

CHAPTER II

Self-awareness in animals

David DeGrazia

I. INTRODUCTION

Many animals are self-aware. At any rate, I claim, the cumulative force of various empirical data and conceptual considerations makes it more reasonable to accept than to deny this thesis. Moreover, there are importantly different sorts of self-awareness. If my arguments are on the right track, then scientists and philosophers have significantly underestimated the case for animal self-awareness.

2. TYPES OF SELF-AWARENESS

The most primitive type of self-awareness is *bodily self-awareness*, an awareness of one's own body as importantly different from the rest of the environment – as directly connected with certain feelings and subject to one's direct control. Because of bodily self-awareness, one does not eat oneself. And one pursues certain goals. Bodily self-awareness includes *proprioception*: an awareness of body parts, their position, their movement, and overall body position.¹ It also involves various *sensations* that are informative about what is happening to the body: pain, itches, tickles, hunger, as well as sensations of warmth, cold, and tactile pressure. These forms of awareness are essential to any creature that can feel features of its body and environment and act

Thanks to Robert Lurz, Marc Hauser, and my colleagues in the Department of Philosophy – especially Tad Zawidzki and Eric Sidel – for feedback on a draft.

¹ For an outstanding discussion of proprioception and its relationship to self-awareness, see Bermúdez (1998, chapter 6). Addressing how self-awareness is possible without language, the book gradually develops a case – drawing primarily from developmental psychology – that the most sophisticated forms of self-awareness can be built up from primitive beginnings. Thus, Bermúdez represents a notable departure from common assumptions about self-awareness challenged in this essay. Notably, Robert Mitchell has also emphasized different types of self-awareness among animals, but he apparently assumes that there are literally different *selves* associated with a given individual (e.g., Mitchell 1994). By contrast, I maintain that (at least in non-pathological cases) there is just one self, the individual, who may be self-aware in various ways.

appropriately in response. In sum, bodily self-awareness includes both an awareness of one's own bodily condition and an awareness of one's *agency*, of moving around and acting in the world. Somewhat radically, I suggest that most or all sentient animals have this type of self-awareness.

Social self-awareness – awareness of oneself as part of a social unit with differing expectations attaching to different positions – is present in highly social creatures. It enables such animals to interact with each other effectively. By understanding the expectations that come with one's position, and the ways in which particular interactions among group members affect how one can best deal with them, an animal improves her chances of surviving and passing along her genes. Wolf X, for example, understands that he is subordinate to wolf A, the alpha, and that wolf B has recently formed an alliance with A, so X had better not attack B for fear of A's retribution. Social self-awareness in animals presupposes bodily self-awareness insofar as deliberate social navigation is possible only in creatures aware of their own agency.

Introspective awareness is awareness of (some of) one's own mental states such as feelings, desires, and beliefs. Is this phenomenon exclusive to language-users? After all, it requires not just having mental states, but awareness of having them; one might suppose that such mental reflexivity requires the conceptual rocket of language. On the other hand, assuming a rabbit can be hungry – can *have* the sensation of hunger – it may be plausible to hold that the rabbit is also *aware that she has the sensation*. Indeed, insofar as bodily self-awareness rests partly on having various sensations, and noting their connection with one's body, bodily self-awareness may implicate a basic sort of introspective awareness. I leave that possibility open. As we will see, there is independent evidence from metacognition studies involving monkeys that certain non-linguistic creatures are introspectively aware.

Let us turn to the arguments.

3. DESIRES AND INTENTIONAL ACTION

Many animals have *desires*. That is, they *want* certain things such as food, refuge, or access to a mate. Given a choice between two substances to eat, or two places to sleep, they often *prefer* one to the other. The thesis that desire abounds in the animal kingdom seems strongly supported by common sense. But further support is available.

There is a strong case that all animals capable of having pleasant and unpleasant experiences – let's reserve the term *sentient animals* for

them – have desires. To find X pleasant entails, *ceteris paribus*, wanting that the experience of X continue. To find Y unpleasant entails, *ceteris paribus*, wanting the experience of Y to discontinue. Hence a conceptual connection between desires and hedonically valenced experiences, assuming many animals have the latter, provides a good reason to believe they have desires.

