Department of History and Philosophy of Science

Anthropomorphism: Cross-Species Modeling

Sandra D. Mitchell
University of Pittsburgh

Anthropomorphism has been considered to be a bad word in science. It carries the stale dust of 19th century anecdotal evidence for the continuity of humans with non-human animals. Darwin claims that *"There can, I think, be no doubt that a dog feels shame… and something very like modesty when begging too often for food."* (1871/1981, p. 42 quoted in Knoll, p.14). But anthropomorphism is neither prima face bad, or necessarily non-scientific. It can be both, but it need not be either.

There has been a recent resurgence of interest in anthropomorphism, attributable no doubt to two developments - the rise of cognitive ethology and the requirements of various forms of expanded, environmental ethics. The first, investigation into the mental life and behavior of animals has a history clearly traceable from Darwin through Romanes and Tinbergen to modern ethologists. This is a straightforward scientific trajectory of theories and experiments designed to access causal explanations of behavior. Some of the most interesting and relevant work in this area has been directed at explaining the behavior of chimpanzees. Since it is generally agreed, that the chimp is our phylogenically closest relative, it makes evolutionary sense that the features of that species are more likely to be similar to features of our species than of those whose connection is more attenuated.

Darwin's and our love of dogs, notwithstanding, it is in primate research that the most plausible anthropomorphic theses are to be found. Or, as Daniel Povinelli claims in Folk Physics for Apes *"... if the argument by analogy cannot be sustained when it comes to behaviors that we share in common with our nearest living relatives, it can hardly be expected to survive more general scrutiny."* (p. 9). Indeed, as I will report later, Povinelli argues just this - that a strong version of anthropomorphism cannot be sustained in explaining some chimpanzee behaviors. This strong version of anthropomorphism found in cognitive ethology aims to explain behaviors of nonhumans by appeal to mental states similar to the ones we take to explain our own behavior. Of particular interest is the thesis that chimps have a "*theory of mind*", i.e. beliefs about the beliefs of others. Such second order beliefs are invoked to make sense of behavioral variation. For example a human would respond differently to two actors on the basis of beliefs about what those actors could see. If one of them had a clear view of a source of food, while the other's view of the food was blocked by a barrier, then it would make sense to follow any indication of food given by the one who you believe can see it and hence know where it is. Do chimps do the same thing? Do they do it for the same reasons? As I will discuss below, arguments from analogy and experimental results are brought to bear on answering this type of question.

The second source of interest in the similarities of humans and non-human animals arises from the animal rights and environmental ethics movements which have sought to transform the criteria by which we determine what beings merit moral consideration. Animal welfare and animal rights ethical positions make the nature of non-human experience determinate of who and what we must count in judging the moral correctness of our actions. The utilitarian version of animal welfare ala Peter Singer designates sentience or the ability to feel pain as sufficient for having interests that must be

considered. The deontological animal rights position, like that of Tom Regan, grants intrinsic value and moral rights to any being that has *"emotion, memory, desire, intentionality, sense of future, and some self-awareness."* Thus the feelings and cognitive states of non-human organisms is no longer just an academic question of whether or not the Rumbaugh's Kwansi has language, or dolphins can recognize themselves in a mirror (NYTimes May 1, 2001), but is rather a set of facts about the world that we need to know to ethically decide what to eat, and what to wear. Thus, the manner and degree to which non-human animals are similar to human beings becomes an even more pressing scientific problem in a context in which the very morality of our actions depends on the answer.

At its basis, anthropomorphism involves claims about the similarity of non-human objects or beings to humans. Strong anthropomorphism asserts that some description of a feature of human beings applies in the same way to a feature of a nonhuman animal. Critics of anthropomorphism often attack the presumptive character of such claims, like Darwin's lack of doubt of the internal nature of a dog's experience. Observers have been too willing to characterize non-humans using descriptive language that has humans as its primary referent. By describing a dog as feeling shame when it walks away with its tail between its legs, one is not gathering neutral data with which to test the myriad of theories about the nature of dogs, but lurking in the very description is an account that presumes dogs have mental or emotional states like human mental and emotional states. But what is at fault here? Is it the presumptiveness, or the anthropomorphism?

