The minds of animals has been an abiding topic in philosophy since its earliest beginnings. Some may find this surprising. After all, a fairly common picture of the philosopher is someone (in a darkened study) ruminating on the nature of the human mind, or on the mind of God, or on some other abstruse idea, but certainly not on the minds of cats, dogs, and honeybees. As common as this picture may be, however, it does not paint an entirely accurate portrait. Philosophers have thought long and hard about the minds of animals and have held and defended significant and influential views on the topic. Moreover, in the past ten years or so there has been an unprecedented amount of interest among philosophers in animal minds, with numerous publications and conferences dedicated to the subject. The level of interest and publication has reached a critical mass and has sustained itself long enough that it is now appropriate to say that the philosophy of animal minds is a field in its own right. The purpose of this volume is to highlight the state of the art in the field by bringing together a collection of new and cutting-edge essays by the best and brightest philosophers.

Since the essays in this volume have been shaped by various lines in the rich history of philosophical thought on animal minds, I provide a brief (albeit, vastly incomplete) sketch of some of the most important and influential ideas and arguments in this history, as well as a road map to the volume itself.

2. HISTORICAL DEBATES

Historically the central issue in philosophy concerning animal minds was whether animals possessed thought and reason. Aristotle famously defined
The philosophy of animal minds

human beings as rational animals thereby denying thought and reason to non-human animals (see Nicomachean Ethics, book 1, chap. 7 and De Anima, book 3, chap. 10). His chief reason for doing so was his belief that animals, due largely to their failure to speak, were incapable of grasping genuinely abstract concepts and general propositions. For centuries afterward, Aristotle’s denial of thought and reason to animals provoked numerous debates among philosophers. Some, such as Chrysippus, Augustine, Aquinas, Leibniz, and Kant largely argued in favor of the Aristotelian position, while others, such as Theophrastus, Porphyry, Galen, Gassendi, and Locke, largely argued against it. The debate reached its fullest expression in the seventeenth and eighteenth centuries in the figures of René Descartes and David Hume.

Descartes went to the extreme of the Aristotelian position and notoriously denied all types of mentality – reason, thought, and consciousness – to animals. Animals, for Descartes, were nothing more than sophisticated biological machines. Descartes gave two general arguments in support of his “beast-machine” doctrine (1637/1988, chap. 5). The first rested upon his observation that animals appear incapable of transferring their skills in one domain (e.g., migration) to any other domain (e.g., locating new habitats). This failure to transfer or generalize, Descartes argued, proved that the intelligent-like behaviors of animals were due not to reason – that is, “a universal instrument which can be used in all kinds of situations” – but to naturally endowed instincts and reflexes. Descartes’s second reason rested upon the claim that animals lack speech. Like Aristotle, Descartes denied speech to animals, pointing to the fact that they never “use words or put together other signs, as we do in order to declare our thoughts to others.” The absence of speech, Descartes reasoned, could only be explained in terms of animals lacking what speech

1 Aristotle acknowledged that animals possessed sensory perception, desires, memory, imagination, and even emotions. According to Sorabji (1993), Aristotle was able to ascribe emotions to animals, despite their lacking thought, by taking emotion in animals to be a kind of perception. See Roberts (chapter 12) for a recent perception-based account of emotions in animals.
2 See Sorabji (1993) for an excellent account of the early history of the philosophical debate over animal thought and reason.
3 It is consistent with the Cartesian position to allow various types of learning in animals (e.g., imprinting, song learning, associative learning, etc.) so long as they do not involve thought or consciousness. The fact that some kinds of associative learning have been demonstrated in the spinal cords of rats without brains strongly suggests that such forms of learning exist in animals (see Grau [2002]).
4 Descartes was aware that animals produce various vocalizations but argued (much in the Aristotelian tradition) that their function was merely to express “passions,” not thought. See McAninch, Goodrich, and Allen (chapter 7) for the latest on the modern debate over expressive versus referential theories of animal calls.
expressed – thought. From this, he went on to conclude that animals also lacked all forms of consciousness, since, for Descartes, thought was taken (in the unusually wide sense) as the very object of conscious awareness – it was, as he writes in the *Principles*, “everything which we are aware of as happening within us, in so far as we have awareness of it” (1644/1988, p. 162).

