'Theory of mind' in animals: ways to make progress

Elske van der Vaart · Charlotte K. Hemelrijk

Abstract Whether any non-human animal can attribute mental states to others remains the subject of extensive debate. This despite the fact that several species have behaved as if they have a 'theory of mind' in various behavioral tasks. In this paper, we review the reasons of skeptics for their doubts: That existing experimental setups cannot distinguish between 'mind readers' and 'behavior readers', that results that seem to indicate 'theory of mind' may come from studies that are insufficiently controlled, and that our own intuitive biases may lead us to interpret behavior more 'cognitively' than is necessary. The merits of each claim and suggested solution are weighed. The conclusion is that while it is *true* that existing setups cannot conclusively demonstrate 'theory of mind' in non-human animals, focusing on this fact is unlikely to be productive. Instead, the more interesting question is how *sophisticated* their social reasoning can be, whether it is about 'unobservable inner experiences' or not. Therefore, it is important to address concerns about the setup and interpretation of specific experiments. To alleviate the impact of intuitive biases, various strategies have been proposed in the literature. These include a deeper understanding of associative learning, a better knowledge of the limited 'theory of mind' humans actually use, and thinking of animal cognition in an embodied, embedded way; that is, being aware that constraints *outside* of the brain, and *outside* of the body, may naturally predispose individuals to produce behavior that looks smart without requiring complex cognition. To enable this kind of thinking, a powerful methodological tool is advocated: Computational modeling, namely agent-based modeling and, particularly, cognitive modeling. By explicitly simulating the rules and representations that underlie animal

E. van der Vaart (✉) · C. K. Hemelrijk
Behavioural Ecology and Self-Organisation, University of Groningen, Groningen, The Netherlands
e-mail: e.e.van.der.vaart@rug.nl

E. van der Vaart
Institute of Artificial Intelligence, University of Groningen, Groningen, The Netherlands

performance on specific tasks, it becomes much easier to look past one's own biases and to see what cognitive processes might actually be occurring.

Keywords Theory of mind · Animal cognition · Computational modeling

1 Introduction

Does the chimpanzee have a 'theory of mind'? Premack and Woodruff asked this question in 1978, and it is still provoking a torrent of response. The search for 'theory of mind', the ability to attribute mental states to others, has driven dozens of studies, on a variety of species (Emery 2005). Typically, scientists start by asking whether a specific skill, such as 'reasoning about what others can *see*', is present in their subjects, and then design an experimental setup where subjects should behave differently if they possess the skill than if they lack it (Povinelli and Vonk 2004). Thus, the explicit objective is to test for the presence of human-like social reasoning. At present, many such experiments have had positive results, with animals acting in ways consistent with 'theory of mind' (Emery and Clayton 2009). However, opinions differ on the interpretation of these findings. Some reviewers draw the conclusion that some animals[1], at least, understand some mental states (Lyons and Santos 2006; Emery and Clayton 2009; Premack 2007; Seed and Tomasello 2010; Byrne and Bates 2010); others find the evidence unconvincing (Penn and Povinelli 2007; Lurz 2011; Shettleworth 2010a; Bolhuis and Wynne 2009; Barrett 2010; Shettleworth 2010b; Penn et al. 2008; Lurz 2009).

Upon closer inspection, contemporary reviewers tend to motivate their skepticism with three different concerns. The first is that results that seem to indicate 'theory of mind' come from experiments that are of *the wrong kind*, that cannot differentiate between 'mind readers' and 'behavior readers' (Lurz 2011; Penn and Povinelli 2007). The second is that results come from experiments that are *insufficiently controlled*, in the sense that they do not explicitly test for alternative, 'less mentalistic' explanations (Bolhuis and Wynne 2009; Heyes 2012). The third is that certain results are seen as evidence of 'theory of mind' because *our own biases* favor that interpretation (Shettleworth 2010a; Barrett 2010; Bolhuis and Wynne 2009). None of these concerns are new (e.g., Kummer et al. 1990; Heyes 1993, 1998; Hemelrijk 1996). Therefore, we can ask: Why have they not been resolved yet? Or refuted? And what can animal cognition researchers do to move past them? That is the focus of this paper. To this end, we first provide a brief overview of the literature on 'theory of mind' in animals, then discuss each concern in turn. We conclude with an introduction to the methodological approach that we think offers the most promise for progress: Computational modeling.

2 Theory of mind in animals

Outside of humans, scientists have primarily looked for 'theory of mind' in apes and monkeys, as well as in corvids, a family of birds known for its cognitive prowess

---

[1] In this paper, we use the term 'animals' to refer to 'non-human animals'.

