

Theory of mind in dogs? Examining method and concept

Alexandra Horowitz

© Psychonomic Society, Inc. 2011

Abstract In line with other research, Udell, Dorey, and Wynne's (in press) finding that dogs and wolves pass on some trials of a putative theory-of-mind test and fail on others is as informative about the methods and concepts of the research as about the subjects. This commentary expands on these points. The intertrial differences in the target article demonstrate how critical the choice of cues is in experimental design; the intersubject-group differences demonstrate how life histories can interact with experimental design. Even the best-designed theory-of-mind tests have intractable logical problems. Finally, these and previous research results call for the introduction of an intermediate stage of ability, a *rudimentary* theory of mind, to describe subjects' performance.

Keywords Comparative cognition · Theory of mind · Domestic dog · Attention · Cognitive ethology · Social learning

Why test for theory of mind in nonhuman animals? Conceptually, the interest is comparative: Do any animals behave in a way that suggests that they are, like humans, aware that other creatures have minds just as they do? Experimentally, research looks for evidence that an animal can predict the behavior of others (Premack & Woodruff, 1978) or can attribute mental states, particularly desire or knowledge, to others (Carruthers & Smith, 1996; Heyes, 1998).

The many tests for theory of mind in humans are linguistic in form and are passed by normally developing children by around age three. By contrast, a definitive test

for theory of mind in non-language-using animals has been elusive, and no animal has been uniformly successful at those that exist. For instance, Call, Hare, and Tomasello (1998) found that no chimpanzee was able to follow a human's gaze, and Povinelli and Eddy (1996) found that chimpanzees were unable to use eye information. However, Hare, Call, Agnetta, and Tomasello (2000) found that chimpanzees used gaze of conspecifics to get information about whether food was accessible.

Adding to this literature, Udell, Dorey, and Wynne (in press) report that canids—hand-raised wolves, pet domestic dogs, and nonpet domestic dogs—perform equivocally on one putative theory-of-mind test. In this target article, Udell et al. (in press) used a *begging* paradigm, a so-called *perspective-taking* task originated by Povinelli and Eddy (1996). Each subject group passed the test in at least one design (back turned/facing forward), but all also failed on one or more of the other designs. (Interestingly, the wolves' performance, considered passing in the target article, is much like the “failing” chimpanzees in Povinelli and Eddy's original begging experiment.)

What explains the canids' (or chimpanzees') divergent performance on theory-of-mind tests, as well as the canids' occasional outperformance of nonhuman primates? I will concentrate on methodological and conceptual issues that emerge from the target article that are relevant to interpreting behavior in theory-of-mind tests.

Relevance of design and life histories

The first significant contribution of the target article is that it draws attention to the importance of the objects used in a task given to a nonhuman animal. To the human eye, the difference between the trials was trivial: Each was designed so that one experimenter could see the subject (experimenter's eyes were

A. Horowitz (✉)
Department of Psychology, Barnard College,
3009 Broadway,
New York, NY 10027, USA
e-mail: ahorowit@barnard.edu

visible) and the other could not see the subject (experimenter's eyes were obstructed). To the canid eye, apparently, the difference between the designs was not trivial. The subjects' performance changed as the obstructing item changed. This fact points to the salience of every element of the design: To the subjects, a face hidden by a *book* is interpreted differently than a face hidden by a *bucket*.

It is safe to assume that this finding reflects life history differences between the different subject groups and, perhaps, even members of the same group. Animals' knowledge of objects and the objects' properties is tied to their exposure to the object in development. It is likely that a dog who, exposed to a book-loving owner, comes to understand that a *book* is opaque, for instance, may not extend that property to a novel object like a bucket. If so, rather than testing the subjects' understanding of "seeing," the experimental design is testing the subjects' understanding of buckets and books. In a similar vein, the background of tested dogs in other recent research was found to be material to the rate of production of behaviors associated with the "guilty look": Those dogs who had participated in obedience classes showed more of these behaviors, across conditions, than did dogs who had not (Horowitz, 2009a). Perhaps the expression of submission is learned from exposure in the training context—just as the import of a *book* to a human's attention may be learned by exposure thereto in the owner's home. A comparable point could be made of the primate studies as well, which variously test enculturated, lab-raised, or captive zoo animals with varied human, conspecific, and experimental backgrounds.

Relatedly, the performance of the shelter dogs and hand-raised wolves points to the importance of inter- or intraspecific social experience in experimental performance. In the case of the shelter dogs, the authors have almost no information about the dogs' life histories and experience with humans. It is possible, though, that in shelters, dogs may learn that the attention of humans is *actually irrelevant* to them, since they are regularly ignored by workers and visitors who pass by.

