

Animals, Consciousness, and I-thoughts

Rocco J. Gennaro, University of Southern Indiana

[The final version of this essay can be found in *Philosophy of Animal Minds*, R. Lurz (ed.), Cambridge University Press, 2009, pp. 184-200.]

1. Introduction

I-thoughts are thoughts about one's own mental states or about "oneself" in some sense (Bennett 1988). They are closely linked to what psychologists call "meta-cognition"; that is, "cognitions" about other cognitions or mental representations (Metcalf & Shimamura 1994; Koriat 2007). There seems to be growing evidence that many animals are indeed capable of having I-thoughts as well as have the ability to understand the mental states of others (Terrace & Metcalfe 2005; Hurley & Nudds 2006).

There is also a relevant philosophical theory of consciousness; namely, the "higher-order thought" (HOT) theory of consciousness which says that what makes a mental state conscious is the presence of a suitable higher-order thought about that state (Gennaro 2004a; Rosenthal 2005). For various reasons, such thoughts are typically understood to take the form "I am in mental state M now." A higher-order thought, then, is a kind of meta-cognition. It is a mental state directed at another mental state. So, for example, my desire to write a good book chapter becomes conscious when I am (non-inferentially) "aware" of the desire. Intuitively, it seems that conscious states, as opposed to unconscious ones, are mental states that I am "aware of" in some representational sense (Lycan 2001). In a case of subliminal perception, I am not aware that I am in that perceptual state. Thus, it is unconscious. However, when I become aware that I am having that perception, it becomes conscious.

An often cited problem, however, is that the HOT theory rules out animal consciousness because animals (or at least most animals) are incapable of having such thoughts; they do not possess such sophisticated “I-concepts” and “mental concepts.” This is a common objection normally offered by non-HOT theorists, such as Fred Dretske (1995), Robert Lurz (2002), and Bill Seager (2004). Moreover, one prominent HOT theorist, Peter Carruthers, actually embraces this alleged consequence of the HOT theory (Carruthers 2000, 2005). I have had my say elsewhere on Carruthers’ contention that animal consciousness is very unlikely given the truth of some form of HOT theory (Gennaro 1996, 2004b, 2006). I won’t repeat those arguments here except to say that a higher-order thought need not be as sophisticated as it might seem. Since most of us believe that many animals have conscious mental states, a HOT theorist must explain how animals can have the higher-order thoughts necessary for such states. One reason most of us believe that animals have conscious states is simply because our folk psychology is a theory of conscious mental states and it works well in explaining and predicting much of animal behavior.

Thus, there is a three-way tension among the following claims that needs to be relieved: (a) Most animals have conscious mental states, i.e. there are generally positive common sense grounds for believing that animals have conscious states, (b) The HOT theory is true, which, in turn, entails having I-thoughts, and (c) Few (if any) animals are capable of having I-thoughts based on various empirical and theoretical considerations.

Carruthers rejects (a) and embraces (b) and (c), while Dretske, Lurz, and Seager endorse (a) and (c) but reject (b). I reject (c) and accept (a) and (b). Thus, this chapter has a double purpose: to discuss and elaborate upon the evidence for higher-order

thoughts (or I-thoughts) in animals but also to show that the HOT theory is indeed consistent with animal consciousness. In section 2, I will argue that recent experimental evidence on animal memory and meta-cognition strongly suggests that many animals have the self-concepts and mental state concepts necessary to form I-thoughts. In section 3, I reply to the claim that having I-thoughts requires having thoughts (and thus concepts) directed at others' mental states. The stakes are high because if the HOT theory is true, any evidence indicating the absence of I-thoughts would also cast doubt on animal consciousness itself.

It is also crucial at the outset to note one important subtlety of the HOT theory. When a conscious mental state is a first-order world-directed state, the higher-order thought is *not* itself conscious. When the higher-order thought is itself conscious, there is a yet higher-order (or third-order) thought directed at the second-order state. In this case, we have *introspection* which involves a conscious higher-order thought directed at an inner mental state. When one introspects, one's attention is directed back into one's mind. For example, what makes my desire to write a good book chapter a conscious *first-order* desire is that there is a (unconscious) higher-order thought directed at the desire. In this case, my conscious focus is directed at the entry and my computer screen, so I am not consciously aware of having the higher-order thought from the first-person point of view. When I introspect that desire, however, I then have a *conscious* higher-order thought (accompanied by a yet higher, third-order, thought) directed at the desire itself. An additional rationale for this aspect of the HOT theory is to serve as a reply to the objection that it is circular or leads to an infinite regress. It might seem that the HOT theory results in circularity by defining consciousness in terms of higher-order thoughts.

It also might seem that an infinite regress results because a conscious mental state must be accompanied by a higher-order thought, which, in turn, must be accompanied by another higher-order thought *ad infinitum*. The standard reply is to remind the objector that the initial higher-order thought need not itself be conscious.

Of course, the very concept of ‘consciousness’ is notoriously ambiguous, but perhaps the most commonly used contemporary notion of a ‘conscious’ mental state is captured by Thomas Nagel’s famous “what it is like” sense (Nagel 1974). When I am in a conscious mental state, there is “something it is like” for me to be in that state from the subjective or first-person point of view. This is how I will use the term.