After all, similarity between humans and non-human animals is just what we should expect, on the basis of an evolutionary account of the origin and diversification of life on the planet. But not any similarity willy nilly. As a scientific claim about the facts of the world, any specific similarity between human immune systems, say, and mouse immune systems, or human beliefs and chimp beliefs must be grounded in more than a general truth of the continuum of life, and backed by more than an imposition of the same descriptive language.

In what follows I will consider a range of stances toward anthropomorphism from global rejections to specific models. The bumper sticker version of this talk could be: science made too easy is bound to be wrong. In the end I will argue that specific anthropomorphic theses are supported or not supported by the same rigorous experimental and logical reasoning as any other scientific model.

However, even though anthropomorphic models can be treated as science as usual, unique problems for these models still will remain. These problems have to do with the way in which language descriptive of our experiences travels back and forth between scientific to social domains.

I will first consider some global objections to anthropomorphism, these purport to be logical or conceptual transgressions that the act of describing nonhumans in human terms is supposed to commit. I will then look at empirical arguments for and against anthropomorphism. Finally I will consider some social contextual concerns that arise from the scientific anthropomorphic models.

Logical Objections

A. Anthropomorphism entails a category mistake. To speak of dogs with feelings of shame is like referring to a Bach Partita as being purple. This objection is easily dismissed as a relic of the view that humans are a separate and unique species - either created to be such, or so far evolved that no predicates true of us could be true of other organisms. Surely the evolution of life on the planet tells against this being a logical claim. For a Cartesian (see Spada) it might hold sway, but we are centuries beyond that.

B. Anthropomorphism is defined as the overestimation of the similarity of humans and nonhumans, and hence by definition could not yield accurate accounts. (see Guthrie and Lehman). But this is humpty-dumptyism. *"When I use a word,"* Humpty Dumpty said, in rather a scornful tone, *"it means just what I choose it to mean-neither more nor less."* (from Lewis Carroll, *Through the looking glass*)

If we choose to let "*anthropomorphism*" to be so defined, they we merely shift the question to be WHEN is it anthropomorphism and WHEN is it possibly a legitimate similarity? I.e. when does a relevant similarity hold such that describing a cognitive state as *"believing Sue cannot see the banana"* could be equally true of an adult human, a human infant and a chimpanzee. Just by definition, one cannot answer the substantive questions.

C. Anthropomorphism is necessary or unavoidable, since there is no amorphism or neutral language with which to describe behavior. If we don't use the predicates that describe us like *"believing X, wanting Y, deceiving Z"* for describing non-human animals, then we have to use language appropriate for machines, like *"moving towards the object, picking up the banana, looking towards the gate"* (see Spada). This position makes two mistakes. The first is that it presupposes a conceptual and linguistic impoverishment that is not justified. It underestimates our ability to discriminate and refer to multiple states of a system, many valued parameters. As recent research has suggested, we may end up thinking that chimpanzees don't have the same kind of mental representations that WE have, but nevertheless think they have mental representations that mediate their behavior. They are not input-output machines, but cogitating organisms. They just don't do it the way we do (see Povinelli). The second mistake is to confuse anthropocentrism with anthropomorphism. It is true that the descriptions we apply to anything are created BY us but they need not be OF us. That is we are the source of the terms and predicates but they need not be terms and predicates that apply to our behaviors.

If Anthropomorphism is not bad for logical reasons, then the extent of the acceptability of claims of similarity must be empirically grounded. This indeed, is the conclusion that many recent commentators on anthropomorphism have reached. Do chimpanzees have language like us; do they have beliefs about the beliefs of other chimps or of humans? Testing for the presence or absence of mental states, representations internal to the cognizing agent that are causing the behavior we can observe is no easy matter. I will now turn to the two main types of observational evidence that are used to justify anthropomorphism - the argument by analogy and experimentation.