(Emery and Clayton 2004). In terms of the kinds of mental state that have been studied, most attention has been directed at *intending*, *seeing,* and *believing*. Although in the nineties findings were mostly negative (Tomasello et al. 2003a), recent results have convinced many that both chimpanzees and rhesus macaques understand the goals and perceptions of others, as well as what others know, but not what others falsely believe (Lyons and Santos 2006; Premack 2007; Byrne and Bates 2010; Seed and Tomasello 2010). Increasingly, the same holds for jays and ravens (Emery and Clayton 2009; Byrne and Bates 2010). A variety of experimental setups have contributed to this emerging consensus (Call and Tomasello 2008; Emery and Clayton 2009; Byrne and Bates 2010); in this section, we briefly introduce three influential examples. Later, we shall use these examples to illustrate key points of debate.

2.1 The 'begging paradigm'

The 'begging paradigm' (Povinelli and Eddy 1996) is an early setup designed to study chimpanzees' understanding of visual access. Every trial, subjects were given the opportunity to beg from a 'seeing' and a 'nonseeing' experimenter. When the 'seeing' experimenter was facing them, while the 'nonseeing' experimenter's back was turned, the chimpanzees immediately chose correctly. However, they did not differentiate between experimenters with buckets over their heads versus on their shoulders, or with blindfolds over their eyes versus over their mouths. Furthermore, once they had learned to beg correctly with respect to 'buckets' and 'blindfolds', the chimpanzees still failed to differentiate between experimenters with their eyes open versus their eyes closed (Povinelli and Eddy 1996). When the same subjects were retested a few years later, the results were similar (Reaux et al. 1999). In contrast, a different set of subjects immediately performed above chance on the first trial of all these conditions (Bulloch et al. 2008), ranging from 71 % correct when the choice was 'eyes open' versus 'eyes closed' to 100 % correct when the choice was 'facing forwards' versus 'facing backwards'. Other studies have also produced conflicting results (Hofstetter et al. 2007; Tempelmann et al. 2011; Kaminski et al. 2004), so that it remains unclear what affects the performance of chimpanzees in this setup.

2.2 The 'competitive paradigm'

The 'competitive paradigm' (Hare et al. 2000) was the first paradigm to produce positive results for visual perspective taking in chimpanzees. Here, two chimpanzees were presented with two pieces of food; one visible to both subjects, one visible only to the more subordinate of the two. This was due to one piece being out in the open, while the other was hidden behind a barrier, on the subordinate's side. When released with a small head start, the subordinate chimpanzee obtained more of the food that was hidden from the dominant (Hare et al. 2000). A subsequent study failed to replicate this result (Karin-D'Arcy and Povinelli 2002), but this finding was later explained as a consequence of placing the food too close to the subordinate, so that the competition from the dominant was reduced (Bräuer et al. 2007). In a later experiment, there were two barriers, with a piece of food behind only one, on the subordinate's side

(Hare et al. 2001). In this case, the subordinate approached the food more often if the dominant had not witnessed the baiting than if the dominant had. These results were interpreted as evidence that the subordinate understood what the dominant *saw* (Hare et al. 2000) and what it *knew* (Hare et al. 2001). However, the subordinate failed to approach more often if the food was moved after the dominant saw the baiting than if the food was not moved, suggesting that the subordinate did not understand the dominant's *false belief* about the food's location (Hare et al. 2001). Furthermore, if one piece of food was hidden on the subordinate's side of one barrier, while the dominant was watching, while another piece of food was hidden on the subordinate's side of a different barrier, while the subordinate was in private, then the subordinate chose to approach the two food pieces indiscriminately, indicating that it could not take into account which food piece the dominant knew about (Hare et al. 2001).

2.3 The 'caching paradigm'

The 'caching paradigm' is employed with Western scrub jays. This setup makes use of the birds' tendency to bury food items for future consumption, and the fact that conspecifics will steal such items if they see them being hidden. Emery and Clayton (2001) tested whether cachers take measures to prevent such theft. They gave their subjects ice cube trays to cache worms in, with a competitor in an adjacent cage. If the subjects could see this competitor, they re-cached their worms in new sites once they were alone, but only if they had previous experience being pilferers themselves; conversely, if their view of the competitor was blocked, they re-cached less frequently (Emery and Clayton 2001). Furthermore, when the birds cached in one tray in front of one competitor, and in another tray in front of another competitor, they later re-cached more of the worms cached in front of the competitor that was present at that time (Dally et al. 2006). Finally, if one of the two trays was less visible to the competitor—because it was behind a barrier, or further away, or in shadow—the birds not only cached more worms in this tray, they also re-cached less from it later (Dally et al. 2004, 2005).