2. I-thoughts and HOTs in animals

It is sometimes said that all or most (non-human) animals cannot “mindread”; that is, do not understand that others (or even they) have mental states. And adherence to the so-called “theory-theory” view of mindreading, whereby understanding mentalistic notions presupposes having a “folk-psychological” theory of mind, seems to rule out that animals have I-thoughts (or higher-order thoughts).¹ In addition, Carruthers (2005), for example, cites experimental work suggesting that chimps (and very young children, for that matter) lack the mental concepts ‘appear’ or ‘see’ (Povinelli 2000), which is then treated as necessary to have I-thoughts about *one’s own* experiences. Such experiments are designed to determine if chimps take notice of whether or not the experimenter is looking at something (e.g. food) or unable to see something (e.g. due to blindfolding). However, there seems to be growing evidence that many animals can indeed have some

I-thoughts and read (other) minds in some sense. Moreover, as we will see in section 3, it is not at all clear that having I-thoughts requires reading *other* minds.

In any case, there are two main concepts in an I-thought (or higher-order thought); namely, a self-concept (“I”) and mental state concept (“M”). Let us consider them in turn.

2.1 Self-concepts and episodic memory in animals

Episodic Memory is an explicitly conscious kind of remembering involving “mental time travel” (Tulving 1983, 2005). It is, for example, often contrasted with *semantic* memory, which need only involve knowing that a given fact is true or what a particular object is, and *procedural* memory whereby memory of various learned skills is retained. Tulving also uses the term “autonoetic consciousness” for that kind of episodic remembering. The link to I-thoughts is fairly clear: some notion of self or “I” seems necessary to have a genuine episodic memory. I recognize the memory as *mine* and representing an event in *my* past. Tulving speaks of episodic memory’s “dependence on a remembering ‘self’ [and]... relation to subjectively apprehended time” (Tulving 2005, p. 14).

However, Tulving himself strongly resists the idea that non-human animals have episodic memory. But his case is less than convincing and he even seems to retreat from this position in the very same essay. For example, he concedes “that some of [his] assertions may be too strong (2005, p. 27).” In speaking, by analogy, of amnesic patient KC, he explains that “it is possible that he [KC] has a little left of one or more of the properties of episodic memory and that we are dealing with a case of severe impairment

in episodic memory rather than its total absence” (27-28). Now this naturally makes one wonder why the same should not be said for most non-human animals if indeed KC has at least a limited ability for episodic memories as well as an ability to think about the short term future. It is hard to see how, for example, KC can play various card games and even chess without having *some* ability for mental time travel (both into the past and future). Tulving also tells us that KC’s short term or “working memory” is “normal; he remembers what happened a short-term while (1 to 2 minutes) ago...” (p. 23)²

Turning more directly to non-human animals, Tulving often qualifies his strong negative claim by speaking approvingly of the more cautious conclusion (reached in Roberts 2002) that chimps’ ability to imagine their *extended* future is in doubt and that their ability to mentally travel into the future and past is *limited* (p. 39). Perhaps most telling is Tulving’s concession that Clayton and Dickinson and their colleagues (in e.g. Clayton et al. 2003) have “reported ingenious and convincing demonstrations of memory for time in scrub jays” (p. 37). Scrub jays are food-caching birds and when they have food they cannot eat, they hide it and recover it later. Because some of the food is preferred but perishable (e.g. crickets), it must be eaten within a few days, while other food (e.g. nuts) is less preferred but do not perish as quickly. In cleverly designed experiments using these facts, scrub jays are shown, even days after caching, to know not only *what* kind of food was *where* but also *when* they had cached it (see also Clayton et al. 2006).

There is even more positive recent evidence for episodic memory in animals not mentioned by Tulving. Evidence in primates is discussed in the same volume in which Tulving’s 2005 chapter appears. Menzel (2005), for example, describes experiments in

which a female chimp (Panzee) recovers food hidden by a trainer by coaxing a different trainer (who was unaware of its location) to let her get it. This was done even after quite a bit of time delay between the observation and the time at which the second trainer appeared. A similar result can be found in Schwartz (2005) with respect to a gorilla named King. And Hampton (2005) found that a monkey was able to successfully match-to-sample even after delays. The subject had to decide whether to submit to a test of the sample *after* the stimulus display had been removed but *before* the test has been presented. The results indicate that monkeys both know when they remember and when they have forgotten, indicating both a capacity for episodic memory as well as a form of meta-cognition.

Additional evidence is presented in Eichenbaum et al. (2005) whereby rats can remember the temporal order of the odors of various objects. And Dere et al. (2006) extensively review the expanding literature strongly suggesting episodic memory in a variety of animals, including dolphins, birds, and rodents (such as mice and rats). Finally, it is interesting to note that there are also reports of animals able to plan for the future, such as is found in western scrub-jays (Raby et al. 2007). They show that the jays plan for future need by preferentially caching food in a place where they have learned that they will be hungry the following morning. It is often observed how an ability to think about the future (i.e. “prospective memory”) is closely related to the capacity for episodic memory (a point Tulving repeatedly stresses in his 2005; cf. Zentall 2005).

