Chimpanzee minds: suspiciously human?

Daniel J. Povinelli and Jennifer Vonk

Cognitive Evolution Group University of Louisiana at Lafayette, 4401 W. Admiral Doyle Drive, New Iberia, Louisiana 70560, USA

Chimpanzees undoubtedly form concepts related to the statistical regularities in behavior. But do they also construe such abstractions in terms of mental states – that is, do they possess a 'theory of mind'? Although both anecdotal and experimental data have been marshaled to support this idea, we show that no explanatory power or economy of expression is gained by such an assumption. We suggest that additional experiments will be unhelpful as long as they continue to rely upon determining whether subjects interpret behavioral invariances in terms of mental states. We propose a paradigm shift to overcome this limitation.

Why do the minds of chimpanzees seem so human-like? We consider two answers; one of which has something profound to say about their minds, and another, which has something profound to say about our own. (In this article, we restrict our comments to chimpanzees, including both *Pan troglodytes* and *P. paniscus*, but the reader will apprehend that, in most cases, our remarks generalize beyond comparisons of humans and chimpanzees. Structurally similar problems face researchers who compare humans to other taxa, or who compare developmental transitions in infants and children.)

Mental similarity: real or apparent?

The first possibility is that the chimpanzee's mind seems similar to ours precisely because it *is* similar. Biological parsimony would seem to support such an assumption: chimpanzees and humans arose from a common ancestor about six million years ago. Alas, invoking biological parsimony will not help. After all, humans and chimpanzees are different in several other important ways, but this in no way denies their evolutionary relatedness. By way of analogy, the fact that some bats echolocate but their closest living relatives do not, hardly constitutes a crisis for evolutionary theory [1,2].

A second possibility exists, however: the human mind may have evolved a unique mental system that cannot help distorting the chimpanzee's mind, obligatorily recreating it in its own image. This idea should be taken seriously. After all, from change-blindness, to false memories, to cognitive dissonance, don't we already know the various ways in which the human mind systematically distorts its *own* workings?

It is worth noting that of the two lineages, the human one has undergone dramatic evolution since its divergence from the common ancestor. From head to toe, the human body is stamped with the legacy of those changes – changes that dwarf modifications that occurred in the chimpanzee lineage. Humans have resculpted the pelvis for bipedalism, evolved unique muscles in the hands for precision gripping, lost an opposable toe, tripled the size of the brain [3,4], and evolved the most complex system of communication that the planet has ever seen – natural language. Would it be any wonder, then, to discover that our minds also changed?

To gain some appreciation of the scope of evolutionary diversification that went on in the human lineage compared with the great apes, conduct the following thought experiment. Line up all the living representatives of the great ape/human clade, and then throw in the common ancestor. One of these six is immediately perceived as being not like the others. And that's not just from the human vantage point. As Alan Wilson and colleagues showed several decades ago [5], a frog would probably notice the differences as well.

Thinking about mental states

Let us examine these issues in the context of one of the most hotly-contested questions under comparative investigation: is the ability to conceive of the mental world a peculiarly human ability, or is it shared by other species, including, perhaps, chimpanzees?

To begin, chimpanzees (like humans), probably form abstract representations of the behavior of others. Each instance of another chimpanzee pursing his lips, hair bristling, need not be separately represented and understood. Rather, a concept of 'threat display' can be formed. Further, having witnessed a 'threat display,' the chimpanzee probably has a good sense of the sorts of things that will follow ('charging', 'being hit', etc.). This 'behavioral abstraction hypothesis' posits that chimpanzees: (a) construct abstract categories of behavior, (b) make predictions about future behaviors that follow from past behaviors, and (c) adjust their own behavior accordingly.

The question of theory of mind, then, becomes an *additional*, and quite focused one: do chimpanzees construe behavior in terms of mental states? Is the concept of 'gaze', for example, represented in both a behavioral code (abstracted spatio-temporal invariances) and a non-behavioral code (an attributed experience of 'seeing')? Although the former may well do much, if not most, of the actual work in supporting our behavioral interactions with others (and hence ought to be a greater focal point for research), if the latter is not present, then we have no business invoking the phrase 'theory of mind' [6].