3 Three types of concern

3.1 The wrong kind of experiment

One reason skeptics give to interpret the results outlined above as insufficient evidence of 'theory of mind' is that they come from 'the wrong kind of experiment' (Hurley and Nudds 2006; Lurz 2009; Penn and Povinelli 2007; Heyes 1993, 1998; Povinelli and Vonk 2003, 2004). According to this claim, all experimental setups reported in the literature are fundamentally incapable of establishing whether animals can reason about mental states, or only about behavior. In this view, the defining feature of 'theory of mind' is that it involves thinking about *unobservable inner experiences*, and not just states of the physical world. Thus, there is a crucial difference, on the one hand, between understanding that others experience the mental state of 'seeing' and, on the other hand, understanding that lines of sight are important. For instance, for a chimpanzee to succeed in the 'begging paradigm', it can think about which experimenter

can *see* it, or it can think about which experimenter is oriented towards it, without any obstructing barriers (Heyes 1998; Povinelli and Vonk 2004). Both strategies are sufficient to solve the task, and there is no way to tell them apart. For the 'competitive paradigm', the same reasoning holds (Povinelli and Vonk 2004; Lurz 2009; Karin-D'Arcy and Povinelli 2002). For a scrub jay to re-cache the worms most at risk of being stolen, it can remember what competitors *know* about its cache locations, or it can recall which competitors were visible when it was caching, and how far away those competitors were (Penn and Povinelli 2007). Thus, in these experiments, subjects can always succeed by reasoning from what they *themselves* have seen, to what they themselves should *do*, without considering the mental states of others.

3.2 Insufficient controls

A second reason to be skeptical of results that seem to demonstrate that animals can attribute mental states is that they may come from experiments that are 'insufficiently controlled' (Bolhuis and Wynne 2009; Shettleworth 2010a; Heyes 2012). Although comparative researchers studying 'theory of mind' often take care to exclude as many alternative hypotheses as possible, relevant control conditions may also be overlooked. For instance, with respect to Hare et al. (2000) 'competitive paradigm', Karin-D'Arcy and Povinelli (2002) posited that subordinates might have preferred the 'hidden food' simply because they preferred eating near barriers. Although their own follow-up experiments showed this hypothesis to be false—subordinate chimpanzees reacted differently to food pieces fully and partially concealed from a dominant, despite the fact that they could eat by a barrier in both cases (Karin-D'Arcy and Povinelli 2002)— it was not initially discussed as a possible alternative explanation. This indicates that many other results may *also* come from insufficiently controlled experiments, as skeptics claim (Bolhuis and Wynne 2009; Shettleworth 2010a; Heyes 2012).

3.3 Interfering intuitions

A third reason to question whether positive results on certain tasks actually imply 'theory of mind' is that our own intuitions might be biased in favor of that interpretation (Shettleworth 2010a; Barrett 2010; Bolhuis and Wynne 2009; Kummer et al. 1990; Hemelrijk 1996). Various reasons have been given for why such a bias might exist. One is that the popular press seems to prefer stories of how animals are 'unexpectedly human-like', creating an incentive for researchers to interpret their data that way (Shettleworth 2010a). Another is that finding 'theory of mind' in our closest relatives might be seen as a strengthening of evolutionary theory itself (Bolhuis and Wynne 2009; Shettleworth 2010a; Povinelli and Vonk 2003). To convincingly demonstrate that the cognitive complexity of humans could be the result of gradual natural selection, so the reasoning goes, it would be helpful to have a living illustration of every stage that the human mind might have gone through. Consciously or unconsciously, this idea could be motivating researchers to interpret their results more 'cognitively' than necessary, to 'close the gap' between humans and the rest of the animal kingdom. A final idea is that human 'theory of mind' is expressly designed, in an evolutionary sense, to infer men-

tal states —and mental states about mental states—from observable behavior (Barrett et al. 2007). Therefore, we automatically infer them from the observable behavior of our experimental subjects. Thus, our own 'folk psychology' could be hindering our ability to objectively evaluate results concerning 'theory of mind' in animals.

4 'The wrong kind of experiment': a closer look

In sum, unsuitable experimental setups, a lack of sufficient controls, and our own intuitive biases are all possible reasons to doubt positive results on tests for 'theory of mind'. However, these three arguments for skepticism have different implications. If all the experimental setups currently employed are fundamentally incapable of differentiating between 'mind readers' and 'behavior readers', then other concerns seem moot. Why care about the results, or the interpretations, of uninformative experiments?

4.1 The right kind of experiment

The only way to resolve the concern that all results come from 'the wrong kind of experiment' is to do 'the right kind of experiment'. As an example of what such an experiment could look like, Povinelli and colleagues have adapted an earlier suggestion by Heyes (1998) to create the 'opaque visor paradigm' (Povinelli and Vonk 2003, 2004; Penn and Povinelli 2007), a variation of the 'begging paradigm'. In this setup, chimpanzees are first exposed to two buckets, a yellow one and a blue one, which both contain a visor. From the outside, the visors look similar, but through personal experience with wearing the buckets, the chimpanzees will discover that the yellow bucket's visor blocks vision, while the blue bucket's visor does not. They will be given the opportunity to beg from two experimenters, one wearing a yellow bucket, the other wearing a blue one. If, from their own experience, they understand through which bucket one can *see*, they should gesture preferentially toward the experimenter wearing the blue bucket; conversely, if they have no concept of 'seeing', they should gesture indiscriminately—as two chimpanzees did in a pilot study by Vonk and Povinelli (2011). The argument is that this setup *can* distinguish between 'mind readers' and 'behavior readers' because it does *not* provide subjects with any behavioral cues that correspond with the experimenter's mental state. Thus, given that the chimpanzees have no experience with others wearing buckets, the claim is that there is no 'non-mentalistic' way for them to predict the experimenters' behavior.