Some interpretations of the above data are, to be sure, not uncontroversial in some circles and we should of course not jump to unwarranted conclusions. However, there is little reason to hold the very strong view that animals have *no* episodic memory or have

no ability at all to mentally time travel. If this is correct, then there is also no reason to deny an animal's ability to form at least some minimal self-concept which, in turn, can figure into a higher-order thought.

Three other points are worth making here. First, it is important not to equate having episodic memories with having *accurate* episodic memories. That is, if an experiment really shows a lack of the accuracy of an episodic memory, we should not conclude that the animal in question has no episodic memory at all. If we compare this to human adults, we quickly see the problematic inference. Human eyewitness reports, for example, are often mistaken but we would not and should not infer any general lack of episodic memory on the part of the subject. We must be careful not to hold animals to a higher standard than we hold ourselves.

Second, I have previously argued at length that there is a compelling *a priori* Kantian-style argument showing that having at least some form of episodic memory is necessary for even being a conscious creature (Gennaro 1996, chapter nine). The basic idea is that having concepts of outer objects involves some understanding of those objects as enduring through time (since we do not take them to be mere fleeting subjective states of mind), which, in turn, requires us to think of ourselves as temporally enduring subjects with a past (mainly because we recognize that those objects are the same objects at different times). That is, if a conscious organism can re-identify the same object at different times, then it implicitly understands itself as something which endures through time.

Third, we should recognize that there are degrees of self-concepts. We might distinguish the following four I-concepts, moving from the least to most sophisticated (initially presented in Gennaro 1993):

Level 1: I *qua* this thing (or “body”), as opposed to other physical things.

Level 2: I *qua* experiencer of mental states.

Level 3: I *qua* enduring thinking thing.

Level 4: I *qua* thinker among other thinkers.

Interestingly, one also finds a willingness to talk of a continuum of self-consciousness in the experimental literature (e.g. Kinsbourne 2005) as well as corresponding levels of consciousness in human development, including the purely physical “self-other contrast” (Nelson 2005). In any case, all that is needed for having most higher-order thoughts is the kind of minimal “bodily self-consciousness” self-concept found in level 1; that is, being able to distinguish one’s own body from other things. And it is surely fairly uncontroversial that most animals at least have this most unsophisticated I-concept. My own view in the end is that most animals are at least capable of a level 3 self-concept and perhaps even a level 4 concept, as is evidenced by the results and arguments offered in this and the next section.

2.2 Attributing mental states to oneself and others

It is also important to point out some of the evidence for *mental state* attributions, i.e. the “M” side in a higher-order thought. Some of it suggests that animals have meta-

cognitive states but only directed at themselves; thus, I-thoughts. We have already seen some significant evidence for this in sec 2.1 in what is sometimes termed “meta-memory.” If an animal has a meta-memory state, then it not only has a self-concept but it is also able to form a thought directed at a memory, which is itself a mental state.

In addition, there is the much discussed work on uncertainty monitoring with animals such as monkeys and dolphins (Smith et al. 2003; Smith 2005). For example, a dolphin is trained in a perceptual discrimination task, first learning to identify a particular sound at a fixed frequency (the “sample” sound). Later, he learns to match other sounds to the sample sound. When presented with a sound which is either the same or different in pitch as the sample sound, he has to respond in one way if it is the same pitch (e.g. press one paddle) and another way if it is a different pitch (press another paddle). Eventually, the dolphin is introduced into a test environment by being forced to make very difficult discriminations. To test for the capacity to take advantage of his own uncertainty, the dolphin is presented with a third “uncertain” response that is rewarded if he is uncertain. He is presented with a third paddle, the Escape paddle, which is virtually equivalent to declining the trial. The dolphin does indeed choose the Escape paddle with expected frequency and a similar response pattern to humans and rhesus monkeys, which in turn is taken by many to suggest that he is aware of his state of uncertainty; that is, he has some knowledge of his own mental state. This is a meta-cognitive state: the dolphin is aware that he doesn’t know something, e.g. whether or not a sound matches (or is very close to) the sample sound. It seems reasonable to seek a common underlying explanation for all of the subjects involved (Browne 2004).

A related paradigm has to do with a subject's (e.g. a monkey's) ability not to accurately respond to a stimulus, but rather on the appropriateness of its level of confidence in the accuracy of a response (Son & Kornell 2005; Hampton 2005). Such "meta-confidence judgments" are treated as evidence of meta-cognition because the experiments are designed to elicit a "betting" judgment of the form "I am confident that I know." I won't describe this experiment in any detail, but, for example, two rhesus macaques were tested in this way using a system of low and high risk bets. In brief, the monkeys did indeed tend to bet "high risk" much more often when they were able to make accurate confidence judgments, and bet low risk more often when responding incorrectly. Thus, they seem able to express feelings of confidence or lack of confidence about their cognitions, and so display meta-cognitive ability.

