

8 August 2006

Davies, M. & Weiskrantz, L. Eds. *Frontiers of Consciousness*. Oxford University Press

BEAST MACHINES? QUESTIONS OF ANIMAL CONSCIOUSNESS

Cecilia Heyes

Department of Psychology

University College London

Gower Street

London WC1E 6BT

United Kingdom

e-mail: c.heyes@ucl.ac.uk

Phone: 020 7679 5378

URL: <http://www.psychol.ucl.ac.uk/celia.heyes/netintro.htm>

When the philosopher-scientist Rene Descartes coined the term ‘beast machines’ he is understood to have been denying that nonhuman animals have souls, minds or, in contemporary language, ‘consciousness’ (Descartes 1637; Huxley 1874). His views fuelled a debate that, not only continues to the present day, but may be more lively, more heated, and apparently more tantalisingly close to resolution, than at any other time.

This debate is often said to relate to ‘the question of animal consciousness’, but naturally it comprises many questions (Carruthers 2005; Griffin 1976). For example: What, in this context, is meant by ‘consciousness’? Why is the question of animal consciousness a focus of such broad interest? Can it be addressed through scientific studies of animal behaviour? What has been shown by scientific studies that appear to address the question of animal consciousness? This chapter examines these questions, focussing on the claim that animal consciousness is subject to scientific investigation.

1. What and why

Three senses of the terms ‘conscious’ and ‘consciousness’ are commonly distinguished. First, we often say that a person is conscious meaning that they are awake rather than asleep or in a coma. Consciousness of this kind is an intriguing phenomenon (Zeman, this volume), but it is not the focus of the animal consciousness controversy. No one doubts that the vast majority of animals can be conscious in this sense. Second, we sometimes say that a person is conscious of an object or event when we mean that they

are able to detect and react to the object or event. For example, if a wasp lands on my shoulder while I am speaking, I might say parenthetically that I am conscious of the wasp, meaning that I have detected the wasp and can deal with it appropriately. Again, there is no doubt that the term conscious can be applied to animals in this second sense. Most people would agree that a mouse sometimes is, and sometimes is not, conscious in this sense of a cat sneaking up on him.

The third sense of ‘consciousness’ is the one at the core of the animal consciousness debate, and it is famously difficult to capture or define. In this third sense, consciousness is a property, not of an individual, person or animal, but of a mental state – a percept, thought or feeling. A mental state that is conscious in this third sense, ‘phenomenally conscious’, is said to have ‘subjective’ properties. If a creature has phenomenally conscious mental states, then there is, in Nagel’s (1974) words, ‘something that it is like to be’ that creature (Davies, this volume).

A number of factors contribute to contemporary interest in the question whether animals are phenomenally conscious. At the most general level, a powerful motivation for much contemporary enquiry about phenomena consciousness, in human and nonhuman animals, is the desire to demonstrate that the scientific, naturalistic world view is comprehensive – that there is nothing that, in principle, eludes its explanatory reach. In this respect, phenomenal consciousness has the character of a final frontier. Investigating animal consciousness is part of this frontier project, and, in this area, as in many others,

the similarities and differences between humans and other animals can tell us about 'man's place in nature' (Huxley 1873).

More specifically, the question of animal consciousness is relevant to biomedical research. For example, it is questionable whether animals that are not phenomenally conscious can provide adequate substitutes for humans in research leading to the development of analgesic drugs. Turning from pragmatics to ethics, many philosophers and members of the public regard the question of phenomenal consciousness in animals as having substantial moral significance; strong implications regarding the way in which animals should and should not be treated, not only in biomedical research, but also in food production, as sources of labour, and as pets or 'companion animals'. Questions regarding animal welfare policy do not depend solely on the assumption that many animals are conscious, but it is undoubtedly an important consideration (Carruthers 2005, Shriver & Allen 2005).

Reference to companion animals reminds us that concern about animal consciousness is not only academic and political in nature, it is also deeply personal. We engage in a great deal of what Carruthers (2005) has called 'imaginative identification' with animals throughout our lives. In the style of Aesop, we ascribe to them human thoughts and feelings, and derive from our observations and engagements with them both moral lessons and emotional support. The possibility that, like Sony's robotic dog, Aibo, they are automata – impressively complex machines without phenomenal consciousness – puts

these sources of companionship under threat. It raises the spectre of a profound, species-wide, human loneliness.