Their behavior also suggests that many animals have desires. Why does your dog zoom into the kitchen when she hears you pouring food into her dish? Presumably, because she wants to eat. Why does she jump excitedly and head to the back door, where the leash is, when you look at her and say it is time for a walk? Presumably, because she wants to go for a walk. Appeals to animal behavior as evidence for desires, however, must be advanced carefully. Behavior *alone* might suggest that all animals have desires, but that inference would be unwarranted. We would be on questionable ground saying that the spider builds a web because it wants to or, worse, that the jellyfish follows its desire in swimming around. At least as I am using the term “desire,” one must be capable of conscious states, and in particular pleasant and unpleasant feelings, in order to have desires; unconscious desires are possible, but only in beings capable of having conscious desires.² So desire-like behavior requires independent evidence that the creature in question is sentient, for responsible attribution of desires. Here I simply assume that such independent evidence is available in the case of mammals, birds, and probably at least reptiles and amphibians.³

Let us now consider studies focusing on animals’ preferences. Marian Stamp Dawkins has studied *what* animals want in choice situations and *how much* they want it: “For instance, when a pigeon has learnt to peck a key for food, will it still keep pecking when instead of having to give just one or two pecks per item of food, it has to peck four, eight or even 50 times?” (Dawkins [1993], pp. 147–148). Of course, a pigeon can prefer pecking for food over resting with no food only if he *has* preferences or desires. Similarly for any animal who prefers to go into one enclosure over another. For example, when hens were offered a choice between standing on wire floors and standing on a floor of wood shavings, even those hens who had never before encountered the second sort of floor chose it, consistently, as soon as they had the option (*ibid.*, p. 153). Their preference or desire was evident.

² More formally, A desires X only if (1) A is disposed to bring X about, (2) this disposition is potentially conscious, and (3) A is disposed, *ceteris paribus*, to have pleasant feelings upon attaining X and unpleasant feelings at prolonged failure to attain X (DeGrazia [1996], p. 130).

³ See *ibid.*, chapter 5 for arguments and citations to empirical evidence.

Much behavior among sentient animals suggests desires. Much of this same behavior, I submit, is best understood as reflecting *beliefs* that, together with the relevant desires, produce *intentional action*. Your dog heads for the kitchen upon hearing you pour food into her dish not only because she wants to eat, but because she believes that by going to the right place she will be able to eat. Thus she intentionally heads for the kitchen. She goes to the back door when anticipating a walk not only because she wants one, but also because she believes going there in this situation will enable her to take a walk. So she intentionally heads there. To be sure, there are other ways of interpreting such familiar behaviors without attributing either desires or beliefs, much less intentional action. But these alternative interpretations seem strained in view of the evidence.⁴ Better to maintain that cognitive ethology – the study of animal behavior, in the context of evolutionary theory, that attributes intentional states such as beliefs and desires to animals (Jamieson and Bekoff [1993]) – is on the right track. If so, then a belief-desire model of intentional action (Davidson [1980]) supports the attribution of the latter to animals.⁵

But now we face an important objection. In the philosophy of mind, desires and beliefs are classified as *propositional attitudes*, mental states that take propositions or sentences as their objects. For example, in desiring food, I desire *that I eat food*. Similarly, I believe *that there is food in the kitchen*. But can my dog, or any nonlinguistic creature, mentally entertain such propositions? To do so, he would seem to need *concepts*. Does he really have the concepts of food, eating, kitchen, and so on? Presumably he doesn't have *our* concepts of food as nourishing, eating as applicable to all creatures with mouths, and kitchens as rooms used for cooking. But perhaps he has *his own* concepts that pick out these items even if the conceptual scaffolding differs from that of our concepts – and, from the building blocks of his concepts, we could in principle construct the relevant propositional attitudes. That's one possibility I would take seriously. But one might doubt it. One might reasonably suppose that possession of concepts requires capacities for abstraction that surpass non-linguistic beings. In that case, my dog, lacking concepts, would also lack desires and beliefs as propositional attitudes.

⁴ See DeGrazia (1996, chapter 6) for my full case.

⁵ I attribute the model, not its extension to animals, to Davidson. Eric Sidel persuasively argues that this model incontrovertibly applies wherever animals' behavior reveals an ability to consider novel means to an end or familiar means to a novel end (Sidel, this vol., chapter 2). Some of the examples I present meet even this conservative standard.

Suppose that's correct. We should still agree that behaviorist and stimulus-response interpretations of relevant animal behavior are less credible than interpretations informed by cognitive ethology. Thus, even if we don't attribute to animals full-blown desires and beliefs, which are concept-laden, we may in good epistemological conscience attribute to them *proto-desires* and *proto-beliefs*. These states, we may say, interact in producing *proto-intentional action*. The idea is that these mental states, though not conceptual, nevertheless have *content*.⁶ Their content is supplied by something like *generalized features of perceptions*. My dog's proto-desire for food will pick out relevant instantiations of food as things suitable for him to eat even if it does not employ a universal concept of food as stuff that nourishes.

