summarised as follows: (i) to have a belief, one must have the concept of belief; (ii) to have the concept of belief is to have command of the subject-object contrast; (iii) one way to gain command of the subject-object contrast is through having the notion of intersubjective truth; (iv) linguistic communication demonstrates command of the subject-object contrast, because linguistic communication requires a prior notion of intersubjective truth; so (v) being part of a linguistic community is sufficient evidence that an individual possesses the subject-object contrast, and, therefore, the concept of belief; and therefore (vi) being part of a linguistic community is sufficient evidence that an individual has beliefs.

But the conclusion stated in (vi) is actually weaker than the one that Davidson would like. The intended message of the paper was that one has beliefs if and only if one has language. However (vi) says that language is *sufficient* for belief, not that it is *necessary and sufficient*. The shortfall occurs because Davidson has not shown that the *only* way that a creature could come to have command of the subject-object contrast is through language. There are actually two stages to the relevant part of the argument: Davidson needs to show that the *only* way that one could come to have command of the subject-object contrast is through having the concept of intersubjective truth, and then that the *only* way that one could have the concept of intersubjective truth is through having a language. Davidson takes it that the second stage of this argument is relatively uncontentious. The real problem (he thinks) is to show that having the notion of objective truth (i.e., having the subject-object contrast) depends on having the concept of intersubjective truth.

Let's home in on what might be another problem with Davidson's argument. The real work is being done not by the demands of holism per se, or merely by the observation that linguistic behaviour is complex, but by the claim that in order to have beliefs at all, one must have the concept of belief. It is only because of that claim that Davidson can make the link with the subject-object dichotomy, and, therefore, with language. The idea that to have beliefs one must have the concept of belief is supposed to be made plausible by the analysis of surprise. But the consequence of this analysis (taken in the context of the whole argument) is that animals cannot be surprised (because that requires the creature in question to have the concept of belief, which in turn requires mastery of the subject-object contrast, which in turn requires a language, which animals don't have). This puts the burden on Davidson (and anyone else who shares similar views) to tell us how to describe situations in which it looks, for all the world, as if animals are genuinely surprised.

Dennett on animals

We can sharpen up our understanding here if we interpret Davidson's position on animals in terms of a framework due to the influential American philosopher Daniel Dennett. In 'Intentional Systems in Cognitive Ethology' Dennett proposes a *ladder of intentionality*. ('Intentionality' here does not mean the property of being intentional in our usual, everyday sense of intentional. For present purposes, you can think of 'intentionality' as a philosophical term of art for the having of propositional attitudes). A zero-order intentional system is pure tropism, with 'no mentality, no intelligence, no communication [and] no intentionality'. A first order intentional system can be attributed states with the form 'x believes that p' where p contains no intentional terms. A second order intentional system has beliefs and desires about its own beliefs and desires, and it has beliefs and desires about the beliefs and desires of others (e.g. 'x believes that y expects x to run away'). On this view, the level to which such intentional terms can be recursively embedded has no theoretical restrictions, only practical limitations. But now notice that on Davidson's view, there are no purely first-order intentional systems, because to have a belief, a creature must have a belief about that belief. If one is convinced by the Davidsonian argument, an enormous gap opens up between Dennett's zero-order intentionality and his second-order intentionality. This gap, which potentially contains most animal behaviour, needs to be filled.