The chief problem for Descartes’s position has always been to explain animal behavior without attributing mentality. As noted above, Descartes attempted to do this by appealing to mechanical reflexes and instincts in animals. But it seemed to a number of his contemporaries (as well as to a number of philosophers and scientists today) that this cannot be done for certain types of animal behaviors. As Arnauld famously replied to Descartes,

> It seems incredible that it can come about, without the assistance of any soul [i.e., without any conscious awareness], that the light reflected from the body of a wolf onto the eyes of a sheep should move the minute fibers 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. (Quoted in Descartes [1641/1984], p. 144.)

However, as we shall see later in this introduction and in a number of the essays in this volume, it is anything but obvious what (if any) types of behaviors in animals require an explanation in terms of thought, reason, or consciousness.

Equally as famous as Descartes’s denial of mentality to animals was David Hume’s (1739/1978) proclamation that “no truth appears to me more evident, than that beasts are endow’d with thought and reason as well as men” (p. 176). The type of thought that Hume had in mind here was belief, which he defined as a “lively idea” or “image” caused by (or associated with) a prior sensory experience (p. 94). Hume defined “reason” as a mere disposition or instinct to form associations among such ideas on the basis of past experience. He argued by analogy that since animals behave in ways that closely resemble the behaviors of human beings, which we know introspectively to be accompanied by the association of ideas, it is likely that animals too behave as a result of forming similar associations among ideas in their minds. This argument, Hume claimed, not only described the type of thought process that ordinary persons go through in ascribing thought and reason to animals, it provided an “incontestable” proof that animals possess thought and reason in the same manner as human beings. The only difference between the
The philosophy of animal minds

thought and reasoning of animals and humans, Hume held, was a matter of degree.5

There are two well-known problems with Hume’s theory. First, although it may be true, as Hume held, that belief and reasoning in animals are constituted by lively images and their associations, many types of beliefs and reasoning in humans are not. We can, for example, form beliefs about all sorts of abstract matters (e.g., about numbers, complex relations, non-Euclidean spaces) of which it is impossible to form an image, as well as engage in many forms of non-associative reasoning. It is difficult, therefore, to understand how human belief and reasoning could be, as Hume held, just a difference in degree from animal belief and reasoning. The other problem with Hume’s theory is that his analogical argument appears to lead to an objectionable form of anthropomorphism. After all, toy robotic dogs, computers, radios, heat-seeking missiles, and even animated circles and squares on a computer screen behave in ways that (at times) strike us as resembling the ways that we behave when we have associated ideas presented to our consciousness. But few would take such resemblances alone as incontestable proof that these objects too act as a result of associated ideas presented to their consciousness.

3. QUESTIONS AND APPROACHES IN THE PHILOSOPHY OF ANIMAL MINDS

From this very brief historical sketch, two distinct types of questions and approaches emerge which, as we shall see, carry through to the contemporary debates on animal minds and are used (in part) to organize the essays in this volume. Philosophical questions concerning animal minds generally come in two (non-exclusive) types: metaphysical and epistemological. Metaphysical questions are questions about what kinds of minds animals possess or could possess consistent with the facts about their behaviors, brains, environments, and histories. As we have seen, the focal metaphysical question in the history of the field has been whether animals could think and reason given that they cannot speak. Epistemological questions, on the other hand, are questions about our knowledge and understanding of animal minds. Such questions can be either normative or descriptive.

5 Hume went on to argue on similar analogical grounds that animals possess various types of emotions, such as pride, humility, love, and hatred (1739/1978, book II, part 1, sect. 12 and part 2, sect. 12). See Roberts (chapter 12) for an importantly different account of the existence and limits of emotions in animals.
The philosophy of animal minds: an introduction

in import. They can be about what justifies our attributions of mentality to animals, for example, or about how we in fact go about attributing mentality to animals – although the answers may be the same for both, as was seen in the case of Hume’s analogical argument.

When considering these types of questions, philosophers tend to take a top-down approach (as exemplified by Aristotle and Descartes) or a bottom-up approach (as exemplified by Hume), or some combination of both. The top-down approach begins with reflection upon mentality in humans, with special emphasis on how such mentality is expressed and described in language. The approach takes the human mind as the paradigm and judges animals more likely to have minds the closer they approximate this paradigm. The bottom-up approach, on the other hand, begins with taking our intuitively plausible (or, in some versions, our scientifically informed) ascriptions of mentality to animals at face value and proceeds with designing a theory of animal minds, such as Hume’s theory of ideas, that aims to explain the truth of such ascriptions.