Empirical Questions

Argument by analogy

An argument by analogy is invoked to support a claim about the unobserved features of one system - the target of the analogy - based on the presence of that feature in another system - the model system. The relevant similarities between the two systems are what justify the inference. Traditional analyses of analogical arguments render them fairly weak. They go something like:

P1. System M is observed to have features a, b, c
P2. System T is observed to have features a, b, c
P3. System M also is observed to have feature d
C. Therefore system T must have feature d

The strength of the analogy is sometimes rendered in terms of the number of similarities between the two systems. The more features in common, the more likely the target system will have the projected unobserved feature. But quantifying over similarities is notoriously difficult, and quite frankly, beside the point. Number of similar features does not portend the relevance of the similarities for the presence or absence of the feature of interest. Humans and mice are pretty different - and yet we are comfortable using the results of drug tests on mice to infer

the consequences of those drugs on our biochemistry. A more sophisticated rendering of the logic of analogical arguments, developed by Weizenfeld, suggests that the inference to the presence of the unobserved feature in the target system is deductive not inductive. The claim that system T has feature d is entailed by an assumption of the isomorphic causal structure in both systems that govern the generation of that feature. Thus when using information about the model system - the mice in drug experiments or, as we shall see human beliefs in anthropomorphic inferences to chimp beliefs - what determines the relevant similarities is the causal structures in those two systems - if they have the same structure, then the inference is sound. This is all rather abstract philosophy. Let's bring it back to the case at hand.

The analogical argument for inferring that chimps are like us is reconstructed by Povinelli as follows:

P1: I exhibit bodily behaviors of type B (i.e. those normally thought to be caused by 2 nd order mental states)
P2: Chimps exhibit bodily behaviors of type B
P3: My own bodily behaviors of type B are usually caused by my second order mental states of type A

C: Therefore bodily behaviors of type B exhibited by chimps are caused by their 2nd order mental states of type A, and so a fortiori chimps have 2 nd order mental states of type A.

On the traditional analysis of analogical arguments - what would make the conclusion strongly supported is the number of similarities between us and chimps. Here the phylogenetic proximity is brought to bear to say we have more similarities with chimps than other species since we are historically closer to them. Divergence occurred more recently - than to other species and hence we expect them to be more like us than, say toads or amoeba. But notice how weak this support actually is. Divergence is presumed, and distinction is required for humans and chimps to NOT be the same species. Lots of features may be shared, but just the ones we are interested in - 2nd order mental states, e.g. - may be just those that constituted the break in the lineage. So, evolutionary proximity may entail more similar features but not necessarily the relevant features.

The more sophisticated reading of analogical inference suggests a different understanding of the argument. Here what makes our experience relevant to inferences to chimp experience is not number of similarities, but isomorphic causal structures. What causes our behavior B is, supposedly our second order mental state A. But is this the same causal structures in chimpanzees? If it is, then even though we cannot ask the chimp what their belief is that caused behavior B, we can be justified in thinking that if the behaviors are the same, then the beliefs that cause them are the same. However, the $64,000 question remains as to whether or not the causal structures generating behaviors in humans and in chimpanzees are isomorphic. That is the subject of the second type of empirical evidence which I will discuss shortly.

To summarize, so far: Anthropomorphic theses can be seen as instances of analogical inferences. We project onto other organisms the features we take to be true of us. Phylogenetic relatedness seems to render weak support to the conclusion of such inferences, so weak that they can only garner some modest plausibility for the conclusions. However, a stronger analogical inference is supported when there is justification for isomorphism of causal structures in the two systems generating the features we are interested in. That analogy requires a different type of evidence than evolutionary history alone. Statistical and experimental data are required to support the premises that would entail this inference. So how can empirical evidence help?

Argument from experimental data

Advocates of cognitive ethology cry foul when their opponents reject the enterprise

from the beginning just for being anthropomorphic. Rather let the facts decide. But this is not as easy as it might sound. The controversial anthropomorphic theses project onto nonhumans just those sorts of features that are not directly accessible to observation. Bekoff and Allen want to know whether apes have a concept of death. Premack and Woodruff explore whether apes have a 'theory of mind' that is invoked in generating behaviors that appear to be acts of deception. We cannot just look at a chimpanzee and see its internal mental state. We cannot ask a chimpanzee to report to us the content of its cognition. We have access experimentally and observationally only to the very behaviors we take as the effects of the projected mental causes. So how can observation and experiment help decide this issue?