2. Methods

The question whether animal consciousness can be investigated scientifically seems to be answered with a resounding ‘yes’ in the literature on animal behaviour. Many books and journals in comparative psychology, ethology and behavioural biology report empirical work probing the nature and distribution of consciousness in the animal kingdom.

Broadly speaking, this work is of four kinds, distinguishable by the combination of its empirical approach and underlying theory, or assumptions, about consciousness.

2.1 Empirical approaches

The majority of empirical studies of animal consciousness are based on analogical reasoning. They start with an assumption that humans are phenomenally conscious when they exhibit a specified behaviour, B. The focal behaviour is sometimes simple, such as withdrawing from a potentially damaging stimulus, and sometimes more complex and protracted, such as learning from experience. The study then demonstrates, typically through observation of spontaneously occurring behaviour, rather than by experiment, that animals of a particular species also exhibit behaviour B. It is then concluded that, when they exhibit behaviour B, members of the studied species are phenomenally conscious.

The other empirical approach, which is adopted more rarely in research concerned specifically with animal consciousness, uses experimental methods to test against one another two or more alternative explanations for a focal behaviour, B. One of these hypotheses suggests that the behaviour is mediated by phenomenally conscious states or processes, similar to those that mediate the behaviour in humans, whereas the other suggests that B is mediated by alternative, nonconscious states and processes. The method of testing alternative hypotheses is a pervasive feature of scientific investigation (Sober 2000).

Both of these empirical approaches, analogical reasoning and testing alternative hypotheses, depend on assumptions about the kinds of behaviour that are associated with phenomenal consciousness in humans. This is obvious in the case of reasoning by analogy, but it is also true of studies that test alternative hypotheses. Typically the alternative hypotheses each postulate functional states and processes, i.e. processes defined in terms of what they do within the information-processing system, but not in terms of how it feels to be in the states or engaging in the processes. The functionally defined states and processes are then classified as consciousness-related or consciousness-unrelated on the basis of assumptions about which functional states and processes are and are not associated with phenomenal consciousness in humans.

2.2 Theories of consciousness

There is no consensus, among scientists or philosophers, about the nature of the behaviour, or the functional states and processes, that are associated with phenomenal consciousness in humans. However, there are, broadly speaking, two schools of thought on this issue.

The first school asserts that, in humans, consciousness is associated with first order representation; that we are conscious of those sensations, emotions and thoughts that contribute, or have the potential to contribute, to the rational control of action (Dretske 1995, LeDoux, this volume; Tye 1997). Since rational control of action is traditionally understood to result from the interaction of beliefs and desires, this view implies that, when looking for evidence of animal consciousness, we should be looking for behaviour indicating that animals have beliefs and desires.

The other school of thought is more demanding. Commonly known as higher order theory, it suggests that we humans are phenomenally conscious only of those sensations, emotions and thoughts that we think about, or conceptualise (Carruthers 2000; Rosenthal 2005). According to this view, it is not enough for sensations and emotions to be capable of influencing beliefs and desires, they must themselves be objects of thought. Thus, for this camp, behaviour indicative of animal consciousness is behaviour indicative of higher order thought; of thinking about mental states. Higher order theory is consistent with the

approaches to the investigation of human consciousness pursued by Frith (this volume) and Milner (this volume).

2.3 Four methods

Although students of animal behaviour rarely make explicit reference to any theory of consciousness, their assumptions tend to conform to one of the two positions outlined above. When these are combined with the two empirical approaches, four methods of investigating animal consciousness are discernable: analogical reasoning based on the assumption that phenomenal consciousness is a property of first order representations; testing alternative hypotheses while assuming that first order representations are conscious, analogical reasoning based on the assumption that phenomenal consciousness is a property of the content of higher order thought, and testing alternative hypotheses while assuming that conscious content is the content of higher order thought.

Many studies using the first of these methods – analogical reasoning with first order representation - are reviewed in Griffin's early work on 'animal awareness' (1976; 1984) and in Tye's more recent analysis of 'simple minds' (1997). Taking one example, Tye refers to a study by Russell (1934) in which he observed the responses of sticklebacks to food placed in a jar at the bottom of their tank. Initially the fish attempted to seize the food through the glass, but after a while they nearly all learned to solve the detour problem: to move away from the food, up the side of the jar, and over the rim to the inside. This study, among others, is taken by Tye to indicate that fish can learn and

‘make cognitive classifications’, which, according to Tye’s first order theory, is sufficient to qualify them as creatures with phenomenal consciousness.