Again, I attribute real desires and beliefs to many animals. But even if I am wrong on this point, I'm on solid ground in attributing at least proto-intentional states to them. And that will suffice for my arguments. (Although I will hereafter drop cumbersome references to *at least proto-intentional states*, let us bear the qualifications in mind.)

Desires to do certain things and intentional actions that involve doing them suggest at least some rudimentary awareness of oneself as persisting through time. Your dog's intentionally running to the back door with a desire to go for a walk requires that she represent herself as being around long enough to go outside. The very desire to do something, even if the action is obstructed, is similarly future-oriented and self-implicating. For the desire and intention amount to a rudimentary plan, which necessarily includes a representation of completing the intended action. If this is correct, then a commonsense appreciation of the ordinary behaviors of many animals suggests a kind of self-awareness – namely, bodily self-awareness, here with an emphasis on the agency aspect.

Strengthening the case for intentional action, and therefore for bodily self-awareness, is evidence of more sophisticated behaviors in animals involving planning, complex problem-solving, and/or tool use. In such cases, denying that animals perform intentional actions seems absurd, because the actions are so obviously deliberate. Consider some examples.

Chimpanzees use natural objects in pursuing certain goals – for example, moss as a sponge, rocks as nut-crackers, and stems to probe for insects (McGrew [1992], pp. 44–46). They have also been observed using sticks to cushion the soles of their feet when climbing or walking over thorns (Stanford [2001], p. 126). Such behaviors are clearly intentional, not to

⁶ See Bermúdez (1998, chapter 4) for a way of developing this idea.

mention intelligent. For those who reserve the term “tool use” for instances in which one *fashions* an inanimate object before employing it, several feats will make the grade. Chimpanzees have been observed doing all of the following (Hauser [2000], pp. 35–36): stripping leaves off sticks and inserting them into the homes of ants and termites, waiting for them to climb aboard only to become primate dinner; chewing leaves to create an absorbent sponge, which is used to soak up excess water or sap from tree holes; fashioning sticks to produce a sort of dental probe for an unusual method of grooming group members.

Dolphins also furnish examples of highly deliberate problem-solving. Some dolphins wear cone-shaped sponges over their beaks, apparently a protective measure as they nose along the bottom in search of food (Connor and Peterson [1994], pp. 195–196). They engage in cooperative hunting that is responsive to immediate circumstances (Mann *et al.* [2000]). There are even reports of dolphins apparently “asking” humans in the water for assistance in removing a fishing hook from a group member’s mouth or tail (White [2007], pp. 93–94).

Some examples come from birds, specifically New Caledonia crows, who fashion two different types of twigs to extract insects from different sorts of holes (Hauser [2000], p. 36). One crow was videotaped making a complexly shaped tool out of a wire:

Betty tried to obtain the food with the straight wire but when that proved impossible she took it away, wedged it in a crevice in her tray and bent it to the appropriate hook shape. Then she went over and used it as a tool. She even corrected the shape of the hook. (Anderson and Kacelnik [2004], p. 46)

Your own experience with pets may furnish an impressive feat of problem-solving (even if not tool use). My best example features my family’s Labrador retriever, who, apparently frustrated at being confined to the study, reared on her back legs and attempted to turn the doorknob.

Such planning and problem-solving are instances of intentional action. Again, intentional action is possible only if the animal agent has some sense of herself as persisting long enough to complete the action or plan. This sense of self involves, most basically, a sense of one’s own body as importantly distinct from the rest of the world and as subject to one’s direct control.

4. FEAR

Few will doubt that many animals experience the primitive emotion of fear. Like anger and sexual arousal, fear is associated with the sympathetic autonomic nervous system. This system facilitates action in what we may

broadly call *emergency* situations. Fibers in the system increase heart rate, sweating, and general arousal while decreasing digestion and other processes associated with rest. Also implicated is the limbic system, a group of neurological structures (e.g., the amygdala, hippocampus, thalamus) that are essential to motivation and emotion. Importantly, the sympathetic autonomic nervous system and limbic systems are evolutionarily primitive, common to all vertebrates. It seems responsible to assume that any creature that is endowed with these two systems, and sentient, can experience fear.