However, others have claimed that the 'opaque visor paradigm' *can* be solved without reasoning about the mental states of others (Andrews 2005; Hurley and Nudds 2006; Lurz 2011, 2009). Once the experimenters are wearing their yellow and blue buckets, with their opaque and transparent visors, there are no 'observable signs' of 'seeing'. In fact, if the chimpanzees have never experienced what others are like with the buckets over their heads—as the protocol calls for— then they have had no opportunity to directly learn a rule like 'others only respond to events and things if they are not wearing yellow buckets'. However, the chimpanzees could have learned that their *own* behaviors were limited by wearing yellow buckets; 'with the yellow bucket on, I cannot do things'. From this, the chimpanzees could, in theory, reason that they want

the experimenter to do things (namely, pass them food) and that they should, therefore, beg from the experimenter wearing the bucket that allows for doing things (the blue one). This is complex, self-other mapping ('I cannot do things with the yellow bucket on, so she cannot either'), but it does not involve reasoning about 'unobservable inner experiences' in the strictest sense.

Lurz (2009, 2011), at the very least, finds this argument against Povinelli and colleagues' paradigm compelling. Even if chimpanzees *could* pass the 'opaque visor paradigm', that would still not prove that they are capable of reasoning about mental states. Therefore, Lurz argues for different, more complex experimental setups. His fundamental claim is that unequivocally demonstrating that any animal understands that others experience the mental state of *seeing* is impossible. The problem is that the mental state of *seeing* correlates too well with its observable indicators. A well-lit unobstructed line of sight to an object implies that one can *see* it; conversely, to *see* an object requires a well-lit unobstructed line of sight. There is no way to test whether a nonverbal animal can attribute the mental state of *seeing* without offering it the behavioral cues through which it can make the attribution, and once one does that, there is no way to exclude that the animal is *only* reasoning about the behavioral cues. According to Lurz, the only way to avoid this problem is by focusing on a slightly more complex mental state: That of *seeing-as*, rather than *seeing*. If chimpanzees understand that things can *look* different than they *are*, then it is possible to affect their mental state inferences without affecting anything observable at the same time.

Lurz (2009, 2011), Lurz and Krachun (2011) proposes several experimental paradigms that rely on this principle, but we use only one to illustrate the reasoning. It involves lenses that distort images, and it relies on the distinction between *appearance* and *reality*. A subordinate chimpanzee is first exposed to a version of the 'competitive paradigm', with the twist that it involves larger and smaller food pieces. This should teach the subordinate that the dominant usually approaches larger food pieces over smaller ones. Then, the subordinate chimpanzee is trained to retrieve objects from behind two kinds of barriers: A red-rimmed barrier that makes objects look bigger, and a blue-rimmed barrier that makes objects look smaller. Crucially, these objects should not be food pieces, and the chimpanzee should have no reason to prefer larger over smaller versions—so that there is no opportunity for the chimpanzee to associate the different barriers with different degrees of 'desirability'. Then comes the crucial test, where the red and blue-rimmed barriers are put to use in the 'competitive paradigm', and an equally-sized food piece is placed behind each, on the subordinate's side. If the subordinate chimpanzee understands the mental state of *seeing-as*, as well as how the barriers work, it should prefer to approach the food piece behind the blue-rimmed barrier, as this is the food piece that should *look* smaller to the dominant. Conversely, if it *does not* understand the mental state of *seeing-as*, then it should approach randomly.

4.2 No need for the right kind of experiment?

Although a version of 'the right kind of experiment' was first proposed over a decade ago (Heyes 1998), only one team of researchers have reported any results, involving just three juvenile chimpanzee subjects (Vonk and Povinelli 2011). As far as we are

aware, the assertion that existing experimental setups are incapable of differentiating between 'mind reading' and 'behavior reading' has never been directly countered. However, two related challenges are often made. Firstly, it is argued that 'behavior reading' is not a plausible alternative to 'mind reading' at all; secondly, it is claimed that the proposed 'right kind of experiment' is fundamentally unsuitable for chimpanzees.