Studies of the first kind have been conducted with great care in scientific institutions, by people with scientific training, and published in scientific journals. Thus, according to institutional criteria, they are scientific investigations of animal consciousness. But it is strange to regard them in this way because scientific institutions are largely incidental to their genesis and development. One does not need scientific training, or a laboratory, or the peer-review processes of scientific publication, to observe animal behaviour suggesting, via analogical reasoning, the presence of phenomenally consciousness mental states. The fourteenth century Scottish king, Robert the Bruce, needed none of these to attribute perseverance to the spider that tried repeatedly to cast his thread from one beam to another (Barton 1912). Arguable, analogical reasoning based on first order theory is the method that we all use spontaneously in our day-to-day dealings with animals, in an attempt to understand and to anticipate their behaviour.

Because the first method is not distinctively scientific, and the focus of the present discussion is the question whether animal consciousness can be investigated scientifically, research using this method will not be discussed further. Examples of each of the other three methods are considered below.

3 Examples

3.1 Alternative hypotheses / first order representation

In everyday life, we commonly assumed that nonhuman animals have beliefs and desires. When we see a cat prowling around its food bowl at feeding time, we naturally assume that the cat wants (or desires) some food, and that he believes that prowling in the vicinity where food is delivered will, if not make it come sooner, then ensure that he can tuck-in as soon as possible after it arrives. But this is not the only plausible explanation for the cat's prowling behaviour. Instead, he may simply have developed a blind habit of approaching his dish when he is in a physiological state of hunger. According to this hypothesis, the cat does not believe that approaching the dish will hasten the moment when he can eat, or indeed have any thoughts about what the outcome of his behaviour might be. He is just drawn towards the dish when hungry because that behaviour has been rewarded in the past; a stimulus-response link, or habit, has been established (Heyes & Dickinson 1990).

Dickinson and colleagues have tested the belief-desire hypothesis against the habit hypothesis in experiments, not with cats approaching dishes for food, but with rats pressing levers for food (Balleine & Dickinson 1998). The seminal experiment of this kind, by Adams and Dickinson (1981), had three stages. In the first, rats pressed a lever and received two kinds of food, A & B. A pellet of food A was delivered whenever the rat pressed the lever. It was the outcome of that action. Food B was delivered equally

frequently but free of charge; its delivery did not depend on what the rat was doing. Over time, the frequency of lever presses went up. Was this because the rats had formed a belief that lever pressing caused the delivery of food A, and desired food A?

To find out, in the second stage of the experiment, Adams and Dickinson gave half of the rats an experience that would devalue food A. They allowed them to eat some of A and then injected them with lithium chloride (LiCl), an emetic. Rats typically avoid eating foods that they have eaten just before administration of LiCl. The other half of the rats, the controls, ate food B, the free food, before being injected with LiCl. In the final stage, several days later when they were no longer nauseous, all rats were given a test in which they could press the lever but this action had no outcome; neither food A nor food B was delivered.

If stage 1 experience had established lever pressing as a blind habit, one would not expect devaluation to affect the rats' propensity to press the lever. If the behaviour was based on a habit rather than a belief about the outcome, then changing the value of the outcome, should not change the behaviour. In contrast, if the rats had formed a belief in stage 1 – the belief that lever pressing causes the delivery of food A – then devaluation of food A should reduce their propensity to press the lever. Believing that lever pressing causes A, and *not* desiring A, they should press the lever on test less than the controls who, by hypothesis, believe that lever pressing causes A and *do* desire A. In accordance with the second hypothesis, the results showed that the rats averted to A in stage 2 pressed the lever less often in stage 3 than the controls, who had been averted to food B in stage 2.

This is one example of many experiments by Dickinson and colleagues testing belief-desire hypotheses against habit-drive hypotheses. Others, focusing on the desire versus drive distinction, show something remarkable: that animals are not merely driven to seek what their body needs – food when hungry, fluids when thirsty – but must experience the commodity, the food or drink, in the relevant physiological need state before it can function as an incentive (Balleine & Dickinson 1998).