What is fear? It is an emotional response to the perception of danger, a response that facilitates attention to promote protective action. What is the object of fear? It is something one perceives to pose a threat of harm to oneself. When? Some time in the future. These mundane implications of the concept of fear suggest that those who can be afraid have some sense of themselves as persisting into the (possibly very near) future. That is, anyone who can fear has at least a rudimentary bodily self-awareness. Moreover, if the subject who fears perceives the harm threatened as *being hurt*, this would entail an awareness of the possibility of having pain in the future, a type of introspective awareness. These basic points suggest that an enormous range of animals are self-aware in some way and to some degree.

5. ANTICIPATION OF ONE'S OWN FUTURE

Like intentional action involving a plan, fear requires some awareness that one will continue into the future. Is there independent evidence that animals can anticipate their own futures?

Note that, from an evolutionary standpoint, a sense of time would be highly adaptive for creatures capable of complex behaviors in a changing environment. Anticipation is useful for getting a jump on predictable events and selecting behaviors accordingly. Moreover, anticipation would presumably work in conjunction with memory. Animals who can anticipate the movement of prey and predators, based partly on memory of their past behavior in similar circumstances, would enjoy a major advantage in determining what to do.

A skeptic might reply, however, that what is adaptive is the capacity to *encode* information gained from experience and *use* that information in modifying future behavior. There is no additional adaptive value, the challenge continues, to representing that information *consciously*, as would be required for any forms of memory or anticipation relevant to self-awareness.

This objection misfires, for two reasons. First, there is good reason to think some self-representing memories and anticipations are conscious.

There is probably additional adaptive value to being able to represent one's own past and future consciously. In humans, the ability to manage complexity and novelty, to improvise in unfamiliar circumstances, is associated with conscious mental states, whereas unconscious information-processing often suffices in familiar terrain. But, again, animals must deal with variable environments: moving predators and prey, changing weather and food supplies, and – in social species – evolving social dynamics. So conscious anticipation and remembering would be advantageous in novel or highly variable situations. Second, even if the relevant self-representations were all unconscious, they would still manifest self-awareness. There is no reason to require that self-representations be conscious to count as manifesting a rudimentary self-awareness (at least in creatures conscious of some features of their world through pleasure, pain, proprioception, and external senses such as vision and hearing).

In addition to evolutionary considerations, there is specific empirical evidence suggesting certain animals' ability to anticipate their own futures. Consider two examples. In one, squirrel monkeys were initially given a choice between one piece of date (a fruit they like) and four. Naturally, they took four. Then the scientists began to withhold water – for three hours if the monkeys chose four pieces of date, for thirty minutes if they chose one piece. The monkeys learned to anticipate the consequences of their choice. Although not thirsty when choosing, they anticipated becoming thirsty and chose a smaller bounty of food in order to drink more readily when thirsty (Zimmer [2007]). In another study, scrub jays were permitted to feed freely during the day for six days, except in the morning, when they were confined either in a compartment where they got breakfast or in another where they did not. On the evening of the seventh day, the jays received extra food. Although not then hungry, they stashed the surplus in the compartment where they had learned they would not receive breakfast, anticipating the possibilities for the next morning (Raby *et al.* [2007]). Importantly, in these instances, the animals not only anticipated future events, but anticipated their own situation in the future – being thirsty or hungry or not – providing further evidence of bodily (and perhaps introspective) self-awareness.

6. MEMORY OF ONE'S OWN PAST

Consider now some data suggesting animals' awareness of their own recent behavior. Researchers trained a dolphin to understand a particular gesture as meaning "repeat": do again what you just did before. The dolphin was able to execute this command, repeating his immediately preceding action

when this command was given but not when other commands were given (Mercado *et al.* [1998]). In an earlier study, rats were trained to press one of four levers right after hearing a buzzer, the correct response depending on their behavior – immobility, face-washing, walking, or rearing – at the time of the buzzer. The rats demonstrated their ability to discriminate among their immediately preceding behavior types (Beninger *et al.* [1974]). Awareness of what one just did would seem to implicate memory and bodily self-awareness.

Also of interest are studies in which researchers focused on the possibility of *episodic* memories. The latter involve conscious recollection of *experiences from one's past*. Implicit memories, by contrast, are stores of information based on past experiences, where those experiences are not consciously recollected, though the information can influence present behavior. You may remember (having seen) the face of a character in a movie without remembering when you saw it, what the movie was, or whose face it is – implicit memory. If you recall the experience of watching the movie, you have an episodic memory.