Each of these approaches has its potential problems and distinctive virtues. What is yet to be seen is whether one of them, or some combination of both, can yield an account that can satisfactorily answer the various metaphysical and epistemological questions that define the field. With this in mind, let us turn to the contemporary philosophical debates over animal minds.

4. THE CONTEMPORARY DEBATE OVER ANIMAL THOUGHT AND REASON

There are three main types of arguments in contemporary philosophy for animal thought and reason: the argument from intentional systems theory (Dennett [1987]), the argument from common-sense functionalism (Carruthers [2004]; Fodor [1987]; Stich [1979]), and the argument from science (Allen and Bekoff [1997]; Bermúdez [2003a]).

The intentional systems theory consists of two general ideas. The first is that our ordinary mental-state concepts, such as our concepts belief, desire, and perceiving, are theoretical concepts whose identity and existence are determined by a common-sense theory or folk psychology. In many cases, we apply our folk psychology to animals to predict and make sense of their behaviors, and when we do, we view animals as intentional systems and take up what Dennett calls the intentional stance toward them. The second important idea of the intentional systems theory is its instrumentalist interpretation of folk psychology. On the instrumentalist interpretation, what it
The philosophy of animal minds

is for a creature to have mental states is simply for its behaviors to be richly and voluminously predicted by the principles of our folk psychology. Therefore, according to the intentional systems theory, the fact that much of animal behavior is successfully predicted from the intentional stance makes animals genuine thinkers and reasoners. The chief objection to this argument, similar to the one raised against Hume’s analogical argument, is that it seems to lead to an objectionable form of anthropomorphism. One can, after all, effectively predict the behaviors of heat-seeking missiles and even the class of thermostats by taking up the intentional stance toward them.\(^6\)

Common-sense functionalism, on the other hand, avoids this problem by taking a *realist* and *functionalist* interpretation of folk psychology. On this interpretation, for a subject to have mental states is for the subject to have in his brain a variety of discrete internal states that play the causal/functional roles and have the internal structures that our mental-state concepts describe. On this view, if Fido (for example) believes that the cat is up the tree, then he has in his brain an individual state, \(s\), that plays the causal/functional role that beliefs play according to our folk psychology and has an internal structure similar to the “that”-clause that we use to specify its content (i.e., \(s\) has the structure \(R_{xy}\), where “\(R\)” represents the two-place relation up, “\(x\)” represents the cat, and “\(y\)” represents the tree). Since the internal state \(s\) is seen as having an internal structure similar to the sentence “the cat is up the tree,” common-sense functionalism is often taken to support the view that thinking (including animal thinking) involves a language of thought (Fodor [1975]). It is argued that since animal behavior is often successfully predicted and explained by our folk psychology, we have defeasible grounds for supposing that animals actually have such internal states in their heads. It has been questioned, however, whether our everyday ascriptions of thought to animals really do presuppose the existence of sentence-like representations in their heads (Stalnaker [1999]) or whether there is any empirical evidence that shows that animals have and process such representations as opposed to iconic (non-sentential) representations.\(^7\)

---

\(^6\) Some (e.g., Searle [1994]) would take this to be a *reductio* of the intentional systems theory, but not Dan Dennett and John McCarthy: both claim that a “thermostat is one of the simplest, most rudimentary, least interesting systems that should be included in the class of believers” (Dennett [1995], p. 114).