It is also crucial to note here that some authors do speak of degrees of meta-cognition or self-awareness in ways arguably similar to the distinction between conscious and unconscious higher-order thoughts (as described in section 1). For example, Son and Kornell (2005, pp. 300-1) interestingly speak of "*implicit* meta-cognition" which sounds very much like "unconscious higher-order thoughts." They refer to the "*tacit* meta-judgment of uncertainty" as opposed to *explicit* meta-cognitions or *self-reflective* consciousness, which clearly has an affinity to the more sophisticated conscious higher-order thoughts. We may not often have evidence of such further awareness (i.e. self-reflective consciousness) because monkeys, for example, cannot verbally express their judgments. "If we did, we would then have evidence of *meta*-metacognition..." (p. 318). Again, a clear example of meta-metacognition would be a *conscious* higher-order thought, but this is more than is required for one to have a conscious state. Thus, animals

can still have unconscious higher-order thoughts continuously accompanying all of their conscious states. As Son and Kornell (2005) put it: “according to this view, we make meta-cognitive judgments *constantly* and without explicit knowledge of them” (p. 317, emphasis added).

Let us turn now to the ability of animals to attribute mental states *to others*, which is another kind of higher-order thought, albeit a thought about *another's* mental state. Despite the Povinelli style experiments noted briefly early in this section, the evidence seems to be growing that at least some animals can "mindread" under other conditions. For example, recent work by Laurie Santos and colleagues show that rhesus monkeys attribute visual and auditory perceptions to others in more competitive paradigms (Flombaum & Santos 2005; Santos et al. 2006). Rhesus monkeys preferentially attempted to obtain food silently only in conditions in which silence was relevant to obtaining food undetected. Monkeys would take grapes from a silent (i.e. not noisy) container while a human competitor was looking away; thus, apparently understanding that hearing leads to knowing on the part of human competitors). Subjects reliably picked the container that did not alert the experimenter to the fact that a grape was being removed. This suggests that monkeys take into account how auditory information can change the knowledge state of the experimenter. In addition, Rhesus monkeys also chose to take food from human competitors who could not see them, either because the humans' eyes were facing away or because their faces were blocked by an opaque barrier.³

It is important to point out that controls were used to eliminate various non-mindreading (or "behavior-reading") interpretations of the data. For example, it is shown

that monkeys do not prefer to steal grapes from the non-belled containers simply because the sound of bells frightens them. There was also no historical link between the competitor's observable features and his future actions; that is, subjects had no past experience hearing the bell make noise which could have been used to associate with the competitor's likely response.

Now, it may be the case that a non-mindreading interpretation *could* still be given for virtually all such experiments. It may indeed *always* be *possible* to construct creative and very elaborate alternative first-order mental or even purely behavioral explanations for any given set of animal behaviors (Povinelli & Vonk 2006; Carruthers 2008). A very thorough reply would be a topic for another paper. I'll only say here, first, that just because an alternative explanation is *possible*, it doesn't follow that it is the best or most reasonable explanation. Second, there comes a point where such deflationary interpretations might even work for much of *human* behavior or for, say, the behavior of a deaf mute human who is incapable of verbal communication. Third, as Browne puts it: "it is parsimonious to explain similar, complex, stimulus-response patterns by similar psychological mechanisms" (2004, p. 648). Thus, I side with those who hold that attributing higher-order mental states (and consciousness for that matter) is often the more parsimonious move to make. This is all the more clear when one considers other similarities between humans and other animals, such as evolutionary history and brain structure.⁴

Finally, it is worth mentioning that results similar to those described in section 2.1 regarding caching and episodic memory are applicable here as well. For example, many crows and scrub jays return alone to caches they had hidden in the presence of others and

re-cache them in new places (Emery & Clayton 2001). This suggests that they know that *others* know where the food is cached and thus, to avoid having their food stolen, they re-cache the food.

Taken together with the earlier evidence presented, it seems reasonable to think that many animals have I-thoughts of some kind or other. Although many subjects of the above experiments are primates, many other animals are also tested, such as dogs, pigs, and even mice and rats.

In the next section, I'll consider two problematic claims which underlie some of the opposition to the above conclusions.

3. Two problematic claims

Some insist that if an animal cannot pass a given mind-reading task (directed at another), it is therefore incapable of having any thoughts about its own mental states.

This is a very problematic claim. The thesis to be challenged, then, is:

(1) Having I-thoughts requires having thoughts directed at others' mental states.

The obvious corollary of (1) is:

(2) If an organism O cannot form *concepts* of another's mental states, then O also cannot have I-thoughts of any kind.

3.1 A reply to thesis (1)

To be sure, despite the recent evidence presented in section 2, many animals do not pass various tests designed to show the ability to mindread (such as the Povinelli style experiments mentioned early in section 2). Nonetheless, I disagree with the notion that

one should conclude from these negative results that most animals therefore do not have *any* higher-order thoughts. That is, I think that (1) is false. First, it is not clear that so much should be read into the failure of animals in such experiments. For one thing, failure might be explained in such situations because they don't typically arise in their native environment. As we have seen, many primates, at the least, do much better in similar tests when performed in more natural or competitive settings. Moreover, it is odd to treat such experimental results as if the paradigms used indicate an undeniably clear *necessary* condition for having other attributing mental capacities. This is somewhat reminiscent of treating failing the Turing Test as indicating the utter absence of any intelligence or even consciousness.