The first challenge is that, given all the results already collected, any explanation other than 'theory of mind' has begun to seem 'unparsimonious' (Tomasello and Call 2006; Emery and Clayton 2008). To maintain that chimpanzees and scrub jays cannot reason about mental states, the argument goes, one needs to assume that they are using many different behavioral rules instead. 'Beg from experimenter whose face is visible'; 'Go to food that a dominant has not oriented towards'; 'Re-cache items cached close to a conspecific'. While it cannot be *excluded* that subjects possess such behavioral rules —either because they are genetically equipped with them, or because they learn them in some unspecified way—it does not seem *likely*, given the sheer number required, and the lack of clarity with regards to where they come from. Therefore, the claim is that 'theory of mind' is the 'simpler' explanation in the sense of requiring fewer behavioral rules, and less complicated prior learning (Tomasello and Call 2006; Emery and Clayton 2008).

The second challenge is that 'the right kind of experiment' suffers from 'low ecological validity' (Hare 2001; Emery and Clayton 2008; Tomasello et al. 2003b; Lyons and Santos 2006). The idea is that whatever 'theory of mind' a species is capable of, it presumably evolved in the context of solving its every day problems. Thus, if we want to investigate the limits of its cognitive abilities, we should present it with tasks that are as similar to its daily life as possible. In this view, it is no surprise that the 'competitive paradigm' and the 'caching paradigm' have proven successful at extracting 'theory-of mind-like' behavior from chimpanzees and scrub jays, respectively; competing for food and preventing cache theft are problems with 'high ecological validity', but the 'begging paradigm' and the 'opaque visor paradigm' have 'low ecological validity'. Cooperation and food sharing are already relatively alien notions to a chimpanzee, and adding unnatural objects such as buckets and visors only makes the task more difficult (Hare 2001).

4.3 Concern refuted?

However, we find it difficult to dismiss the concern that all existing experiments are of 'the wrong kind', on the grounds of either 'parsimony' or 'low ecological validity'. First, it is not clear that 'mind reading' is actually a 'simpler' explanation than 'behavior reading', as many authors have argued (for an overview, see Heyes 2012). It is certainly *more complex* in the sense of being cognitively more complex, and it is not necessarily less complex in the sense of requiring less behavioral rules or less learning (Povinelli and Vonk 2004). Even if chimpanzees do not have 'theory of mind', that does not mean that every specific behavioral pattern that chimpanzees exhibit must be explained by a different associative rule (Penn and Povinelli 2007). Instead, chimpanzees, and indeed scrub jays, might know that 'others only respond

to events and things that they had unobstructed lines of sight to'; a single behavioral rule that guarantees success on virtually every experiment designed to measure understanding of what others *see*. Furthermore, understanding that others *see* is not enough to recognize *what* others *see when*; one still needs to group observed behaviors into categories, and to learn which categories of behavior correspond to which mental state. Therefore, 'mind reading' requires all the learning that 'behavior reading' does (Lurz 2011; Povinelli and Vonk 2004).

Second, 'low ecological validity' is not a counterargument to the claim that all results come from 'the wrong kind of experiment'. It is a reason to object to the *specifics* of the 'opaque visor paradigm', but versions of it with higher ecological validity can be constructed. Instead of requiring subjects to beg for a piece of food from the 'seeing' experimenter, one could give them the opportunity to steal a piece of food from the 'non-seeing' experimenter (Flombaum and Santos 2005). This makes the task competitive, but preserves its structure. More generally, it has been argued that 'low ecological validity' is a strange objection for three reasons. Firstly, the higher a task's ecological validity, the greater the odds that natural selection has equipped subjects with a built-in response, rendering mental state attribution unnecessary (Povinelli and Vonk 2004; Vonk and Shackelford 2012). Secondly, it is not clear that chimpanzees *are* systematically worse at cooperative tasks (Povinelli and Vonk 2004; Penn and Povinelli 2007), and furthermore, although the use of strange objects is common in experiments on tool use, in that context similar complaints are absent (Penn and Povinelli 2007; Vonk and Shackelford 2012). Nevertheless, Penn and Povinelli (2007) also propose an alternative to the 'opaque visor experiment' with 'high ecological validity'. The suggested setup is presented as a 'systematic version' of the 'competitive paradigm', with five barriers and two food pieces, a smaller and a larger. Then, there are many different ways to manipulate which chimpanzee knows what; allow the dominant to watch as one food piece is placed, but not the other, allow the dominant to watch as both pieces are placed, but then swap them while it cannot see, and so on. According to Penn and Povinelli, this makes the task computationally intractable to solve without 'theory of mind'. However, as Lurz (2009) points out, 'computationally intractable' is not the same as 'impossible'; chimpanzees *could* succeed in this setup by avoiding food pieces still located in places to which dominants have had direct lines of sight. Thus, it is the 'wrong kind of experiment' in the sense of 'possible for subjects to solve without the attribution of unobservable inner experiences'.