It is worth noting that the reason one suggests that concepts and second order beliefs might be the causes of nonhuman behaviors is because we believe that they are the causes of our own behaviors. So there is a causal structure or mechanism we can explore which generates behaviors as the effect of beliefs. When I think Raine is joking about where the dinner will be held, but Gregg is telling me the truth, then I don't go to the restaurant mentioned by Raine and I do go to the restaurant mentioned by Gregg. I hear the utterances of each of them, and my behavior is caused by not just those utterances, but also by my beliefs about the beliefs of the speakers. How do I know this? It is introspection or personal self-knowledge that gives me insight into the causal structure that underlies my actions. What if Elliott is standing next to me when both Raine and Gregg pronounce the restaurants where dinner is supposed to be held? Furthermore, when it is time for dinner, while I go out the door in the direction of Gregg's restaurant, Elliott goes the other way in the direction of Raine's restaurant. What do I think caused Elliott's behavior? I don't have the kind of subjective access to his internal mental states that I have to my own - I can only reason by analogy that the causes of his behavior is the same kind of things that causes my behavior. I don't think Elliott is a machine, I don't think he is being physically pulled in the direction he is moving, etc. I think that he must think that Raine was being serious and not joking and that this belief is a contributing cause of his behavior.

This is the same type of argument by analogy outlined above only applied to Elliott rather than to a chimpanzee. If the evidence for beliefs being the cause of behavior is solely the subjective experience of the believer/actor, then I need to project onto other human beings the possession of an unobservable mental cause to explain their reasoned behaviors. This is the well-known philosophical problem of "*other minds*". But the projection of unobservable mental causes to humans seems to be very much like the projection onto nonhuman beings. Why should it be sanctioned in the one case - other humans - but not sanctioned in another case - say, for honeybees - and where does that leave the inference when directed toward chimpanzee behavior?

There are two places to look for an answer to this question - background assumptions about the nature of intra and inter species similarity and behavioral experiments. I will first consider the background assumptions. There are good grounds to assume that basic causal structures or mechanisms are the same for different members of the same species of organism. Although different individual organisms are spatio-temporally distinct, and harbor all sorts of variation in specific features, the basic biological mechanisms most directly connected to surviving and reproducing are most likely to be the same. The reason is that these are the features upon which evolution by natural selection will have been quickest and strongest to act. Variations that have relatively negative effects on survival and reproduction are not kept around. That is how evolution by natural selection works. Even with the caveat of recognizing continual generation of variation within a

species, it nevertheless is a safe assumption that there will be little variation in the basic functioning of organisms within a species. The species is the correct boundary for this degree of similarity, because it is the interbreeding population that is the receptacle for the consequences of natural selection.

As I remarked above, a cause/effect relationship observed in one system (my subjective access to the beliefs that cause my action), can be inferred to apply to another system (seeing Elliott's actions, and inferring the mental cause) if I have grounds to think the two systems are causally isomorphic. Species membership, for the biological reasons given above, gives us the grounds for that inference. Indeed, there are other, observable physical similarities that we take to also substantiate the isomorphism of the psychological causal system within our species. These include the neurophysiological substrate of psychological causation - the structure of the brain and nervous systems, the sensory apparatus that detect features of the world what are then represented in beliefs, etc. In addition, the indirect access to internal mental states by the verbal reports of actors give us additional grounds to think that causal isomorphism holds. This is still no guarantee, as all the arguments about color perception, inverted spectra, etc. will attest, when a human being says "I believe the apple is red" there is no proof that their subjective experience of red is the same as any other individual who when looking at the same object utters the same sentence. Self-report of why one acts the way one does can also be challenged (see A. Gopnick). Nevertheless, there are good, fallible grounds for believing that other human beings have the same sort of second order beliefs that are causally relevant to their actions.

What is the objection to extending this inference from humans to nonhumans? First of all we have fewer types of supporting evidence than was the case in human to human inference. There is no self-reporting to be acquired from the chimp about the reasons for its actions. There is no shared species membership from which to support causal isomorphism. However, we can look to the similarity or dissimilarity of neurophysiological structure, sensory apparatus etc. And, importantly, we can look to behavioral observations and experimentation.