More to the point here, even if some or most animals cannot, say, engage in intentionally deceptive behavior (to use just one example) and so arguably do not have thoughts about the mental states *of others*, it still does not follow that they cannot have *unconscious* higher-order thoughts about *their own* mental states. Recall that only self-directed unconscious higher-order thoughts are required for conscious states, according to the HOT theory. Thus, I agree with Ridge (2001, p. 333) that the opposing view rests on the false assumption "that there could not be an agent capable of having HOTs about its own mental states but incapable of having HOTs about the mental states of others." The latter at least seems far more sophisticated and does not seem necessary for there simply to be conscious mental states, especially simple conscious pains or perceptions. Moreover, as Ridge (pp. 322ff) also points out, the move from "no deceit" to "no HOTs whatsoever" is much too quick and unjustified. Just because evidence of intentional

deception would be the *best* evidence of mindreading, it clearly does not follow that the lack of such evidence indicates a lack of higher-order thoughts.⁵

Finally, it seems to me that some tests for other-attributing thoughts in the theory of mind literature are often aimed at determining whether or not animals (or infants) can have *conscious* thoughts directed at another's mental state. Speaking of "intentional deception," for example, seems to suggest that the animal is *consciously* intending to cause a false belief in another animal. To the extent that this is what experimenters have in mind, then there is again no reason to think the HOT theory is in trouble even if animals lack this ability. As we have seen, the HOT theory allows for the presence of conscious states even in the absence of *any* (either self-attributing or other-attributing) *conscious* higher-order thoughts. If a higher-order thought is itself conscious, then one is in a more sophisticated *introspective* state which is not necessary for having a more primitive first-order conscious state. Thus, thesis (1) is, at best, on very shaky ground.

3.2 Concept possession and the generality constraint

One might reply to the above that adherence to the so-called "generality constraint" dictates that if one can self-attribute a mental concept then one should be able to other-attribute that mental concept. In an effort to show this to be a serious problem for the HOT theory, Seager (2004) also cites evidence suggesting that animals do not other-attribute mental states, and then says the following:

"...all animals also lack the ability to attribute mental states to themselves because those who can self-attribute will be a subset of those who can other-attribute....It is...[a]

doubtful logical possibility that a being which lacked the ability to attribute mental states to others could attribute them to itself....such an asymmetry would seem to run counter to Evans's (1982) 'generality condition' on concept possession (the claim that one cannot have a concept C unless for any object O, one can have the thought that 'O is C')....What is incoherent is the notion that one could conceptualize one's own possession of mentalistic attributes while being completely unable to form a thought of another being having mental states." (Seager 2004, pp. 264-5)

For the sake of clarification, the generality constraint is sometimes put as follows: the attribution of thoughts to any organism of the form "a is F" and "b is G" commits us to the idea that the organism should also be able to think that "a is G" or "b is F." Moreover, we might even agree that "the content of all propositional attitudes is said to be subject to this constraint." (Toribio 2007, p. 446) Thus, we might also think of the generality constraint as involving a commitment to the idea that belief and desire states (which are also composed of concepts) can be recombined with other such states, and perhaps even that the organism can make appropriate simple inferences among them.⁶

In any case, Seager's quotation, at minimum, indicates an endorsement of thesis (2) above: If an organism O cannot form *concepts* of another's mental states, then O also cannot have I-thoughts of any kind. However, the generality constraint is a very strong condition on concept possession. For example, Tye (2006) calls it a "stronger requirement" than others he considers because it requires that I am capable of thinking *any* thoughts that can be formed from combining a concept with other concepts I possess. This makes all concept users into idealized rational agents with a complete grasp of all its

concepts and capable of combining into thoughts all concepts that one possesses. One might instead opt for the view that possessing a concept C is one of degree which allows for “a partial understanding of C. On this intuitively attractive view, one cannot possess the concept FORTNIGHT, for example, unless one grasps that a fortnight is a period of time” (Tye 2006, p. 9).

Even among human beings we sometimes distinguish between those who have a “partial” concept of, say, DOG and those who have, on the one extreme, “no concept at all” and, on the other, an expert biologist’s dog concept. The same might be said for the understanding that a young child has of TREE compared to an adult’s increased understanding and then finally to a botanist’s tree concept. Treating concept possession in this way would also not require all concept users to be able to draw all possible connections among one’s stock of concepts. We sometimes might not “see” the connection between two concepts or thoughts containing those concepts because we only have a partial grasp of the concepts involved. This should also be said for various animals; that is, it may be that animals have a (partial) understanding of VISION, SEE, PERCEPTION, and the like without always being able to apply those concepts to others in certain experimental situations. We should not conclude that animals have *no* concept of X if they are unable to pass a test for a more advanced understanding of X. This, I suggest, is what skeptics often do when, say, a chimp doesn’t infer a concept of VISION or SEEING from certain particular movements (or lack of movements) of another’s eyes. Instead, we have seen how various experiments (such as Santos’) indicate at least some partial grasp of the concepts SEEING and HEARING. This notion of concept possession can thus help to explain *why* certain connections between thoughts and other

propositional attitudes are not made by some animals. It can also explain how an animal can fail to attribute a mental concept to another, but still apply that (or a very similar concept) to itself.