4.4 How to move forward

If we accept that all existing setups are unable to differentiate between 'mind readers' and 'behavior readers', the question is whether this concern can be resolved through the more complicated setups suggested by skeptics. We do not think so. All the setups presently proposed—'the opaque visor paradigm' (Povinelli and Vonk 2003, 2004; Vonk and Povinelli 2011), 'the systematic competitive paradigm' (Penn and Povinelli 2007), the 'seeing-as competitive paradigm' (Lurz 2009)—build on existing setups that we already know chimpanzees have difficulty with. The 'opaque visor paradigm', for instance, asks subjects to choose between two exper-

imenters wearing buckets with visors, of which one allows seeing and the other does not. However, chimpanzees in the earlier 'begging paradigm' do not even consistently prefer experimenters with buckets over their heads to experimenters without buckets over their heads (Povinelli and Eddy 1996; Reaux et al. 1999; Bulloch et al. 2008). This makes it improbable that they will choose correctly when faced with experimenters wearing buckets equipped with visors, and in fact, preliminary evidence suggests that they do not (Vonk and Povinelli 2011). Similarly, the 'systematic competitive paradigm' requires chimpanzees to flexibly respond to a dominant's beliefs about the locations of two different food pieces. However, subordinates already behave incorrectly in simpler versions of the setup (Hare et al. 2001); they fail to approach more often when the dominant has a false belief about the location of food, and if they must choose between a food piece which the dominant has seen and one that it has not seen, they approach indiscriminately. Although this last result has been ascribed to motivational, rather than cognitive issues (Hare et al. 2001), and chimpanzees have actually passed a similar test in a slightly different setup (Kaminski et al. 2008), the fact remains that more challenging versions of the 'competitive paradigm' are unlikely to be successful. The 'seeing-as competitive paradigm' suffers from the same problem, in addition to requiring that subjects understand the workings of size-altering barriers. Given that Krachun et al. (2009) found evidence of such understanding in only 4 out of 11 chimpanzees, the odds of successfully running the 'seeing-as paradigm' with any animal are small. In fact, the best use of this type of novel setup may be with human children. A variation on the 'opaque visor paradigm' conducted with toddlers has already led to new insights with respect to how 'theory of mind' develops (Teufel, Clayton & Russell, in press).

Furthermore, although we have been referring to these experimental setups as shedding light on 'whether animals have a 'theory of mind' or not', this is actually an overstatement: Success might be sufficient to prove some understanding of mental states, but failure tells us very little, as Penn and Povinelli (2007) and Lurz (2009) all acknowledge. For instance, with respect to the 'opaque visor paradigm', chimpanzees could be perfectly capable of thinking about 'seeing' without being able to map their own experience with the buckets to that of someone else; they might forget which bucket allows seeing and which does not; they might assume that humans can see through any type of visor, and so on. The proposed paradigm is so intricate that failure can be ascribed to any number of causes. For the 'seeing-as competitive paradigm', there is the further complication that it tests for a more advanced mental state than is necessary. If chimpanzees understand true beliefs, but not false ones (Byrne and Bates 2010; Premack 2007; Seed and Tomasello 2010), they might understand accurate perceptions, but not inaccurate ones. Thus, it seems to us that the existing evidence suggests that chimpanzees are unlikely to consistently pass any of the paradigms proposed as 'the right kind of experiment', and furthermore, that such lack of success would not shed much light on their ability to reason about mental states. Therefore, it seems that there is no way to resolve the concern that all results come from 'the wrong kind of experiment'. Instead, we think that the most promising way forward is simply to ignore it.

5 Beyond 'the wrong kind of experiment'

We have argued that there is little progress to be made by addressing the concern that existing experimental setups cannot conclusively establish whether animals are reasoning about mental states. Instead, the more answerable question is how *sophisticated* their reasoning is, whether it is about 'unobservable inner experiences' or not. To establish that, however, is still difficult. A lack of control conditions, or a bias to see 'theory-of-mind-like' behavior where it is not present, would still be valid concerns. In this section, we tackle both issues.

5.1 Controls and biases

Whether, on the whole, experiments attempting to establish 'theory of mind' in animals are less well-controlled than is possible is a difficult question. In Heyes' (1998) paper, she offered a series of critiques of specific experiments, but since then, few concrete control conditions have been explicitly identified as being missing [one exception is Karin-D'Arcy and Povinelli (2002) analysis of the 'competitive paradigm'; another is Heyes' (2012) discussion of an experiment on prosociality in chimpanzees (Horner et al. 2011)]. As a consequence, this is a rather difficult concern to refute; conversely, it is also a rather difficult concern to substantiate.

Whether researchers might intuitively favor 'theory-of-mind-like' explanations is a different matter. Here, we feel that the evidence certainly suggests such a bias might exist; we discuss a number of examples from the research paradigms discussed earlier. One indication that evidence for 'theory of mind' is implicitly the desired outcome of some experiments is that there is a tendency to split subjects and behaviors into smaller and smaller subsets, until a subset is found that is consistent with 'mental state attribution'. For instance, in one version of the 'begging paradigm', five different begging behaviors were measured across three different species (Kaminski et al. 2004). None of these begging behaviors were consistently directed more at an experimenter with her eyes open than at one with her eyes closed. However, it was noted that one orangutan lip-begged longer in the former condition than in the latter, and that another spat more in the latter condition than in the former, and this was considered to be evidence of some understanding of the importance of the eyes in relation to visual attention.