	Evolutionary relatedness	Self-reporting	Neuro-physiological and other physical features	Behavioral statistics
Human to Human inference	**Strong support**	**Strong support**	**Strong support**	**Strong support**
Human to chimp inference	**Weak support**	**N/a**	**Some support**	**Mixed support**

The experimental data on whether or not chimpanzees have second order beliefs, unfortunately, permits of multiple interpretations. Poveinelli's recent book *Folk Physics for Apes* reports a number of experiments done on captive chimpanzees over a five year period to investigate how they conceive of the physics that underlies their use of tools in particular, or more generally, to *"elucidate the nature of the mental representations that guide this behavior"* p. 1. In service to this goal, Povinelli provides evidence against the strong argument by analogy. A series of experiments were done to determine whether chimpanzees have the concept that others 'see'. This is a basic second order belief. I look at you and have a visual experience of you. I look at your eyes and notice that they are directed at the door. I form a belief that you see the door - i.e. a belief about your internal representation. I can then act on the basis of what I believe that you see or don't see. Povinelli's group studied whether chimpanzees engage in the same kind of cognitive process.

Povinelli begins his discussion by pointing out a number of components to an organism's forming a belief that another organism sees a particular object. The organism must notice the eyes of the other, and then follow the gaze of the other towards

the object under perception. Interest in eyes and gaze direction are present in a wide range of species and these abilities may well have emerged as adaptations to predation and social interactions. But how much like humans are the internal states of other organisms that engage in these behaviors? Povinelli puts the point clearly:

"…some researchers interpret the mutual gaze that occurs between infants and adults, as well as among great apes during complex social interactions as prima facie evidence of an understanding of the attentional aspect of seeing. And admittedly, there is a certain allure to the idea that, because mutual gaze in adult humans is often attended by representations of the mental states of others, comparable behavior in human infants (or other species) is probably attended by similar representations. But is mutual gaze in apes (for example) really attended by the same psychological representations as in human adults, or is this just a projection of our own way of thinking onto other species?" p. 22. In short, is this just wishful anthropomorphism, or can we get evidence that it is in fact the case?

The first step in Povinelli's study was to establish whether chimpanzees had the same behavioral abilities, i.e. gaze following, as do human infants and human adults. For the analogical argument to work the effects - behaviors in this case - expressed in the two systems have to be the same, and then one infers that the causes of these effects are also the same. Experiments show that chimps and 1½ yr. old human infants similarly responded to head movement, eye movement, left/right specificity, gaze following outside of visual field, etc. So chimps and humans engage in similar responses to a series of eye movement stimuli presented to them. Behaviors are the same. But what more is going on?

Povinelli devised ingenious experiments to try to test if chimp's gaze-following behavior indicated the possession of 2nd order mental states. The possible explanations considered were: a low-level and high-level account. The low-level account interprets the chimp's gaze following behavior to express cognition about behavioral propensities of the person whose gaze they followed, where the high-level account claims chimps form concepts about the internal mental states of the person whose gaze they are following. That is, the low-level model is similar to if you or I were following a billiard ball being hit by the cue ball and traveling down the felt. The high level model is as if you and I were watching someone look in the direction of moving car. The first case we just record correlated actions, in the second we form beliefs about what that person is seeing. The Povinelli group hypothesized that the high and low level accounts would make different predictions in cases where the observed individual's gaze was obstructed by an opaque barrier. If the low level account is right, the observing chimp would just scan a line from the eyes of the observed being until something was noticed. This is based on eyes looking right - something is right correlation, barrier is irrelevant. If the high level account were right, the observing chimp would walk around the barrier to see what was being seen. This would be based on eyes looking right - there must be something that is seen which is on the other side of the barrier. The results of an opaque-barrier test were unambiguously in support of the high level model. The chimps walked around the barrier to see what the person in the experiment was looking at. The conclusion naturally drawn is that chimps understand what it is for someone else to see - or represent the world; hence they have second order beliefs just like we do.

However, a dozen other experiments involving seeing supported the low-level model of cognition.. In these experiments the chimps were presented with two humans displaying different capacities to see them and it was observed whether the chimps responded differently to the two humans. The test response was begging behavior - did the chimps beg significantly more to the human

who did not have their gaze obstructed, or not? The conditions of the humans in the test included:

a. one blindfolded, one not;
b. one with a bucket over its head, one not
c. one with hands over their eyes, one not, and
d. one facing front and one facing back.