As the authors state, even the limited task success of the wolves suggests that domestication is not necessary for dogs' social-cognitive skills. But domestication may still be sufficient, with normal social exposure (which the shelter dogs may not receive). A strong case against domestication could be made only if the test were done with very young dogs—when they have the visual acuity and cognitive stamina to be tested, but before they are socialized into human lifestyles.

Theory-of-mind tasks

A second question raised by the target article is whether the begging task is properly described as a *theory-of-mind* test

at all. In fact, in all of the seven described categories of theory-of-mind methodologies (Heyes, 1998), there is a logical problem inherent in the designs. Experiments are designed on the premise that "if one has a theory of mind," then one should do some specified behavior (such as "select a seeing person over a nonseeing person"). But, logically, successful performance allows no valid conclusion to be drawn about theory of mind. There is room for any number of other abilities that might account for the observed behavior. (And it has often been observed that performance on the begging task, in particular, could as easily be explained by noncognitive processes [Heyes, 1998].)

Even failure at a test designed with this premise does not justify a negative conclusion, since, as we have seen, failure might reasonably be explained as indicative of problems in experimental design.

Ultimately, the mixed performance observed across trials, in the target article and in previous studies, suggests that the syllogism that underlies the experimental design should be reconsidered. If a test is a true test of theory of mind, the subject who can pass one trial (such as *book*) should pass a related trial (such as *bucket*), and performance should be 100%. But even the highest-performing subjects never pass *on every trial*, which is what we would expect for subjects with a theory of mind.

None of the canid research to date has made a claim to have proven (or disproven) that dogs have a theory of mind. However, using the experimental paradigms developed thus far, there is no chance that any research *could* legitimately make such a claim.

A rudimentary theory of mind

Dogs (and now, in some cases, wolves) have repeatedly been shown to differentiate others on the basis of attentional state (Call, Bräuer, Kaminski, & Tomasello, 2003; Schwab & Huber, 2006). The target article complements the findings of a naturalistic study of intraspecific play behavior focusing on dogs' use of attention cues in communication (Horowitz, 2009b). In this study, the observed dogs moved to get visual attention before communicating (signaling play intent), and the dogs' use of attention-getters discriminated appropriately between audiences according to state of attentiveness (using more forceful attention-getters when the recipient was most inattentive, as when playing with someone else, and less forceful ones when his attention was mildly diverted). Recent research has shown that captive wolves in a German zoo show similar attention-getting skills in social dyadic play (Ganslößer & Wolf, 2010).

The canids' sensitivity to different attention states and their differential use of attention-getters suggest, as one

possibility, that dogs (and perhaps wolves) recognize that there is an unseen feature of others mediating their behavior and that it is relevant to interacting with others. We call this unseen feature *attention*, and it is often implicated in human cognition. But an animal need not have a human's understanding of the concept to have a functional understanding of its relevance to others' behavior.

An explanation for these results is that there may be *degrees* of theory of mind (Gómez, 1996; Whiten, 1997), and, indeed, the subjects' varied performance across trials suggests that they are demonstrating an intermediate ability. Heretofore, the representation of theory of mind as "all or nothing" has been tacit. But both the fact that humans *develop* a theory of mind and the diverse social abilities seen among species argue that theory of mind is not best considered a unitary ability that one has either completely or not at all (Allen, 1998; Horowitz, 2002).

I propose that the dogs' behavior in play and in the target article may best be described as showing a *rudimentary theory of mind*. This concept fills an explanatory gap. It is important to acknowledge the rather broad territory between an animal who acts with a *theory of behavior* (learning that certain behaviors lead to certain outcomes) and one who has a *theory of mind*. In particular, if the animal seems to be operating with regard to some mediating element between others' appearance and their behaviors, this behavior could be described as a rudimentary theory of mind.

Indeed, it has been suggested that an understanding of "intervening variables" affecting others' behavior is almost entailed by association in complex social groups (Humphrey, 1980). If so, the elaborate socialization not only of primates, but also of some herd and pack animals raises the prospect that an intermediate form of theory of mind may have formed among many animals. It would be advantageous to have such an ability to coordinate with or interpret others' behavior. As an adaptive skill, such an ability might be more easily found in research of naturally occurring behavior, rather than in experiments alone.

A rudimentary theory of mind explains much primate—and now canid—behavior that is equivocal on so-called theory-of-mind tests. For those tests that manipulate experimenters' "attention," the rudimentary theory of mind concept covers the broad territory between various levels of attention, from having attention at all to having an explicit theory of attention. In between, there are many stages of skill at using or understanding attention, including (but not limited to) noticing attention in others (as in mutual gaze), the linking of attention to the provision of information (as in gaze following), the manipulation of others' attention to change their attentional state (attention getting), using attention to change others' informational state (showing), and use of attention in a broader context (attention getting in communication).