Similarly, when a single experimental paradigm is sufficiently rich that it generates many different results, there is sometimes a tendency to overemphasize results that are consistent with 'theory of mind' over results that are less consistent. Take, for instance, a recent study employing the 'begging paradigm' (Tempelmann et al. 2011). In this study, the objective was to clarify an earlier finding by Kaminski et al. (2004), where different species of ape begged more if an experimenter's face was oriented towards them than if it was oriented away, but only if her body was oriented towards them, too. If the experimenter's body was turned backwards, then the apes did not differentiate between trials where the experimenter was also looking backwards, and trials where she was looking over her shoulder, in the direction of the subjects. In Tempelmann et al. (2011) follow-up experiment, the effect of body orientation disappeared for bonobos,

orangutans and gorillas, but not for chimpanzees. Furthermore, it was found that none of the apes made more auditory requests for food when the experimenter's back was turned, suggesting that they did not understand that noise can alert individuals who are visually inattentive. However, neither of these results made it into the final conclusion of the paper, or its abstract, which stated only that the study had shown that all great ape species judge a human's attentional state on the basis of the face, thus supporting the 'theory of mind' hypothesis.

As another example, in the 'caching paradigm', it has repeatedly been shown that when scrub jays are forced to cache in front of a competitor, they re-cache more often afterwards (Emery and Clayton 2001; Emery et al. 2004; Dally et al. 2005, 2006). That is, they dig up their worms and re-bury them in new locations. Furthermore, if they are given two trays to cache in, of which one is better visible to a competitor than the other, they later re-cache more from the tray that was better visible (Emery and Clayton 2001; Emery et al. 2004; Dally et al. 2005, 2006). A common interpretation of this behavior is that scrub jays somehow appreciate which caches are most likely to be pilfered, and then protect those specific caches by moving them to new sites (Clayton et al. 2007). However, what matters in terms of preventing theft is that fewer of the 'high risk' worms are left in their old locations at the end of the trial, and the birds can accomplish this by either re-caching *or* eating them, both of which count as 'recovering'. Thus, it seems like this interpretation predicts that scrub jays should not just re-cache, but also *eat*, more of the caches best seen by other birds. However, this prediction seems to be met only rarely (Emery and Clayton 2001; Emery et al. 2004; Dally et al. 2005, 2006). Taking only statistically significant results, scrub jays re-cached more of the 'high risk' worms in thirteen out of fourteen opportunities reported, but only recovered more of them in two out of nine cases. Yet, the implications of this lack of agreement with the 'theory of mind' hypothesis are rarely considered (van der Vaart et al. 2012).

A final indication of an implicit bias in favor of 'theory of mind' is the fact that occasionally, two conflicting experimental outcomes are both considered evidence of mental state attribution. For instance, for Western scrub jays, it has always been considered a hallmark of their sophisticated social cognition that they only employ their cache protection tactics when they are actively watched; the mere *presence* of another bird, without visual access to the scene, is insufficient (Dally et al. 2005). Conversely, for Clark's nutcrackers, another species of corvid tested in a similar caching paradigm, it was found that they reduced their caching across trials whether a competitor was visible or not; whether a competitor could be seen or just heard made no difference (Clary and Kelly 2011). This result was subsequently interpreted as evidence that the Clark's nutcrackers were generalizing their learning from the 'watched condition' to all conditions where a conspecific was known to be present (Clary and Kelly 2011). While this is certainly a plausible explanation, the implicit message is that this generalization is cognitively impressive, as it was something the Clark's nutcrackers 'were able to do'. Thus, in this case, diametrically opposed results (different behavior because the conspecific is behind a barrier versus similar behavior despite the conspecific being behind a barrier) are both interpreted as evidence in favor of the same hypothesis—that the cache protection strategies of corvids are evidence of complex cognition.

5.2 Designing and interpreting experiments better

Several authors have proposed strategies for improving our ability to design control conditions, and for reducing the effects of intuitive biases. Heyes (2012), for instance, argues that part of the problem is that associative learning—the 'alternative theory' to 'more cognitive' explanations, in most cases—is a challenging subject to get into. Its long history and technical vocabulary make it difficult for nonspecialists to absorb. Thinking in terms of what an animal understands is easier than thinking in terms of what an animal's earlier experiences might have taught it through reinforcement. As a consequence, it is difficult for researchers to conceive of the correct control conditions for their experiments. A solution, according to Heyes, is for everyone to become more familiar with the tenets of associative learning, and she suggests a number of resources for doing so (Pearce 2008; Dickinson 1980).