Most scholars agree that despite cultural diversity, the core ability to think about mental states is a universal feature of the human mind [7]. Crucially, however,

Corresponding author: Daniel J. Povinelli (ceg@louisiana.edu).

http://tics.trends.com 1364-6613/03/$ - see front matter doi:10.1016/S1364-6613(03)00053-6

humans attribute more than just the mere existence of mental states such as thinking, knowing, wanting, and so forth (first-order mental states) to other beings. We also attribute to them the same *ability* to attribute these mental states to themselves and others (second-order mental states). That is, our folk psychology interprets certain behaviors as *prima facie* evidence that other individuals possess a theory of mind. This is why most researchers – and the public – are comfortable with a default hypothesis granting chimpanzees, and other animals, a theory of mind: they exhibit precisely those behaviors that our folk psychology is designed to interpret in that way.

The reinterpretation hypothesis

At this point, some readers may be scratching their heads: 'Okay, you think that humans automatically interpret certain behaviors as evidence of theory of mind, and you agree that this process works reasonably well when the agents are other humans. So why shouldn't it work equally well with chimpanzees?'

The answer is simple: 'Because we can easily imagine that theory of mind uniquely evolved in humans!' Humans and chimpanzees undoubtedly inherited common mental structures for forming behavioral abstractions. During the course of hominid evolution, however, our lineage might have woven a new, 'theory-of-mind system' into our ancestral cognitive architecture in such a way that the new and old systems now exist in perfect harmony alongside each other – just as the numerous systems sub-serving echolocation in bats were woven in alongside mental systems that they share with their closest relatives.

We have labeled this the 'reinterpretation hypothesis' to emphasize that in this scenario, the capacity for behavioral abstraction was already present in the common ancestor, and that humans added another system for coding the behaviors in an additional, mentalistic fashion [8,9].

Consider the full force of this idea. If second-order mental states evolved in this manner, it would mean *by definition* that humans and chimpanzees must each represent, reason about, and ultimately produce, a very similar set of behaviors – but behaviors that humans additionally explain in terms of mental states. The fact that humans interpret certain constellations of behavior as evidence of theory of mind would thus be a trivial byproduct of the fact that theory of mind evolved by exploiting the existing systems for behavioral abstraction. Further, it would virtually guarantee the hegemony of the human theory of mind: there is no way that the human mind could avoid misconstruing the chimpanzee's mind.

Now, let us examine what implications the 'reinterpretation hypothesis' has for the evidence – both anecdotal and experimental – that has been marshaled to support the idea that chimpanzees possess a theory of mind.

Deceived into believing: what's really wrong with anecdotes

The most widely celebrated evidence for second-order mental states in chimpanzees is their ability to manipulate each other [10–12]. The complexity of at least certain instances of chimpanzee 'deception' has frequently tempted the conclusion that the most plausible interpretation is that they are reasoning about what each other see, want, know and believe [12,13]. Other writers have balked, however, noting that it is trivial to imagine how organisms could produce deceptive behaviors without reasoning about mental states [14–16].

Because of this stalemate, the anecdotal approach has gradually fallen out of favor, giving way to a predominantly experimental approach. Unfortunately though, because the most fundamental problem associated with the use of anecdotes was never widely identified, the same conceptual problem has crept, almost unnoticed, into our experiments. Let us therefore briefly examine the spontaneous, non-experimentally-derived behaviors of chimpanzees, and determine what they can and cannot tell us – and why.

Consider any anecdote about deception by a chimpanzee. Question: 'What does the deceptive chimpanzee want?' Answer: 'To get another chimpanzee *to behave* in a certain way' (to move away from a piece of food, to turn in a particular direction, to not approach, etc.). And so, the deceiver does something to bring about this behavior: he turns in a particular direction or walks away, positions himself behind some obstruction, and so on. Thus, on any account of the chimpanzee's mind, the deceiver knows how his or her own behavior will affect the *behavior* of others.

From here, we are typically drawn into a debate over which of two accounts is more parsimonious [10,13–15,17,18]. One side is cast as the skeptic: 'Humans sometimes manipulate others by manipulating their mental states, but chimpanzees do it by learning all sorts of individual stimulus–response chains.'

The other side is cast as the naive believer: 'Yes, it's obvious that chimpanzees and humans *could* be doing it different ways, but it's more biologically plausible to assume that they're doing it the same way. Further, it's more parsimonious (economical) to posit that chimpanzees are reasoning about mental states than to suppose that they are representing individual links between particular behaviors and responses. Thus, *for the same reasons* that we ultimately rejected behaviorism, we should reject the skeptic's assertion.'