Conclusion

With this expanded notion of theory of mind, we can turn from focusing on criterial theory-of-mind tests to compiling a full inventory of the skills seen in a species, allowing us to compare species' behavior directly. There remains a gap between wolf and dog behavior. The animals' learning styles is telling. In the target article, when trained when rewarded for avoiding begging to a bucket-wearing human, half of the naive dogs quickly learned the association. Only one of eight wolves, who had seen the task before, learned the pairing. One dog, exposed to instances when the bucket indicated that a human *would* provide a treat, actually learned this counterintuitive pairing. Similarly, shelter dogs have been shown to quickly learn new skills from an unfamiliar human (Udell, Dorey, & Wynne, 2010).

Perhaps the salient difference between dogs and wolves is the *specificity* of considering the human. In particular, dogs are more predisposed than nondomesticated canids to be *interested in* humans, to look at humans, and to note the details of humans (including our eyes). Notably, researchers who work with hand-raised wolves have reported that the wolves use eye contact differently than does the average dog. They are more likely than wild wolves to make eye contact with (familiar) humans. But like most wolves, they appear to be more susceptible to view sustained eye contact as threatening than pet dogs are (Patricia Goodman, personal communication).

The target article should provoke researchers to consider the life histories of their subjects and the salience of the objects chosen in experimental design. I encourage experimental and observational exploration of theory of mind, with a broadened conception of the range of rudimentary abilities.

References

- Allen, C. (1998). Assessing animal cognition: Ethological and philosophical perspectives. *Journal of Animal Science*, *76*, 42–47.
- Call, J., Bräuer, J., Kaminski, J., & Tomasello, M. (2003). Domestic dogs (*Canis familiaris*) are sensitive to the attentional state of humans. *Journal of Comparative Psychology*, *117*, 257–263.
- Call, J., Hare, B., & Tomasello, M. (1998). Chimpanzee gaze following in an object choice task. *Animal Cognition*, *1*, 89–99.
- Carruthers, P., & Smith, P. K. (Eds.). (1996). *Theories of theories of mind*. Cambridge: Cambridge University Press.
- Ganslöber, U., & Wolf, V. (2010, July). *Attention-getting behaviour and play signals of Canadian wolves (Canis lupus hudsonicus) at Wuppertal zoo*. Paper presented at the Second Canine Science Forum, Vienna, Austria.
- Gómez, J. C. (1996). Nonhuman primate theories of (nonhuman primate) minds: Some issues concerning the origins of mind-reading. In P. Carruthers & P. K. Smith (Eds.), *Theories of*

- theories of mind* (pp. 330–343). Cambridge: Cambridge University Press.
- Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*, 771–785.
- Heyes, C. M. (1998). Theory of mind in nonhuman primates. *The Behavioral and Brain Sciences*, *21*, 101–134.
- Horowitz, A. C. (2002). *The behaviors of theories of mind, and a case study of dogs at play*. Ph.D. dissertation, University of California, San Diego.
- Horowitz, A. (2009a). Attention to attention in domestic dog (*Canis familiaris*) dyadic play. *Animal Cognition*, *12*, 107–118.
- Horowitz, A. (2009b). Disambiguating the “guilty look”: Salient prompts to a familiar dog behaviour. *Behavioural Processes*, *81*, 447–452.
- Humphrey, N. K. (1980). Nature’s psychologists. In B. Josephson & V. Ramachandran (Eds.), *Consciousness and the physical world* (pp. 57–80). Oxford: Pergamon.
- Povinelli, D., & Eddy, T. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development*, *61*, 247.
- Premack, D., & Woodruff, G. (1978). Does a chimpanzee have a theory of mind? *The Behavioral and Brain Sciences*, *1*, 515–526.
- Schwab, C., & Huber, L. (2006). Obey or not obey? Dogs (*Canis familiaris*) behave differently in response to attentional states of their owners. *Journal of Comparative Psychology*, *120*, 169–175.
- Udell, M. A. R., Dorey, N. R., & Wynne, C. D. L. (2010). Wolves outperform dogs in following human social cues. *Animal Behaviour*, *76*, 1767–1773.
- Udell, M. A. R., Dorey, N. R., & Wynne, C. D. L. (in press). Can your dog read your mind? Understanding the causes of canine perspective taking. *Learning & Behavior*.
- Whiten, A. (1997). The Machiavellian mindreader. In A. Whiten & R. W. Byrne (Eds.), *Machiavellian intelligence II: Extensions and evaluations* (pp. 144–173). Cambridge: Cambridge University Press.