\(^7\) For the latest on the instrumentalism versus realism debate regarding mental-state attributions to animals, see Jamieson (chapter 1), Savid (chapter 2), Tetzlaff and Rey (chapter 4), and Sober (chapter 13); and for the latest on the sentential versus iconic debate regarding mental representations in animals, see Rescorla (chapter 3), Tetzlaff and Rey (chapter 4), Carruthers (chapter 5), and Camp (chapter 6).
In the past thirty years or so, due in large measure to the demise of radical behaviorism and the birth of cognitivism in psychology, as well as from the influential writings of Donald Griffin (1976, 2001) and Randy Gallistel (1990), scientists from various fields have found it increasingly useful to propose, test, and ultimately accept hypotheses about the causes of animal behavior in explicitly folk-psychological terms. According to the argument from science, since scientists are finding it useful to test and accept hypotheses about animal behavior in folk-psychological terms, we are justified in believing that animals have such states of mind. Not everyone, however, has found the argument from science convincing. The chief concern is whether explanations of animal behavior in folk-psychological terms are, as the argument assumes, scientifically respectable (see Wynne [2004]; Kennedy [1992]). It is sometimes argued that many (if not all) such attributions of mentality to animals violate a basic methodological tenet of the science of animal behavior called Lloyd Morgan’s canon:

In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale. (Morgan [1894], p. 53)

Since animal behavior can always be explained in psychological or behavioral terms that do not appeal to our folk-psychological concepts, any scientific explanation of animal behavior in terms of such concepts is extravagant and scientifically unwarranted. It is also argued that scientific explanations of animal behavior are objective in that there is typically general agreement among researchers on what counts in favor of or against the explanation, which implies that, since the only generally agreed-upon indicators of consciousness are the verbal reports of subjects, explanations of animal behavior in terms of consciousness are unscientific (see Clayton et al. [2006], p. 206).8

No contemporary philosopher is better known for his criticism of thought and reason in animals than Donald Davidson. In a series of articles (1984b, 1985, 1997), Davidson put forward three interrelated arguments for his denial of thought and reason in animals. The first, known as the intensionality test, argues that our de dicto belief ascriptions to animals – that is, our belief ascriptions that aim to describe how the animal is thinking about some object in the world – are unscientific, since without linguistic

---

8 For the latest on the debate over the justification of Lloyd Morgan’s canon (and other principles of simplicity) used by scientists, see Sober (chapter 13) and Fitzpatrick (chapter 14); for the latest on the debate over the ascription of consciousness to animals, see Gennaro (chapter 10), DeGrazia (chapter 11), and Roberts (chapter 12).
behavior to appeal to, there are always countless different ways of saying how the animal is thinking and no principled method of deciding among them. Davidson's second argument, the argument from holism, claims that we are also unwarranted in our de re belief ascriptions to animals – that is, our belief ascriptions that aim to identify the objects in the world the animal is thinking about independently of how it is thinking about them – since there are always countless different categories of objects about which the animal might be thinking and the only way to decide among them is to know what general background beliefs the animal holds, which, Davidson claims, is impossible without the animal being able to speak. Finally, Davidson's main argument is that if animals really did have beliefs, then they should be subject to surprises on those occasions when their beliefs turn out to be false; but being so surprised, Davidson maintains, involves being aware that one’s former belief failed to fit the facts and thus requires having an idea of a world of objective facts. Davidson goes on to claim that the only way for a creature to come to have an idea of a world of objective facts is by comparing its own beliefs with those of others, which it cannot do, according to the intensionality test and the argument from holism, if it cannot interpret the speech of others – as animals, Davidson maintains, plainly cannot. The upshot is that only creatures capable of understanding speech can have beliefs.

Detractors of Davidson’s arguments have tended to take one of three approaches. Some, such as Bermúdez (2003b), have sought to develop a theory of how to make principled de dicto ascriptions to animals; others, following Armstrong (1973), have sought to defend our de re ascriptions to animals by rejecting Davidson’s radical holism; and still others, such as Carruthers (2008) and Tye (1997), have sought to undermine Davidson's main argument by challenging its pivotal claim that surprise involves beliefs about beliefs.9

5. THE CONTEMPORARY DEBATE OVER ANIMAL CONSCIOUSNESS

Two general approaches to consciousness – the higher-order representational (HOR) approach and the first-order representational (FOR) approach – have played an important role in the philosophical debate over the status of animal consciousness. The dominant HOR approach has been the higher-order thought (HOT) theory of consciousness, according to which a mental

9 See Jamieson (chapter 1) for a critical commentary on Davidson’s arguments, as well as Lurz (2008) for a general list of such critical commentary.
state is conscious just in case the subject has (or is disposed to have) the higher-order thought that he is in such a mental state (Carruthers [2000]; Rosenthal [1986]).