Of course, one central issue then becomes: what is it to have a mental concept 'M' anyway? Again, the answer will in part depend on which notion of concept possession one holds. If we accept the idea one can have a concept C based on a partial understanding of C, then it is very unclear what reason we have for withholding such concepts as PERCEPTION, PAIN, and DESIRE from most animals. To be sure, most animals may not have a very sophisticated concept of these mental states, but yet it seems that they are aware that they are in such states, e.g. a dog seems to be aware that it is seeing as opposed to, say, blind or hearing. Similarly, an animal knows when it is in pain even if it has a difficult time determining or understanding when another creature is in pain. Thus, having at least a partial understanding of the concept PAIN would involve the notion that "this hurts," as opposed, say, to comprehending philosophical writings on pain. An animal with a DESIRE for food understands that it "wants something." And a partial understanding that one is having a VISUAL PERCEPTION involves at least grasping that one is "seeing" as opposed to, say, "hearing." As Allen puts a similar point: "...philosophers have been tempted by the argument that...for example, a dog does not believe there is a squirrel in the tree because it lacks 'the'...concept of squirrel. But there is no reason to think that having [that] belief requires that animals have that specific concept, nor that lacking the canonical concept of squirrel means that they lack any concept whatsoever" (Allen 1999, pp. 35-6).

Thus, another question arises when considering the plausibility of the generality constraint; namely, do we even apply the *same* exact mental concept to ourselves as we would to another? It is assumed by many that the concept PAIN or VISUAL EXPERIENCE that one might attribute to oneself is the same as that which one attributes to another. Then, as we have seen, the reasoning goes that since many animals can't attribute the latter they can't attribute the former. This is a highly questionable claim. For one thing, if the above account of concept possession is viable, then it may just be that an organism O has a *better* understanding of a mental concept C when self-attributing C than when attributing it to another. And so, *contra* Seager and the generality constraint, it would be perfectly reasonable for an animal to be able to conceptualize its own mental states *to some extent* without being able to conceptualize, to the same extent or in the same way, another's mental state.

Moreover, perhaps we should even suppose that there is a difference between the concepts *MY PAIN* and *YOUR PAIN*, or *MY VISUAL EXPERIENCE* as opposed to *YOUR VISUAL EXPERIENCE*. That is, since we only *directly* experience our own mental states, it might just be that the concept MY always implicitly accompanies my mental state concepts. The lion or chimp has more immediate first-person access to its own mental states and thus can acquire concepts of its own mental states in a more direct way. However, acquiring and then attributing mental states to others (such as *YOUR PAIN* or *YOUR BELIEF*) involves an additional, or at least different, inferential process. To put it bluntly: it is normally *much harder* to know about another's mental states than it is to know about one's own mental states. There is a reason why the problem of *other* minds is an age old problem in the history of philosophy. One need not hold some

radical Cartesian infallibility view to appreciate this point, since even the most anti-Cartesian skeptic will typically acknowledge that the admittedly fallible access to our own minds (as opposed to other minds) is at least more immediate in some important sense.⁷ For example, I need not interpret my *behavior* in any obvious way when I think that I have a simple desire for food or a pain in my back. Thus, if there are two distinct concepts involved in the self-attribution and other-attribution of mental concepts (because of the “my” and “your” qualifiers), then it is not even clear that the generality constraint is violated when an animal cannot other-attribute a mental state that it can self-attribute. And this would be yet another reason not to accept the logic that failure of some animals to other-attribute mental concepts means that they cannot attribute them to themselves. Attributing mental states to others seems to involve additional cognitive abilities, such as making certain inferences based on behavioral evidence, which some animals may not have. But this does not mean that they are incapable of having mental concepts at all or self-attributing them.

In any case, if I am right thus far, this bodes well for the HOT theory which only requires an (unconscious) I-thought with a self-attributing mental concept in order for an animal to have conscious mental states.

Now, much of the above raises the very difficult issue of under what conditions is it reasonable to attribute any concept to a non-linguistic animal. I cannot fully address this issue here. Nonetheless, one core idea in the animal cognition literature is that the attribution of concepts is justified if there is evidence supporting the presence of a mental representation that is independent of solely perceptual information (Allen & Hauser 1991). This view fits well with the attribution of concepts in the experiments described

the previous section. Following up on this, Allen (1999, pp. 36-7) proposed the following account:

An organism O may reasonably be attributed a concept of X (e.g. TREE) whenever:

- (i) O systematically discriminates some Xs from some non-Xs; and
- (ii) O is capable of detecting some of its own discrimination errors between Xs and non-Xs; and
- (iii) O is capable of learning to better discriminate Xs from non-Xs as a consequence of its capacity (ii).

As Allen notes, the above is not meant as a definition of ‘concept,’ but rather a criterion for when it is reasonable to attribute a concept X to an animal. Of course, if we are going to use it as a guide for attributing *mental* concepts to animals, then the X in question cannot be TREE or some external object but rather a mental concept (of which an animal arguably has at least a partial conception). We have seen that many animals seem to be able to meet condition (i); that is, they are able to distinguish one mental state from another, such as seeing from not-seeing at all (or hearing), or remembering and not-remembering. For some, this may even be enough for an animal to have a mental concept since it shows at least some understanding of those concepts by way of comparison.

A case could also be made that some of the above experimental results indicate an ability to meet clauses (ii) and (iii). For example, when a chimp or dolphin or rat detects its own error in, say, a meta-confidence or meta-memory task, and then goes on to learn

to perform the task better on that basis, it seems to have met (ii) and (iii). So when an animal learns to improve its performance on memory or confidence tasks, it seems to have understood how to *better* discriminate its own mental states from one another. In addition, Allen's own example (1999, p. 38) of pigs' "backout behavior" seems suited for this purpose. He describes cases where pigs display a self-monitoring of performance. Some pigs would attempt to back away from the choice they had made after committing to a response they had given (on, say, a same/different perceptual task), but before any feedback was provided.