In all of these experiments, the belief-desire hypothesis provides a better explanation for the rats' behaviour, it renders the observations more probable, than the alternative habit-drive explanation. Therefore these experiments may be said to provide evidence of beliefs and desires in rats. There is a certain irony in this. In the minds of many people, lever pressing rats represent behaviourism, the school of psychology that sought to deny mental life of any kind to all animals, human and nonhuman. Yet it is this 'preparation', this animal/procedure combination, that has yielded some of the strongest evidence of mentality, and specifically of first order representation, in animals.

3.2 Analogical reasoning / higher order thought

In 1970 evidence began to emerge that chimpanzees are able to use mirrors to detect marks on their bodies. The most compelling evidence of this capacity, which is known as 'mirror self-recognition', has come from experiments using a mark test (Gallup 1970). In this procedure, chimpanzees are first exposed to a mirror for a period of days. Then they are anaesthesia and, while under the anaesthetic, marked with an odourless, non-irritant

dye, typically on an eyebrow ridge and on the opposite ear. After recovery from the anaesthetic, the chimpanzees touch the marks on their heads more often when a mirror is present than when it is absent. The results of many mark tests could be due to an anaesthetic artefact; being sleepier in the first, mirror absent, observation period, the chimpanzees may have been less active in general than during the second, mirror present, observation period (Heyes 1994). However, this problem was overcome in a more recent study (Povinelli et al 1997), and now there is no reason to doubt that chimpanzees can detect marks on their bodies using a mirror.

It is often claimed that mirror self-recognition is indicative of higher-order thought, and the leading figure in the field, Gordon Gallup, claims explicitly that it implies *conscious* higher order thought (Gallup 1970, Keenan, Gallup & Falk, 2003). The reasoning behind these claims has never been articulated, but it seems to be roughly as follows. 1) When I (a human) use my mirror image, I understand the image to represent my 'self', and I understand my self to be an entity with thoughts and feelings. 2) This chimpanzee uses his mirror image. 3) Therefore this chimpanzee understands his mirror image to represent his 'self', an entity with thoughts and feelings

There is something seductive about this line of reasoning, perhaps because the mirror is a recurrent metaphor for the soul. However, on reflection, the argument is not compelling. Humans indicate by what they say that they understand a mirror image to represent or 'stand for' their body; to be in some sense a symbol. It is not clear that such a rich understanding of the relationship between object and image is necessary for use of a

mirror image as a source of information about the body. For example, it may be sufficient to know that changes in the appearance of the image regularly co-occur with changes in the body. One could learn this by performing exploratory movements while looking at one's mirror image, and exploit the information to guide body-directed movement, without conceptualising the image as a representation. Therefore, it is quite possible that, although chimpanzees can use a mirror to detect marks on their bodies, they either do not understand the image to be a representation at all, or they understand it to be a representation of a body – a body they care about – but not necessarily a body that thinks. In other words, mirror self-recognition need not involve higher order thought.

Research on mirror self-recognition is based on analogical reasoning, and therefore the hypothesis that it depends on higher order thought has not been explicitly tested against leaner alternatives. However, indirect evidence in support of a leaner interpretation comes from studies showing that the probability that a chimpanzee will pass the mark test declines with age (Heyes 1995, de Veer et al 2002). Nearly 90 percent of very young chimpanzees, 1-5 years old, pass the test, but the proportion declines in older age groups until, in mid- to late adulthood, only 40-50 percent of chimpanzees are successful. If mirror self-recognition were based on conceptualisation of the mirror image as a representation, and the 'self' as a thinker, one would expect success on the mark test to be low in infancy and to rise with age, as chimpanzees undergo cognitive development. The observed pattern suggests, in contrast, that mark test performance depends on how much a chimpanzee makes spontaneous movements when looking in the mirror, and thereby

gives itself the opportunity to learn that felt changes in the body co-occur with viewed changes in the image.

3.3 Alternative hypotheses / higher order thought

Studies of metamemory seek evidence that animals, typically monkeys, know when they remember, i.e. that they are capable of higher order thought about their memory states. Most studies of this kind (see Smith, Shields & Washburn, 2003, for a review) use an ingenious ‘commentary key’ method, proposed by Weiskrantz (1986, 1995), and pioneered by Cowey and Stoerig (1995)¹. Two experiments on metamemory will be described. The first is not very successful because it relies on analogical reasoning, but its weaknesses highlight the very considerable strengths of the second experiment, which tests alternative hypotheses.