The question of animal consciousness on this theory becomes the question of whether animals are capable of such higher-order thoughts.

A common argument against higher-order thoughts in animals is that the possession of such thoughts, in virtue of the mental-state concepts they contain, entails certain linguistic or mindreading capabilities that animals appear to lack. Bermúdez (2003a, and chapter 8, this vol.), for example, argues that since animals are incapable of speaking and interpreting a public language, they cannot possess concepts of propositional attitudes (e.g., beliefs and desires) and, therefore, cannot have higher-order thoughts about their own or others’ thoughts. On different grounds, Carruthers (2000) has argued that if animals really do possess mental-state concepts, then they ought to be able to apply these concepts to other animals for the purpose of anticipating and manipulating their behaviors – that is, they ought to be able to engage in mindreading. Carruthers goes on to argue, however, there are no incontestable empirical cases of mindreading in animals and some studies appear to show that not even the chimpanzee – the animal most likely to engage in mindreading – is capable of it (see, for example, Penn and Povinelli [2007b]; Povinelli [1996]). Both of these arguments, however, have been challenged recently. Lurz (2007) has raised a series of objections to Bermúdez’s argument against propositional attitude mindreading in animals, and Gennaro (chapter 10) and DeGrazia (chapter 11) argue that a number of recent studies of animal mindreading and metacognition strongly indicate that apes, monkeys, and dolphins are capable of higher-order thoughts about their own minds and the minds of other animals.

In contrast to HOT theories of consciousness, FOR theories hold that mental states are conscious nor because the subject is higher-order aware of having them but because the states themselves make the subject aware of the external environment (Dretske [1995]; Tye [1997]). Mental states that make subjects aware of the environment do so, according to FOR theories, in virtue of their having (or being poised to have) an effect on the subjects’ belief-forming system. FOR theorists argue that many varieties of animals, from fish to bees to chimpanzees, form beliefs about their environment based upon their perceptual states and bodily sensations and, therefore, enjoy conscious perceptual states and bodily sensations.

FOR theories are at their best, it has been argued, when explaining the consciousness of perceptual states and bodily sensations but have difficulty explaining the consciousness of beliefs and desires (Lurz [2004]). Some FOR theorists have responded by endorsing a HOT theory for conscious belief and desire and a FOR theory for perceptual and sensory states (see Tye [1997]; Dretske [2000], p. 188). However, such a hybrid view would appear to belie the FOR theory’s claim to parsimony and antecedent plausibility regarding ascriptions of consciousness to many lower animals. For it seems rather reasonable to suppose that a perceptual state or bodily sensation would count as conscious only if it has an effect (or is poised to have an effect) on the subject’s conscious belief-forming system. But if that is so, then a creature will have conscious states on this hybrid view only if it is capable of forming higher-order thoughts about its own beliefs. Not only would such a view require higher-order thought for consciousness, as HOT theories do, it would be more demanding than HOT theories by requiring animals capable of conscious perceptual states and bodily sensations to have higher-order thoughts about their own beliefs. What appears to be needed here in order to save FOR theories from this problem is a non-higher-order account of conscious belief (see Lurz [2006] for a sketch of such an account).

6. THE ROADMAP TO THE VOLUME

The essays in this volume are organized around topics as well as whether their focus is metaphysical or epistemological. The first two essays by Jamieson (chapter 1) and Saidel (chapter 2) are on epistemic questions concerning our ascriptions of thought (e.g., belief and desire) to animals. Jamieson proposes to resolve a tension that he identifies as existing between the belief that animals think and the belief that what they think cannot be characterized. After rejecting various eliminativist and realist approaches, Jamieson concludes that interpretivism – the view, roughly, that whether and what animals think is deeply connected to whether we find it useful to attribute thoughts to them – offers the best resolution. Saidel, on the other hand, puts forward a realist argument for beliefs and desires in animals based on evolutionary and empirical considerations. He argues that since human beings satisfy the behavioral criterion for belief-desire attribution

If scientists were to discover, for example, that the perceptual states involved in subliminal perception or blindsight actually caused subjects to form unconscious beliefs about their environment, none but the most committed FOR theorist would conclude from this alone that subliminal perception and blindsight were, after all, just cases of conscious perceptual awareness.