The skeptic is wrong to suggest that the only alternative to attributing a theory of mind is to accept the tenets of behaviorism (i.e. positing that the chimpanzee has no mental representations), but the believer's invocation of parsimony (economy of expression) constitutes an error in logic: for each anecdotal instance of deception in which a chimpanzee might have been reasoning about the mental states of others, the agent must *also* have possessed a corresponding behavioral abstraction that could have done the same work.

Thus, unless the behavioral-abstraction hypothesis is rejected, those who believe that deceptive chimpanzees possess a theory of mind must postulate two things: first, that they possess behavioral abstractions, and second, that they possess representations of mental states! Here the lack of analogy with the behaviorism debate becomes apparent: everyone agrees that the chimpanzee's mind contains mental representations – that is, intervening variables. The question is: are these intervening variables representations of behavioral abstractions *and* mental states (as theoretical entities), or behavioral abstractions alone?

Importantly, we are not denying the possibility of constructing an imaginary dataset in which an economy of expression can be gained by postulating second-order mental states as intervening variables (e.g. see [18]. Rather, we contend that for the total corpus of chimpanzee behavioral data to be explained, positing that chimpanzees possess intervening variables that are *behavioral abstractions* will suffice. Consider instances of 'partially hiding from view', a form of deception that Whiten and Byrne thought might 'rule out the use of…first-order representations' (Ref. [19] pp. 215–216). For example, a subordinate male chimpanzee conceals his erect penis (which signals his intention to mate) with his hand, while monitoring the dominant male who remains present. But is the agent representing both the dominant's behavior (turning away) *and* his mental state (seeing/not seeing), or just his behavior – and how would we know given that reasoning about the mental state presupposes reasoning about the behavior? Further, because each anecdotal instance of deception assumes the presence of behavioral abstractions, we *lose* economy of expression by also assuming the presence of second-order mental states.

Thus, the real problem with the anecdotes is *not* that it is unparsimonious to account for chimpanzee deception by appealing to associative learning models (i.e. that the behaviors were 'shaped', 'reinforced', or otherwise 'learned'). Instead, the problem is that each anecdote presupposes a behavioral abstraction on the basis of which a mental state is inferred, without specifying what unique causal work the second-order mental state performs. Put simply: anecdotes cannot resolve whether the intervening variable is an invariant category of behavior coupled with an inference about a mental state, or a behavioral invariant alone.

Spinning our experimental wheels?

It directly follows that any experiments that rely upon a behavioral abstraction will be of little use, especially when this invariant is one the subject has previously witnessed, or that they are likely to have evolved to detect and exploit. Indeed, contrary to recent speculations [20], behavioral interactions that make the *most* ecological sense to the organism are precisely the ones that will be *least* diagnostic of whether the organism is reasoning about mental states and behavior, or behavior alone [21]. After all, these are the contexts for which evolution is most likely to have sculpted special-purpose, highly-focused behavioral representations for use by the organism.

Consider recent experiments by Hare, Call and Tomasello [22], (and see [23], this issue) who have tried to use the context of food competition to determine whether chimpanzees understand the connection between 'seeing' and 'knowing'. A subordinate and dominant were positioned on either side of an empty room from each other, temporarily prevented from entering by doors which could be opened: either slightly, to let them look into the room, or all the way, to let them enter. An experimenter placed food into one of two cups between them. The subordinate's door was opened first, giving him or her a head start. When the subordinate, but not the dominant, was allowed to observe the baiting, the subordinate frequently approached the food; when *both* the subordinate and dominant observed the food being hidden, the subordinate was less likely to approach the food (but see [24]).

Does this establish that the subordinate understands the connection between seeing and knowing? Unfortunately not, because a behavioral abstraction serves as the basis for a theory-of-mind coding: 'Don't go after the food if that dominant has oriented towards it $<$*because he has seen it, and therefore knows where it is*$>$.'

A 'control' condition was also implemented in which both the subordinate and dominant watched the baiting, but the location of the food was then switched. On half the trials, the subordinate alone observed this switch; on the other half, both observed. The subordinates were more likely to search when only they saw the move (as opposed to when the dominant saw the move as well). But again, the conflation of behavioral abstraction and mental-state attribution is obvious: 'He was present and facing the food when it was placed where it is now $<$*so he saw the food placed and currently knows where it is*$>$ therefore he is likely to go after it'; and 'He was not present when the food was placed where it is now $<$*so he didn't see, therefore he doesn't know…*$>$ therefore he is less likely to go after it.'