On each trial in the first experiment (Hampton, Zivin & Murray 2004), food was placed in one of four opaque, horizontal tubes. On some trials the monkey could see this baiting operation, and on other trials his view of baiting was occluded by a screen. After baiting, the monkey was given access to the tubes and allowed to select one of them to search for food. If the selected tube contained the food, the monkey was allowed to eat it, but if the selected tube did not contain the food, the monkey ended the trial empty handed. The researchers recorded when and how often the monkeys looked down the tubes before selecting one to search, and found that seven out of nine monkeys looked less often when they had seen baiting than when their view of the baiting operation had been occluded.

Analogical reasoning tempts one to interpret this result as evidence of metamemory. For example, the monkeys' behaviour may be seen as analogous to that of a person who wants a telephone number (Hampton 2005). In this situation, we often ask ourselves whether we can remember the number and, if we cannot summon a clear mental image of the sequence of digits, look it up in a directory. This way of interpreting the monkeys' behaviour is enticing, but it is not justified by the data because they can be explained by several alternative hypotheses. For example, the monkeys could have been using the occluding screen (a public perceptual cue) or their own dithering behaviour (a public motor cue) as the basis for the decision whether or not to look. In other words, each monkey may have learned, during the experiment or previously, that in choice situations it is a good idea to look before you leap *if* you were recently behind a screen, or find your arm hovering between the alternatives. These hypotheses account for the experimental findings while assuming that monkeys monitor, not private mental states, but states of the world that are observable to all.

The second experiment (Hampton 2001, Experiment 3) overcame the public/private problem using a procedure that, in effect, asks a monkey to report whether he remembers a picture he saw recently. At the beginning of each trial, the monkey was shown one of four pictures on a computer screen. (A new set of four pictures was used in each test session.) After picture presentation, there was a delay, of variable duration, in which the screen was blank. After the delay, the monkey was usually required to touch one of two flags on the screen. Touching the 'test flag' resulted in the monkey being presented with a display containing all four pictures. If he selected from this array the picture he had

seen at the beginning of the trial, he was rewarded with a peanut, a preferred food.

Touching the other, 'escape flag', resulted in the monkey being given a lesser reward, a pellet of ordinary primate diet, but without him taking the test.

Hampton (2001) found that the frequency with which the monkey chose the escape key over the test key increased with the duration of the delay since the original picture was presented. If he had been reasoning by analogy, Hampton may have stopped there, but in reality he recognised that this pattern of behaviour need not be due to the monkey reflecting on the quality of his memory. The target image was not present when the monkey made his choice between the flags, so his tendency to escape after longer delays could not have been due to use of a public motor cue; he could not have based his decision on his own dithering behaviour. However, the monkey could in principle have used a public perceptual cue, i.e. the duration of the delay since the original image was presented. He may have learned that test taking has a happy result when the trial started a short time ago, and an unhappy result when it started a long time ago. He may have no idea why this is the case, that delay influences memory quality.

To test this lean hypothesis against the metamemory hypothesis, Hampton compared the accuracy of the monkey's performance on trials, like those described above, when he chose to take the test, and on other trials when he was forced to take the test. At each delay duration, approximately one-third of trials were forced rather than chosen. If, in tests he has chosen, the monkey decides to take the test by assessing the quality of his memory, by asking himself whether he has a clear mental image of the picture, then he

should be less accurate on forced than on chosen test trials because the forced trials will sometimes be trials in which he does not have a clear picture. In contrast, if the monkey decides whether or not to take the test just by remembering how long it was since the trial started, he should be equally accurate on chosen and forced test trials because the two types are likely to include a similar mix of trials in which he does and does not have a strong memory. In fact, the accuracy data were consistent with the predictions of the metamemory hypothesis. When tests were chosen, accuracy was high and fairly constant across delay durations, but when the tests were forced, accuracy declined with delay. This is an intriguing result. In providing evidence that favours the metamemory hypothesis over an alternative, public cue hypothesis, this study provides scientific evidence of metamemory, of thinking about memory, in a monkey.