In three of the four conditions, the chimps were as likely to gesture to the person who could not see them as to the person who could. However, in the front facing vs. back-facing case they did beg more to the human with their front facing the chimp. So the low-level account captured experimental conditions a-c, whereas high level account was supported by experimental condition d. To try to distinguish whether it was the seeing that mattered or the front position, Povinelli introduced a fifth experimental set up. This time both humans had their backs to the observing chimp, but one was looking over her shoulder at the chimp, the other was not. *"To our surprise ....and in full support of the low-level model, on the looking-over-the-shoulder trials the apes did not prefer to gesture to the person who could see them."* P 34.

Povenilli's group continued to introduce new seeing/not seeing experimental conditions to the chimps using screens and eyes-open/eyes-shut conditions to try to figure out what was going on. In the end, Povinelli rejected the high level, 2nd order belief model and suggests that through trial and error the apes learned a set of procedural rules about successful gesturing (1. gesture to person whose front is facing forward, 2, if both fronts present or absent, gesture to person whose face is visible, and 3. if both faces visible or occluded, gesture to person whose eyes are visible). The chimpanzees do not appear to be using a concept of seeing to help them decide to whom to gesture. Instead, the chimpanzees after lots of trial and error behaved "*as-if*" they had our concept of seeing. Important for the conclusion is that the behavior at the end of the study was not how the chimps behaved at the beginning of the study. They learned how to gesture to the person we would say could see them. In contrast, human children were compared in these experiments and were shown to have the behaviors appropriate to understanding a concept of seeing from the beginning.

What do these experiments tell us about whether the similarity of chimp and human behavior indicate a similarity of internal mental cognition? Povinelli concludes that it is still open to interpretation. Indeed, he postulates three very different ways to account for the behaviors of the chimps in the experiments. First, they could have entered the test without a concept of seeing but through the testing came to construct the concept. Second they could have entered the test with a general conception of attention, and constructed a notion of visual attention. And third, they could have neither entered nor exited the tests with an understanding of the mental state of visual attention. Rather, they constructed an '*as-if*' understanding of seeing-as-attention. The third option is just clever Hans behavioral modification.

An anomaly for Povinelli's preferred low-level interpretation is that the opaque barrier tests did support the high-level model of cognition for the chimps. Povinelli takes the preponderance of evidence to suggest that the low level model is much better supported, and gives a re-interpretation of the opaque barrier test that would account for this contrary bit of evidence. On the way, he points out that if we walk into the laboratory with an anthropomorphic attitude, we are much more likely to continually refine and retest experimental results that support the low level model and accept on its face the results of tests like the opaque barrier one that support an anthropomorphic high-level model.

What conclusion should we draw from these experiments on nonhumans? Does the fact that their behavior and our behavior are sometimes indistinguishable indicate that the cause of those behaviors in us and in them are also the same? The experimental results are, at best, ambiguous, and according to Povinelli

lean towards a rejection of strong anthropomorphism. Indeed, as you will recall, he said if the similarity of human and nonhuman behaviors does not license the analogical inference to same causes for chimpanzees - it can hardly be credible for other species. At least it should be clear how very difficult it is to get unambiguous experimental results for anthropomorphic models.

What I suggested is that anthropomorphism can be a specific, scientifically accessible claim of similarity between humans and nonhumans. As such, it must be substantiated by evidence that there are similar causal mechanisms responsible for generating the apparently similar behaviors that are observed. If experimental and background theoretical support do provide that evidence, then there should be no objection to using the same descriptive language for both us and them. If that evidence is not provided, then using the same predicate for a full-fledged human behavior to refer to an "as-if" non-human behavior will be misleading and inaccurate.

It is clearly very difficult to get definitive evidence either way for even the simplest second order belief that A sees X. It gets progressively more difficult when the descriptions carry not just causal assumptions but also social and moral baggage. Let's return for a moment to Darwin's dogs. *"There can, I think, be no doubt that a dog feels shame… and something very like modesty when begging too often for food."* What would count as evidence that the behavior was an instance of shame? What is it for us to act out of a feeling of shame? It is not just that we have a belief about the physical world like, that there is a rock 2 meters in front of me, and I move in a way to avoid colliding with the rock. It is not just that we have a belief about another or my own belief, like there is a person 2 meters in front of me who is seeing an object that is blocked from my view. But shame must include something like an internal set of norms of appropriate behavior to which I compare my current behavior and find it wanting. For that to be accessible to behavioral experiments seems a long shot. Nevertheless something of the order of believing that A sees X does seem to be amenable to scientific investigation.