In any case, I also find little reason to accept thesis (2).

In closing, then, it is at the least, very premature to claim that animals cannot have I-thoughts at all. At best, there seems to be growing positive evidence that animals can mindread and have I-thoughts, which include various degrees of conceptual sophistication. And even if mindreading others is not found in very many animals, it doesn't follow that no I-thoughts are present. Finally, this also means that if the HOT theory is true, we need not deny that most animals are conscious.

Acknowledgement: Thanks to Robert Lurz for numerous helpful comments on an earlier draft of this chapter.

References:

Allen, C. (1999). Animal concepts revisited: the use of self-monitoring as an empirical approach. *Erkenntnis*, 51, 33-40.

- Allen, C., and Hauser, M. (1991). Concept attribution in nonhuman animals: theoretical and methodological problems in ascribing complex mental processes. *Philosophy of Science*, 58, 221-240.
- Bennett, J. (1988). Thoughtful brutes. *Proceedings and Addresses of the American Philosophical Association*, 62, 197-210.
- Browne, D. (2004). Do dolphins know their own minds? *Biology and Philosophy*, 19, 633-653.
- Carruthers, P. (2000). *Phenomenal Consciousness*. Cambridge, MA: Cambridge University Press.
- Carruthers, P. (2005). *Consciousness: Essays from a Higher-Order Perspective*. New York: Oxford University Press.
- Carruthers, P. (2008). Meta-cognition in animals: a skeptical look. *Mind and Language*, 23, 58-89.
- Carruthers, P. (this volume). Invertebrate concepts confront the Generality Constraint (and win).
- Carruthers, P., and Smith, P. (eds.) (1996). *Theories of Theories of Mind*. Cambridge: Cambridge University Press.
- Clayton, N., Bussey, T., and Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nature Reviews Neuroscience*, 4, 685-691.
- Clayton, N., Emery, N. and Dickinson, A. (2006). The rationality of animal memory: complex caching strategies of western scrub jays. In Hurley and Nudds 2006, pp. 197-216.
- Dere, E., Kart-Teke, E., Huston, J., and Silva, D. (2006). The case for episodic memory

in animals. *Neuroscience and Biobehavioral Reviews*, 30, 1206-1224.

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

Eichenbaum, H., Fortin, N., Ergorul, C., Wright, S., and Agster, K. (2005). Episodic recollection in animals: “If it walks like a duck and quacks like a duck...” *Learning and Motivation*, 36, 190-207.

Emery, N., and Clayton, N. (2001). Effects of experience and social context on prospective caching strategies in scrub jays. *Nature*, 414, 443-446.

Evans, G. (1982). *Varieties of Reference*. Oxford: Clarendon Press.

Flombaum, J., and Santos, L. (2005). Rhesus monkeys attribute perception to others. *Current Biology*, 15, 447-452.

Gennaro, R. (1993). Brute experience and the higher-order thought theory of consciousness. *Philosophical Papers*, 22, 51-69.

Gennaro, R. (1996). *Consciousness and Self-consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*. Amsterdam & Philadelphia: John Benjamins.

Gennaro, R. (ed.) (2004a). *Higher-Order Theories of Consciousness*. Amsterdam & Philadelphia: John Benjamins.

Gennaro, R. (2004b). Higher-order thoughts, animal consciousness, and misrepresentation: A reply to Carruthers and Levine. In Gennaro 2004a, pp. 45-66.

Gennaro, R. (2006). Review of Peter Carruthers’ *Consciousness: Essays From a Higher-Order Perspective*. *Psyche*, 12.

<http://psyche.cs.monash.edu.au/book_reviews/carruthers3/Carruthers.pdf>

Goldman, A. (2006). *Simulating Minds*. New York: Oxford University Press.

Hampton, R. (2005). Can rhesus monkeys discriminate between remembering and forgetting? In Terrace and Metcalfe 2005, pp. 272-295.

- Hare, B., Call, J., and Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behavior*, 61, 139-151.
- Hare, B., and Tomasello, M. (2004). Chimpanzees are more skilled in competitive than in cooperative cognitive tasks. *Animal Behavior*, 68, 571-581.
- Hurley, S. and Nudds, M. (eds.) (2006). *Rational Animals?* New York: Oxford University Press.
- Kinsbourne, M. (2005). A continuum of self-consciousness that emerges in phylogeny and ontogeny. In Terrace and Metcalfe 2005, pp. 142-156.
- Koriat, A. (2007). Metacognition and consciousness. In P. Zelazo, M. Moscovitch, and E. Thomson (eds.), *The Cambridge Handbook of Consciousness*, pp. 289-325. Cambridge, MA: Cambridge University Press.
- Lycan, W.G. (2001). A simple argument for a higher-order representation theory of consciousness. *Analysis*, 61, 3-4.
- Lurz, R. (2002). Neither HOT nor COLD: An alternative account of consciousness. *Psyche*, 9. <<http://psyche.cs.monash.edu.au/v9/psyche-9-01-lurz.html>.>
- Menzel, C. (2005). Progress in the study of chimpanzee recall and episodic memory. In Terrace and Metcalfe 2005, pp. 188-224.
- Metcalfe, J. and Shimamura, A.P. (eds.) (1994). *Metacognition: Knowing about Knowing*. Cambridge, MA: MIT Press.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435-450.
- Nelson, K. (2005). Emerging levels of consciousness in early human development. In Terrace and Metcalfe 2005, pp. 116-141.
- Nichols, S., and Stich, S. (2003). *Mindreading*. New York: Oxford University Press.