Can additional 'controls' help? Hare *et al.* [22] used a similar setup in which one dominant (Joe) watched the baiting of food, but then when the door opened it was a different dominant – Mary. The authors' *own* theory of mind interpreted their experiment as follows: if the subordinate is more willing to approach the food when the new dominant is present, they must be reasoning, 'Well, Joe *saw* the food placed so he *knows* where it is…But look, it's Mary! She *didn't see* the food being placed, therefore she *doesn't know*…' But, of course, that ignores that an intelligent chimpanzee could simply use the behavioral abstraction (upon which the additional theory-of-mind coding depends): 'Joe was present and oriented; he will probably go after the food. Mary was not present; she probably won't.'

It is tempting to think that we can remedy these failings of the current line of experiments by simply implementing more or better controls. However, the problem is not the ingenuity of the experimenters; it is the nature of the experiments. Techniques that pivot upon behavioral invariants (looking, gazing, threatening, peering out the corner of the eye, accidentally spilling juice versus intentionally pouring it out), will always presuppose that the chimpanzee (or other agent) has access to the invariant, thus crippling any attempt to establish whether a mentalistic coding is also used. The sobering point is that *no experiment in which the theory-of-mind coding derives from a behavioral abstraction will suffice*. Control will chase control with no end in sight, leaving only our intuitions, hopelessly contaminated by our folk psychology, to settle the matter.

A conceptual solution: experiential mapping from self to other

Drawing on isolated research trends, we propose a shift to paradigms that develop and deploy techniques requiring subjects to make an extrapolation from their own experiences to the mental states of others. Subjects must be given an experience that they could not otherwise have predicted from

the environment, and then researchers must determine whether they understand the nature of that experience.

The notion that 'theory of mind' involves, at its foundation, using one's own experiences to model the experiences of others is not new; it forms the basis of certain simulationist accounts of knowledge of other minds (e.g. [25–27]). Gallup was an early advocate of this general approach for assaying the presence of theory of mind in other species [21], and several studies have approximated this design – some with chimpanzees (M.S. Novey, unpublished doctoral dissertation, Harvard University, 1975), some with human infants and children (J.A. Sommerville, unpublished doctoral dissertation, University of Chicago, 2002). We propose that this approach, with key strictures, become recognized as a way of escaping the logical problems inherent in anecdotes and the experimental approaches currently in vogue.

Let us consider an example. Povinelli and colleagues [28,29] gave chimpanzees experiences with opaque buckets and blindfolds and cardboard screens, and then, in a food-begging context, confronted them with two familiar experimenters, one who could see them and one who could not. For example, one choice was between someone with a bucket over her head and someone holding a bucket on her shoulder. Interestingly, they did not prefer to gesture to the person who could see them, *but even if they had*, we would face the same problem we isolated above: the subjects certainly had ample opportunity to form abstractions about how others behave with their face or eyes obstructed. Thus, merely providing the self-experience is not sufficient. Two other conditions must be met: (1) the cue on which the inference to the mental states is made must be arbitrary, and (2) the subject must have no exposure to others behaving in association with that cue.

Adapting a suggestion by Heyes [14], imagine that we let a chimpanzee interact with two buckets, one red, one blue. When the red one is placed over her head total darkness is experienced; when the blue one is similarly placed, she can still see. Now have her, for the first time, confront others (in this case the experimenters) with these buckets over their heads. If she selectively gestures to the person wearing the blue bucket we could be highly confident that the nature of her coding was, in part, mentalistic – that is, that she represented the other as 'seeing' her. Our point is not to advocate a particularly 'good' or 'clever' test. Rather, the test exemplifies a class of experiments that could in theory escape the problem of the power of the behavioral abstraction system.

Although, first and foremost, we advocate this new experimental program, there is another, more subtle change that is needed for further progress: the idea that theory of mind is the 'holy grail' of comparative cognition needs to be abandoned. Neither chimpanzees nor evolutionary theory will be insulted if the very idea of 'mental states' turns out to be an oddity of our species' way of understanding the social world.

Acknowledgements

We gratefully acknowledge that this work was supported by a James S. McDonnell Centennial Award to D.J.P. We also thank Steve Giambrone for thoughtful comments on some of the ideas in this article.