4. Conclusions

At least two conclusions can be drawn from the examples discussed above. First, studies based on analogical reasoning tell us about animal *behaviour*, but provide very little information about animals' *minds*, about the psychological states and processes, conscious or otherwise, that generate behaviour. For example, they show that chimpanzees can use mirrors to detect marks on their bodies, and that monkeys look for food when it is in their interests to do so. However, research based on analogical reasoning is no better than common sense or intuition in telling us what kind of thinking is behind these behavioural achievements. Second, the method of testing alternative hypotheses *can* provide information, not only about behaviour, but also about the

psychology of animals. In the examples above, this method showed that rats have beliefs and desires, and that monkeys can think about their memories.

Does the second of these conclusions imply that animal consciousness can be investigated scientifically, and that the products of this research have shown that rats and monkeys have phenomenally conscious mental states? Many psychologists and biologists would argue that it does, but to secure this conclusion one must make at least three major assumptions.

The first assumption is that further research will not undermine the consciousness-related hypotheses favoured by research to date. For example, the studies by Dickinson and his colleagues have shown that the belief-desire hypothesis (consciousness-related) provides a better explanation than the habit-drive hypothesis (consciousness-unrelated) for the instrumental behaviour of rats. Further experiments are very unlikely to require revision of this conclusion, but they may well favour a new consciousness-unrelated hypothesis over the belief-desire hypothesis. Indeed, Dickinson's (1994) own associative-cybernetic model of learning is a rich potential source of such a rival hypothesis. Thus, the first assumption is tenuous, but it is not unusual. Arguably all conclusions of scientific research rest on the same kind of assumption; they are all, in this sense, provisional.

The second assumption is more specific to the investigation of consciousness. To interpret the studies reviewed above as providing evidence of consciousness in animals, one must assume that one of the major theories of consciousness, the first order

representational theory or higher order theory, is broadly correct. Scientific conclusions typically depend on assumptions about the validity of background theories, but those theories are seldom as contentious as theories of consciousness. In addition to controversy over the question whether phenomenal consciousness is a property of first order or higher order mental states, there are many scientists and philosophers who deny the coherence and validity of the whole project that these theories represent. They argue that consciousness cannot be identified with any functionally defined states, or that such identification completely fails to explain the core features of consciousness, its subjective properties (Chalmers 2002; Nagel 1974; see Hurley & Noe for a contrasting view). This ‘explanatory gap’ (see Davies, this volume; Zeman, this volume), or the suspicion of such a gap, does not accompany other scientific ventures.

The third assumption is related to the second, and is perhaps the most interesting because it concerns the investigation of animal consciousness specifically. Even if we disregard worries about the explanatory gap, and assume that one or other of the functionalist theories of consciousness is correct *for humans*, the studies reviewed above provide evidence of animal consciousness only if we assume that it is also correct for nonhuman animals. In other words, we must assume that consciousness is a property of the same functionally defined states in humans and in other animals.

The presence and pervasive influence of language in humans, but not in other animals (Pinker 1994), is a major obstacle to this third assumption. The absence of language in other animals is sometimes regarded merely as an inconvenience, as a factor that makes it

hard to find out about animal consciousness, but which does not impact on the likelihood that animals are, in fact, conscious. But the absence of language in nonhuman animals does not only create a measurement problem. It also raises the fundamental question of whether creatures that do not have the potential to use language, and have not acquired, through language, the knowledge and cognitive skills of human culture, can have mental states that are phenomenally conscious (MacPhail 1998).

Some theories of consciousness claim explicitly that it is dependent on language (Dennett 1991). Others, such as higher order theory, are more cryptic. They are formulated primarily to account for human consciousness, and they describe the higher order thoughts that render mental content phenomenally conscious as if those thoughts were propositions or 'sentences in the head' (e.g. Carruthers 2002; Rosenthal 2005). However, they do not consider in any detail what kind of non-linguistic, non-propositional mental structure, if any, could fulfill the same higher-order function.