- Nisbett, R. and Wilson, T. (1977). Telling more than we can know. *Psychological Review*, 84, 231-295.
- Povinelli, D. (2000). *Folk Physics for Apes*. New York: Oxford University Press.
- Povinelli, D., and Vonk, J. (2006). We don't need a microscope to explore the chimpanzee's mind. In Hurley and Nudds 2006, pp. 385-412.
- Raby, C., Alexis, D., Dickinson, A., and Clayton, N. (2007). Planning for the future by western scrub-jays. *Nature*, 445, 919-921.
- Ridge, M. (2001). Taking solipsism seriously: nonhuman animals and meta-cognitive theories of consciousness. *Philosophical Studies*, 103, 315-340.
- Roberts, W. (2002). Are animals stuck in time? *Psychological Bulletin*, 128, 473-489.
- Rosenthal, D. (2005). *Consciousness and Mind*. New York: Oxford University Press.
- Santos, L., Nissen, A., and Ferrugia, J. (2006). Rhesus monkeys, *macaca mulatta*, know what others can and cannot hear. *Animal Behavior*, 71, 1175-1181.
- Schwartz, B. (2005). Do nonhuman primates have episodic memory? In Terrace and Metcalfe 2005, pp. 225-241.
- Seager, W. (2004). A Cold Look at HOT Theory. In Gennaro 2004a, pp. 255-275.
- Smith, J.D., Shields, W., and Washburn, D. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and Brain Sciences*, 26, 317-373.
- Smith, J.D. (2005). Studies of uncertainty monitoring and metacognition in animals. In Terrace and Metcalfe 2005, pp. 242-271.
- Son, L., and Kornell, N. (2005). Metaconfidence judgments in rhesus macaques: explicit versus implicit mechanisms. In Terrace and Metcalfe 2005, pp. 296-320.

- Terrace, H. and Metcalfe, J. (eds.) (2005). *The Missing Link in Cognition: Origins of Self-Reflective Consciousness*. New York: Oxford University Press.
- Tomasello, M., and Call, J. (2006). Do chimpanzees know what others see – or only what they are looking at? In Hurley and Nudds 2006, pp. 371-384.
- Toribio, J. (2007). Nonconceptual content. *Philosophy Compass*, 2/3, 445-460.
- Tulving, E. (1983). *Elements of Episodic Memory*. Oxford: Oxford University Press.
- Tulving, E. (2005). Episodic memory and autoevidence: uniquely human? In Terrace and Metcalfe 2005, pp. 3-56.
- Tye, M. (2006). The thesis of nonconceptual content. *European Review of Philosophy: The Structure of Nonconceptual Content*, 6, 7-30.
- Zentall, T. (2005). Animals may not be stuck in time. *Learning and Motivation*, 36, 208-225.

Notes:

¹ Theory-theory is usually contrasted with “simulation theory,” though many theorists hold some form of hybrid theory. There is much discussion about such “theories of mind” in the contemporary literature (Carruthers & Smith 1996; Nichols & Stich 2003; Goldman 2006). One issue centers on those who think that one's concepts of the mental are acquired through a process of simulating another's mental activity with one's own, and those who argue that some kind of background theory of mind is presupposed in the ability to mindread.

² Interestingly, Tulving acknowledges (p. 24) that KC clearly has meta-cognitive abilities and believes that he would pass various tests which measure of understanding of other minds (or “theory of mind”) abilities.

³ Earlier related results using a competitive paradigm are reported for chimps in e.g. Hare et al. 2001; Hare & Tomasello 2004; see also Tomasello & Call 2006.

⁴ See also Smith (2005, p. 257-60) for a very forceful similar line of argument. To give just one example on this point, Carruthers (2008) needs to posit the existence of a “gate-keeping mechanism” (among other capacities) to try to explain how it is possible to interpret much of the evidence cited above as involving only first-order (unconscious) mental states. And as Robert Lurz has impressed upon me (email correspondence), it even remains unclear why the gate-keeping mechanism is not itself meta-cognitive, despite Carruthers' claim to the contrary.

⁵ For much more independently defending the view that self-attribution of mental states is prior to our capacity to attribute mental states to others, see Goldman 2006. A more modest view, offered by Nichols and Stich (2003), is that the two capacities are independent and dissociable.

⁶ For more on this line of argument and varieties of the generality constraint, see Carruthers, this volume.

⁷ *Contra* Carruthers (2008) who cites well-known problems with infallibility (e.g. from Nisbett & Wilson 1977) as one reason to favor the opposing view. This is really not the issue; rather, it is whether or not it is typically *harder* to know about another’s mental state. In other words, aren’t we *more often* wrong when attributing mental states to others?