Now consider another experiment involving scrub jays, who prefer to eat moth larvae rather than peanuts if the larvae are fresh but prefer peanuts if larvae have been dead a few hours (de Kort *et al.* [2005]). The jays were given a chance to hide both kinds of food, and then were removed to a cage. The birds kept away from their caches for four hours tended to dig up larvae whereas those who had to wait five days ignored the larvae and dug up peanuts. (The experiment was controlled to rule out the hypothesis that the birds were following their sense of smell.) If jays appear to have episodic memory, it should be unsurprising that other studies suggest that mammals, including rats (Babb and Crystal [2005]) and gorillas (Schwartz *et al.* [2005]), do as well. Since any such episodic memories would be memories of having done something, or having had a particular experience, they would suggest at least bodily self-awareness.

Does room remain for skepticism? What if the jays, for example, simply remembered *that* they hid food a short time ago or a long time ago without remembering *doing so*? This would be implicit rather than episodic memory. Nevertheless, it would involve an awareness of something one did in the past, manifesting bodily self-awareness.

7. IMITATION

In *imitation*, one intentionally does what someone else has done. More precisely, one individual learns from another *the form of a particular behavior* and copies it. Imitation differs from *goal emulation*, in which one learns

from another a particular goal to pursue; from *observational conditioning*, in which one learns from another in what circumstances to apply a behavior already in one's repertoire; and from *stimulus enhancement*, in which one learns from another what in the environment to attend to, leading to one's discovery of an action that resembles that performed by the other (Whiten and Ham [1992]). In genuine imitation, one's intention implies some representation of oneself. For the imperative, "Do what that individual did," has an implicit subject: oneself. There is no claim here that the intention and associated representations are linguistic, nor even that they're conscious, just that whatever form they may take, their contents imply an awareness of oneself as an agent capable of acting in the same way.

Convincing instances of imitation include the following. An orangutan was observed using kerosene to start a fire and a trash can lid to tend it after observing the same actions by a human (Russon and Galdikas 1993). The chimpanzee Washoe "adopted" a young chimp who eventually mastered thirty-nine signs of sign language, without human instruction, by imitating Washoe (Fouts *et al.* [1984]). Chimpanzees raised in homes have apparently imitated a plethora of actions performed by caretakers (for a list see Whiten and Ham [1992], pp. 263–264). The same is true of Chantek, a language-trained orangutan, who imitated many signs and actions by the time he was two – including in response to the sign "DO SAME" (Miles [1993], p. 49). Dolphins, meanwhile, have an extraordinary capacity to imitate the actions and postures of conspecifics, humans, and seals as well as human speech (Connor and Petersen [1994], pp. 188–91; Herman [2002], pp. 277–278). Perhaps the most remarkable instance occurred when two captive dolphins who had been trained to perform for audiences were accidentally put in each other's shows – which had different cues and required different actions. One performed the other's show correctly, without training, based entirely on having observed the other dolphin's training (discussed in White [2007], pp. 88–90). Whether animals other than apes and dolphins can imitate is uncertain.

8. SELF-RECOGNITION WITH MIRRORS

Since Gordon Gallup's pioneering experiments in the 1970s, self-recognition with mirrors has often been cited as evidence of self-awareness in animals. Before considering those experiments, let us note that mirror self-recognition involves more than perceiving oneself in a mirror. Any dog, for example, can perceive an image in the mirror; and when the image is of herself, she can perceive (what happens to be) herself in the mirror.

Self-awareness in animals

211

But the sort of recognition that is relevant to bodily self-awareness involves perceiving one's own image as *an image of one's own body*.

In Gallup's studies, primates who had become familiar with mirrors were anesthetized and painted with odorless markers on parts of their heads that were visually inaccessible without the use of mirrors. After awaking, the primates used mirrors to pick at the marks (something they did not do without mirrors). Initially, only chimpanzees and orangutans exhibited the ability to use mirrors for self-examination. Only they – and humans – it seemed, could perceive their reflections as reflections of their own bodies (Gallup [1977]).

More recently, representatives of other species have passed the mirror test, though some controversy remains. Bonobos, or "pigmy chimpanzees," have succeeded (Hyatt and Hopkins [1994]). Among gorillas, Koko, the most proficient language pupil of her species, has apparently made the grade (Patterson and Gordon [1993], p. 71), but it is unclear whether any other gorillas have (Parker [1994]; Gallup *et al.* [2002], pp. 326–327). Meanwhile, after some early inconclusive studies, a carefully controlled experiment indicated that dolphins can examine themselves in mirrors and other reflective surfaces (Reiss and Marino [2001]). More recently, a well-regarded study confirmed mirror self-recognition in elephants (Plotnik *et al.* [2006]). Studies on tamarin monkeys suggested they *might* have recognized themselves, but a later attempt to replicate these results failed, leaving uncertain how to interpret the initial data (Hauser [2000], pp. 107–109; Gallup *et al.* [2002], p. 327).