Consider, for example, the monkey that passed Hampton's (2001) test of metamemory. In spite of lacking natural language, he may have a 'language of thought', sometimes known as 'mentalese' (Fodor 1975). If so, the monkey could mentally represent his memory state as a component of a sentence in mentalese. If the monkey does not have a language of thought, then his test performance could be based on an unusual kind of contingency knowledge. It could be that he has encoded his task experience as pairs of mental images of his state, linked to mental images of trial outcomes. For example, an image of himself in state X (a state that we humans would characterize as having a weak

memory), combined with an image of himself touching the test flag, is linked with an image of an unhappy outcome, no food at the end of the trial. In contrast, an image of himself in state Y (a state that we would characterize as having a strong memory), combined with an image of himself touching the test flag, is linked to a representation of a happy outcome, the arrival of a peanut. This is a nonlinguistic, functionalist account of what was going on in the monkey's head during the metamemory test. If it were correct, would the monkey be engaging in higher order thought of the kind that has been associated with phenomenal consciousness in humans? Currently, higher order theories of consciousness do not offer an answer to this question (Shea & Heyes, in prep).

To summarise: When psychological processes are functionally defined – in terms of what they do, rather than how they feel - there is good reason to believe that at least some research, that which tests alternative hypotheses, provides reliable, scientific information about the psychological processes that mediate animal behaviour. To conclude from this that animal consciousness is subject to scientific investigation, one must assume 1) that further research will not favour consciousness-unrelated functional hypotheses over the consciousness-related hypotheses that are currently plausible, 2) that a functionalist theory of consciousness – first order or higher order - is broadly correct for humans, and 3) that the same functionally defined processes are phenomenally conscious in humans and in other animals.

The second and third of these assumptions require one to ignore the explanatory gap, and to take it largely on faith that consciousness does not depend on language. Personally, I

find these demands too great, and therefore doubt that animal consciousness is currently the subject of scientific investigation. However, I assume on nonscientific grounds that many animals experience phenomenally conscious states – that they are not ‘beast machines’ - and I find it plausible that, at some time in the future, the presence and character of these states will be discoverable by scientific methods. To make that possible, we need stronger theories of consciousness; theories that close the explanatory gap, elucidate the relationship between language and consciousness, and that are grounded in the kind of empirical work on human consciousness reported elsewhere in this volume.

Notes

- 1 The seminal experiments of Cowey & Stoerig (1995) are not discussed here because they addressed a subtly but importantly different question. Here we are concerned with whether animals can engage in higher order thought. In contrast, Cowey & Stoerig assumed, plausibly and for methodological reasons, that monkeys engage in higher order thought, and asked whether this capacity is impaired in blindsight. Similarly, Leopold and Logothetis (1996) assume that monkeys are phenomenally conscious of the perceptual states that comprise binocular rivalry in order to investigate the neurological correlates of such states. They do not test this assumption.

Acknowledgements

I am grateful to the Warden and Fellows of All Souls College, and to the McDonnell Centre for Cognitive Neuroscience, for their hospitality and support during the period in which this chapter was written. I wish also to acknowledge the ESRC centre for Economic Learning and Social Evolution (ELSE), the Biotechnology and Biological Sciences Research Council, and a grant awarded by the European Commission under their 'What it means to be human' initiative (Contract number: 012929)

References

Adams, C.D. & Dickinson, A. (1981). Instrumental responding following reinforcer devaluation. *Quarterly Journal of Experimental Psychology*, 33B, 109–122.

Allen, C. (2005) Animal Consciousness. *The Stanford Encyclopedia of Philosophy (Winter 2005 Edition)*, Edward N. Zalta (ed.),
<<http://plato.stanford.edu/archives/win2005/entries/consciousness-animal/>>.

Balleine, B. & Dickinson, A. (1998) Consciousness: the interface between affect and cognition. In J. Cornwell (Ed.), *Consciousness and Human Identity*. Oxford University Press, Oxford.

Barton, B. (1912). Bruce and the Spider. Reprinted from *Historic Ballads and Poems*. Rupert S. Holland (ed.). Philadelphia: George W. Jacobs & Co.

Carruthers, P. (2000) *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge University Press.

Carruthers, P. (2002) The cognitive functions of language. *Behavioral and Brain Sciences*, 25, 657-726.

Carruthers, P. (2005) Why the question of animal consciousness might not matter very much. *Philosophical Psychology*, 18, 83-102.

Chalmers, D. (2002) Consciousness and its place in nature. In S. Stich & T. Warfield (Eds.), *Blackwell Guide to the Philosophy of Mind*. Blackwell.

Cowey, A. & Stoerig, P. (1995) Blindsight in monkeys. *Nature*, 373, 247-249.