References

1 Hutcheon, J.M. *et al.* (2002) A comparative analysis of brain size in relation to foraging ecology and phylogeny in the Chiroptera. *Brain Behav. Evol.* 60, 165–180
2 Legendre, P. and Lapointe, F.J. (1995) Matching behavioral evolution to brain morphology. *Brain Behav. Evol.* 45, 110–121
3 Aiello, L. and Dean, C. (1990) *An Introduction to Human Evolutionary Anatomy*, Academic Press
4 Fleagle, J.G. (1999) *Primate Adaptation and Evolution*, Second Edition, Academic Press
5 Cherry, L.M. *et al.* (1978) Frog perspective on the morphological difference between humans and chimpanzees. *Science* 200, 209–211
6 Premack, D. and Woodruff, G. (1978) Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 515–526
7 Lillard, A. (1998) Ethnopsychologies: Cultural variations in theories of mind. *Psychol. Bull.* 123, 3–33
8 Povinelli, D.J. and Giambrone, S. (1999) Inferring other minds: failure of the argument by analogy. *Philos. Topics* 27, 167–201
9 Povinelli, D.J. and Giambrone, S. (2000) Escaping the argument by analogy. In *Folk Physics for Apes* (Povinelli, D., ed.), pp. 9–72, Oxford University Press
10 Whiten, A. and Byrne, R.W. (1988) Tactical deception in primates. *Behav. Brain Sci.* 11, 233–244
11 Byrne, R.W. and Whiten, A. (1992) Cognitive evolution in primates: evidence from tactical deception. *Man* 27, 609–627
12 Whiten, A. (1997) The Machiavellian mindreader. In *Machiavellian Intelligence II: Extensions and Evaluations* (Whiten, A. and Byrne, R.W., eds) pp. 144–173, Cambridge University Press
13 de Waal, F.B.M. (1991) Complementary methods and convergent evidence in the study of primate social cognition. *Behaviour* 118, 297–320
14 Heyes, C.M. (1998) Theory of mind in nonhuman primates. *Behav. Brain Sci.* 21, 101–148
15 Kummer, H. *et al.* (1990) Exploring primate social cognition: some critical remarks. *Behaviour* 112, 84–98
16 Premack, D. (1988) 'Does the chimpanzee have a theory of mind?' revisited. In *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans* (Whiten, A. and Byrne, R.W., eds) pp. 160–179, Oxford University Press
17 Savage-Rumbaugh, S. and McDonald, K. (1988) Deception and social manipulation in symbol-using apes. In *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans* (Whiten, A. and Byrne, R.W., eds) pp. 225–237, Oxford University Press
18 Whiten, A. (1994) Grades of mindreading. In *Children's Early Understanding of Mind: Origins and Development* (Lewis, C. and Mitchell, P., eds) pp. 47–70, Erlbaum
19 Whiten, A. and Byrne, R.W. (1988) Tactical deception of familiar individuals in baboons. In *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans* (Whiten, A. and Byrne, R.W., eds) pp. 205–223, Oxford University Press
20 Hare, B. (2001) Can competitive paradigms increase the validity of experiments on primate social cognition? *Anim. Cogn.* 4, 269–280
21 Gallup, G. (1985) Do minds exist in species other than our own? *Neurosci. Biobehav. Rev.* 9, 631–641
22 Hare, B. *et al.* (2001) Do chimpanzees know what conspecifics know? *Anim. Behav.* 61, 139–151
23 Tomasello, M. (2003) Chimpanzees understand psychological states: the question is which ones and to what extent. *Trends Cogn. Sci.* 7, 153–156
24 Karin-D'Arcy, R.M. and Povinelli, D.J. Do chimpanzees know what each other see? a closer look. *Int. J. Comp. Psychol.* (in press)
25 Gordon, R. (1986) Folk psychology as simulation. *Mind Lang.* 1, 158–171
26 Harris, P.L. (1991) The work of the imagination. In *Natural Theories of Mind: The Evolution, Development and Simulations of Everyday Mindreading* (Whiten, A., ed.), pp. 283–304, Blackwell
27 Goldman, A. (1993) The psychology of folk psychology. *Behav. Brain Sci.* 16, 15–28
28 Povinelli, D.J. and Eddy, T. (1996) *What Young Chimpanzees Know About Seeing*. Monographs of the Society for Research in Child Development, 61, (Serial No. 247)
29 Reaux, J.E. *et al.* (1999) A longitudinal investigation of chimpanzees' understanding of visual perception. *Child Dev.* 70, 275–290