Thus, certain non-human species are clearly capable of recognizing themselves – as themselves – in mirrors, but we don't know how far into the animal kingdom the capacity extends. Although it is silly to maintain, as some commentators have, that mirror self-recognition is the only valid indication of self-awareness in animals, it is surely one relevant consideration in the case for bodily self-awareness.

9. TAKING INTO ACCOUNT ANOTHER'S SPATIAL PERSPECTIVE

The behavior of some animals indicates that in pursuing particular objectives they can take into account another individual's spatial (and perceptual?) perspective. In one case (Kummer [1982]), a troop of baboons were resting when, over some twenty minutes, a female gradually moved about two meters, ending up behind a rock where she groomed a male. Had the dominant male observed the grooming, there would have been hell to pay. But from where he sat, he could see only the female's back, tail,

and the top of her head. He could not see the male being groomed, who had bent down behind the rock, presumably so the dominant male could not observe the transaction. Jane Goodall (1986, pp. 570, 577–580), meanwhile, provides these instances of suggestive behavior in chimpanzees: A young chimpanzee leads a female out of view of higher-ranking males in order to copulate. A subordinate courting a female covers his erection when a superior male suddenly appears. While fighting a rival, a male hides signs of fear – which might embolden the rival – by suppressing instinctive facial expressions and vocalizations or by manually covering his mouth. A chimpanzee avoids looking at food that only she knows about until other chimpanzees have departed, securing exclusive access to the prize.

In these and similar examples, observers have understood an animal's behavior as evincing an awareness of another individual's spatial perspective, taking it into account in an effort to conceal something about the first animal's situation or behavior – thereby advancing some objective such as a rewarding transaction with a conspecific or exclusive access to food. Wherever such an attribution is correct, it would seem to imply bodily self-awareness. For in each case, the other's perspective is salient *in relation to the agent's own physical position or situation*, of which the agent must be aware for the behavior to be effective.

The most rigorous available evidence of animals' taking into account another's spatial perspective suggests a more radical thesis: that the animals have a "theory of mind," that is, some grasp of other individuals' mental states. In a series of carefully controlled studies (Hare *et al.* [2001]), a subordinate and a dominant chimpanzee competed for food, which was arranged in various ways on the subordinate's side of two opaque barriers. In each setup, the subordinate saw the baiting procedure and could monitor the dominant's visual access to the food. If subordinates could determine what dominants could see, they should preferentially get the food that dominants had not seen hidden or moved. This is exactly what happened. Moreover, when a dominant who witnessed the baiting was replaced with another dominant who had not, subordinates adjusted their behavior accordingly, demonstrating some ability to keep track of who had seen what. A similar set of experiments involving capuchin monkeys (Hare *et al.* [2003]), interestingly, did not furnish evidence that they are sensitive to what conspecifics can see. Yet later, differently designed experiments suggested that not only chimpanzees, but also tamarins and rhesus monkeys, distinguish human investigators' goal-directed and accidental behavior – based on how their actions relate to environmental

constraints – in making inferences about the investigators' goals (Wood *et al.* [2007]). If this interpretation of the data is correct, then monkeys too have a theory of mind.

Besides being interesting in its own right, the question of whether particular animals have a theory of mind is relevant to whether they are capable of true deception (see next section). Moreover, evidence that animals can think about others' mental states makes it more believable that they can think about their own mental states (see discussion of metacognition below, section 12).

10. DECEPTION

Do some of the cases described above involve deception? We might initially define deception as an intentional action – or omission – that is misinterpreted by another to the agent's advantage.

The cases of surreptitious grooming and of resisting the urge to look at food seem to meet this standard. And something like this conception is assumed in Byrne and Whiten's attribution of deception to primates' "Machiavellian" manipulations of each other (Byrne and Whiten [1988], chapters 15, 16). Baboons make distinct gestures of "looking" when they see predators or another baboon troop. Other baboons spontaneously follow the gaze. In one instance discussed by the authors, a male attacked a younger baboon, who screamed, provoking several adults to run toward them. The running adults were making aggressive calls, apparently preparing to attack the offending male, who – seeing their approach – suddenly "looked" into the distance, despite the absence of predators or baboons in that direction. The adults stopped and followed his gaze, at which time he escaped. Very effective manipulation, but was it deception?