Dennett, D. C. (1991). *Consciousness Explained*. London: Penguin Press.

Descartes, René. (1637) *Discourse on Method*. Vol. XXXIV, Part 1. The Harvard Classics. New York: P.F. Collier & Son, 1909–14; Bartleby.com, 2001.

de Veer, M. W., Gallup, G. G., Theall, L. A., van den Bos, R. & Povinelli, D. J. (2002) An 8-year longitudinal study of mirror self-recognition in chimpanzees. *Neuropsychologia*, 1493, 1-6.

Dickinson, A. (1994) Instrumental conditioning. In N. J. Mackintosh (Ed.), *Animal Learning and Cognition*. Handbook of Learning and Cognition, Second Edition. Academic Press.

Dretske, F. (1995) *Naturalizing the Mind*. MIT Press.

Fodor, J. A. (1975) *The Language of Thought*, Cambridge, Massachusetts: Harvard University Press.

Gallup, G. G. (1970) Chimpanzees: self-recognition. *Science*, 167, 86-87.

Griffin, D. R. (1976) *The Question of Animal Awareness*. Rockefeller University Press, New York.

Griffin, D. R. (1984) *Animal Thinking*. Harvard University Press, Cambridge, Mass.

Hampton, R. R. (2001) Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences*, 98, 5359-5362.

Hampton, R. R. (2005) Can rhesus monkeys discriminate between remembering and forgetting? In H. S. Terrace & J. Metcalfe (Eds.), *The Missing Link in Cognition*. Oxford University Press, Oxford.

Hampton, R. R., Zivin, A. & Murray, E. A. (2004) Rhesus monkeys discriminate between knowing and not knowing and collect information as needed before acting. *Animal Cognition*, 7, 239-246.

Heyes, C. M. (1994) Reflections on self-recognition in primates. *Animal Behaviour*, 47, 909-919.

Heyes, C. M. (1995) Self-recognition in primates: Further reflections create a hall of mirrors. *Animal Behaviour*, 50, 1533-1542.

Heyes, C. M. & Dickinson, A. (1990) The intentionality of animal action. *Mind & Language*, 5, 87-104.

Hurley, S. & Nöe, A. (in press) Neural plasticity and consciousness. *Biology and Philosophy*.

Huxley, T. H. (1973) *Man's Place in Nature*. D. Appleton & Co.

Huxley, T. H. (1874) On the hypothesis that animals are automata, and its history. *Fortnightly Review*, 16, 555-580.

Keenan, J. P., Gallup, G. G. & Falk, D. (2003) *The Face in the Mirror: The Search for the Origins of Consciousness*. Ecco/Harper Collins.

Leopold, D. A. & Logothetis, N. K. (1996) Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, 379, 549-553.

Macphail, E. M. (1998) *The evolution of consciousness*. Oxford University Press, Oxford.

Nagel, T. (1974) What is it like to be a bat? *Philosophical Review*, 83, 435-450.

Pinker, S. (1994) *The Language Instinct*. Penguin.

Povinelli, D. J., Gallup, G. G., Eddy, T. J., Bierschwale, D. T., Engstrom, M. C.,
Perilloux, H. K., & Toxopeus, I. B. (1997) Chimpanzees recognize themselves in mirrors.
Animal Behaviour, 53, 1083-1088.

Rosenthal, D. (2005) *Consciousness and Mind*. Oxford University Press.

Russell, E. (1934) *The Behavior of Animals*. London: Edward Arnolf.

Shriver, A. & Allen, C. (2005) Consciousness might matter very much. *Philosophical
Psychology*, 18, 103-111.

Smith, J. D., Shields, W. E. & Washburn, D. A. (2003) The comparative psychology of
uncertainty monitoring and metacognition. *Behavioral and Brain Sciences*, 26, 317-373.

Sober, E. (2000) Evolution and the problem of other minds. *Journal of Philosophy*,
XCVII, 365-386.

Tye, M. (1997) The problem of simple minds: is there anything it is like to be a honey
bee? *Philosophical Studies*, 88, 289-317.

Weiskrantz, L. (1986). *Blindsight: A case study and implications*. Oxford: Clarendon
Press.

Weiskrantz, L. (1995) The problem of animal consciousness in relation to neuropsychology. *Behavioural Brain Research*, 71, 171-175.