Deception, as ordinarily understood, involves not merely intentional action that in fact misleads, but *intentional misleading*. The deceiver intends to misrepresent a situation so that another individual will fail to grasp how things really are. Thus, one who deceives has a theory of mind – specifically, a belief or sense that the targets of deception have mental lives (which can be confounded). But it remains debatable whether animals such as those just described have a theory of mind. Whether or not they really see other animals as *conscious subjects* or *thinkers*, surely they see other animals as unlike inanimate objects: they grasp that certain animals will *respond* in predictable ways to particular provocations, not just *move around* the way a ball or balloon might. One might therefore find congenial a suggestion that these animals have a *proto-understanding of agency*.

[an ability] to recognize that certain things in one's environment, such as conspecifics, prey, or predators, can be manipulated in very specific ways – ways that do not apply to various other things, including many other animals, trees, rocks, and the like. What is emphasized is the recognition of another, not so much as a *thinker*, but as a *doer*. (Güzeldere *et al.* [2002], p. 356)

Even such proto-understanding of agency would suggest a sort of self-awareness: an awareness that one's own behavior can induce certain actions in others. This involves more than the agency aspect of bodily self-awareness, for it implicates an awareness of others as doers or actors, an important component of social self-awareness.⁷

II. COMPLEX SOCIAL UNDERSTANDING

That your own actions can influence those of others, especially if they are members of your social group, is an insight of social understanding. Many mammals have complex social lives featuring group living, dominance hierarchies or more equitable relations, a sense of kin to particular others, shifting alliances, and the like. Individuals may keep track of salient interactions with others, such as whom they have fought with, whom they have groomed or been groomed by, etc. Each group member has to understand her position in the group and her relation to particular others as well as what behavioral expectations follow from these factors. This understanding manifests social self-awareness. To the extent that memory is involved – for example, that so-and-so recently groomed me, or attacked me – such understanding also implicates a non-trivial awareness of one-self as persisting over time. Examples of complex social understanding in particular species will add flesh to these skeletal remarks.

It is common knowledge that wolf packs feature nuanced social dynamics. Moreover, as many human caretakers notice, domestic dogs (a species that evolved from wolves) engage in pack behavior within a human household; even if there is only one dog, he may assess which human is the “alpha” and work to forge an alliance with him or her. Though less actively social than dogs, domestic cats work out dominance hierarchies among themselves.

⁷ We have considered whether non-human primates have a theory of mind. What about dolphins? According to one extraordinary account, a sailor had fallen overboard in rough waters and out of sight of his crew. Later, a group of dolphins surrounded the struggling sailor. Two dolphins from the group swam to the boat and “told” the human crewmates where the lost sailor was by approaching the boat, swimming away in the direction of the swimmer, and repeating this sequence until the boat followed them and found the sailor (White [2007], p. 163).

Self-awareness in animals

215

Primate social life has been the subject of extensive ethological study. Cheney and Seyfarth's leading work on vervet monkeys, for example, demonstrates that vervets know who is a relative, who is a dominant, who's a relative of a dominant, and how other group members rank against each other (Cheney and Seyfarth [1990]). The authors argue that monkeys' innate disposition to group others in hierarchies and family structures evolved to facilitate the ability to predict the behavior of conspecifics (Seyfarth and Cheney 2002). This plausible conjecture may apply as well to other highly social mammal species.

Apes recognize group members, remember favors and grudges, have long-term relationships, and build and shift alliances (Goodall [1986]; Stanford [2001]). The structures of social life characterizing different ape species reveal differences, however. For example, while chimps are very hierarchical and not infrequently violent, bonobos cooperate more, communicate with recreational sex, and excel at building alliances (Stanford [2001], chapter 1).

Though harder to study in the wild than land animals, cetaceans (whales) have been found to have exceptionally complex social lives. Inasmuch as the ocean habitat provides such large animals no refuge from predators, it is unsurprising that these most intelligent of aquatic mammals have evolved elaborate social abilities and group structures: they are utterly dependent on cooperation and mutual defense (Simmonds [2006], p. 109; Connor and Peterson [1994]).

Perhaps the most extensively studied cetacean is the bottlenosed dolphin. As Louis Herman (2002) explains, what young members of the species have to learn about social life is extensive and time-consuming:

To function effectively within these units, the young dolphin must undergo extensive learning about the conventions and rules of the society, about cooperative and collaborative activities, and about the identities and even personalities of group members . . . The protracted period of development and dependence of young dolphins on their mothers and other group members allows the time and opportunity for extensive social learning to take place. (p. 275)

One joint activity is cooperative hunting, which features role specialization: "driver dolphins" herd fish towards the "barrier dolphins." Another example of role specialization is the "broker dolphin" who acts as a link of communication among various subgroups within the larger social unit (Simmonds [2006], p. 110).

These and many other data support the thesis that a wide range of mammalian animals have rich social lives featuring relatively sophisticated

social understanding. Such understanding, I have argued, evinces social self-awareness.

12. METACOGNITION

Our discussion so far has focused on evidence that strongly suggests bodily self-awareness, social self-awareness, and, cutting across these two types, temporal self-awareness (i.e., an awareness of oneself as persisting over time). What about introspective awareness, an awareness of one's own mental states? Does this require such extensive abstraction that only linguistic beings possess it? There is reason to suppose not.

Of special interest are recent studies on metacognition in animals. Metacognition involves having cognitive states about other cognitive states. Strictly speaking, a theory of mind involves metacognition insofar as, say, X has beliefs about what Y believes, sees, or intends. But what is generally meant by "metacognition" in recent ethology literature is having cognitive states about *one's own* cognitive states. Any creature capable of metacognition (in this sense) is capable of introspective awareness because such meta-states involve awareness of the contents of one's own mind.

Some of the best evidence of metacognition in animals comes from studies of monkeys by David Smith and colleagues (for summaries, see Smith and Washburn [2005]; Phillips [2006]). Monkeys learned to control a joystick to choose answers in discrimination tests about visual patterns on a computer screen. They received food pellets for correct responses and timeouts (delays before further trials) – which they hated – for incorrect responses. Then they learned to choose an on-screen icon for "pass" when a test was too difficult. If they chose pass, they received no food and there was no delay; they simply moved to the next trial, a consequence more desirable than a timeout but less desirable than immediate food. Their ability to use the pass option provided initial evidence that they assessed their own level of confidence and understood that they were unsure – an instance of metacognition.

But what if the monkeys were not assessing their own confidence or understanding, but merely doing something to move faster to another trial? Or, conflicted about which answer was correct, simply selecting the pass option by default? Further data renders such skepticism more difficult to maintain. First, less cognitively sophisticated animals, rats, never learned the pass option (Smith and Schull [1989]), suggesting that the monkeys might be doing something special. Second, the researchers changed the monkey trials so that they received food or timeouts only after a series of

trials, rather than after each trial. Third, the investigators found that some monkeys can use the pass option in a brand new cognitive test rather than having to wait to learn its consequences anew (as would seem necessary for a conditioned response). Moreover, recent trials have had the monkeys demonstrate the ability to *remember* previously shown images rather than discriminate among present images (Hampton [2001, 2005]). In these trials the monkeys who master the task apparently attempt to recall an image, compare it with current images, and decide whether they can make a match. In addition to providing evidence for introspective awareness, this cognitive achievement strongly suggests episodic memory. Finally, new research suggests that monkeys who learn a pass response in a perception task can immediately apply it not only to different perception tasks but to memory tasks as well (Kornell *et al.* [2007]). At the same time, it is worth noting that one leading scholar has proffered alternative, “deflationary” explanations for the data (Carruthers [2008]), keeping the issue open.⁸

13. CONCLUSION

Our discussion has supported several claims about self-awareness that are not widely accepted. First, self-awareness is not a single phenomenon; rather, it admits of types that are worth distinguishing. Second, and relatedly, self-awareness can exist in quite humble forms. Any creature with an awareness of its own body as importantly different from the rest of the world – as directly connected with certain feelings and as subject to one’s direct control – has bodily self-awareness. A vast range of animals, it seems, has this sort of self-awareness. A smaller set of animals, members of highly social species including primates, cetaceans, and many other mammals, possess social self-awareness (which presupposes bodily self-awareness). It is therefore abundantly clear that self-awareness is neither exclusively human nor dependent on linguistic competence. There is also good, if not conclusive, reason to believe that certain non-human animals have a degree of introspective awareness.

⁸ Space constraints prevent me from discussing another type of evidence for self-awareness in animals: reports of linguistic self-references by apes trained in sign language. For example, when asked “WHO THAT?” as she gazed in the mirror at herself, the chimpanzee Washoe allegedly signed “ME WASHOE” (Gardener and Gardener [1969]). For several examples involving the gorilla Koko, see Patterson and Gordon (1993).