A second strategy is offered by Shettleworth (2010a). Its purpose is to alleviate the effects of an intuitive bias in favor of 'theory of mind'; a similar sentiment is put forth by Barrett (2010). According to these authors, our 'common sense' view of *human* 'theory of mind' is probably wrong. We think of ourselves as making conscious, deliberate social judgements—and we *can* do that—but most of our interactions we manage without them. Shettleworth, for instance, discusses how human mate choice is affected by simple cues like shoulder width and waist-to-hip ratio (Singh et al. 2010), while Barrett mentions how we predict what others will do based on the personality traits we assign to them, rather than on the mental states we think they have (Andrews 2005, 2008). The human ability to function without 'theory of mind' is also illustrated by studies on infants; already at fifteen months, long before they pass verbal 'false belief' tasks, their looking times show that they expect others to search for objects where they have last seen them, rather than where they currently are (Onishi and Baillargeon 2005). Conversely, even as adults, we do not necessarily use our 'theory of mind' very *well*; when subjects are asked to hand over 'the largest vase', they often reach for the largest vase *they* can see, rather than the largest vase the *asker* can see (Keysar et al. 2000). Clearly, then, even very young humans make smart social inferences without truly understanding others' minds, while even very competent adults do not always think about the mental states of others when completing tasks. Such findings have led to the theory that adult humans may actually possess two systems for predicting and explaining the actions of others (Apperly and Butterfill 2009); one fast but inflexible, based on behavior reading, and one slow but able to handle great complexity, based on explicit mental state ascription. The claim is further that the system for behavior reading is widely shared, with infants and other animals, while the one for mental state ascription is only fully developed in human adults (Apperly and Butterfill 2009). If this perspective was more fully appreciated by animal cognition researchers, so argue Shettleworth (2010a) and Barrett (2010), perhaps we would be less likely to see 'theory of mind' in every smart-looking social behavior exhibited by subjects.

A third strategy, also offered in the spirit of alleviating the risk of interpreting results in an unnecessarily complex manner (Barrett et al. 2007; Barrett 2010), is the one developed by the 'new artificial intelligence' movement (Pfeifer and Scheier 1999) and by researchers of self-organisation (Hemelrijk 2002; Camazine et al. 2001; Sueur and Deneubourg 2011; Couzin and Krause 2003). The basic claim underlying

this 'embodied and embedded' approach is that by focusing on 'cognition' in isolation, we are essentially ignoring many other 'non-cognitive' factors that contribute to how individuals behave. Animals, like humans, have brains that exist in bodies, and bodies that exist in the world, and constraints *outside* of the brain, and *outside* of the body, may naturally predispose individuals to produce behavior that looks smart without requiring complex cognition to accomplish it. The argument is that by taking this perspective, our tendency towards anthropocentric interpretations is automatically lessened, as it forces us to think about how animals actually perceive the world, and what options they have to physically act in it.

6 A helpful tool: computational modeling

One methodological tool that can assist with thinking in an 'embodied and embedded' way is computational modeling. In ethology, most existing computational models are known as *agent-* or *individual-based models* (e.g., Evers et al. 2011; Hemelrijk and Hildenbrandt 2011; King et al. 2011). This type of model simulates every individual separately, with its own characteristics and decision rules. It mostly generates new insights by demonstrating that interactions between individuals, or between individuals and their environment, can generate unexpected patterns across either time or space (Pfeifer and Scheier 1999). As a consequence, they are a powerful tool for generating alternative hypotheses to 'theory of mind' (Hemelrijk 1996; Hemelrijk and Bolhuis 2011); they make it possible to test whether simple assumptions about animals' motivations might lead to self-organised patterns that look like the product of complex cognition. For instance, we have used an agent-based model to show that the 'reconciliation' behavior of primates is not necessarily the product of sophisticated social reasoning (Puga-Gonzalez et al. 2009). In macaques, two former opponents are more likely to groom immediately after fighting than at other times. This has been taken as evidence that macaques understand relationships, and that they are selectively approaching each other in an effort to 'reconcile'. However, in our simulations, the same patterns arise as a consequence of simple rules about fighting and grooming, and their effects on spatial proximity.

In addition to agent-based models, a related kind of model, the *cognitive model* (Sun 2008), is just starting to be used by researchers working in animal cognition. This modeling approach was first designed to study humans (Taatgen and Anderson 2010), and has been applied to a wide variety of psychological questions, ranging from the very fundamental, such as 'how is information retrieved from memory' (van Maanen et al. 2012) to the very practical, such as 'how do car drivers switch lanes' (Salvucci 2006). Like agent-based models, cognitive models simulate each individual separately, with their own characteristics and decision rules. However, typically, the decision rules of a cognitive model involve more detailed representations of the underlying mechanisms, such as those for memory and learning. Thus, compared to agent-based models, the focus is more on how the interactions between different cognitive processes may explain specific patterns of behavior. Therefore, this type of model is exceptionally suited to answering questions in animal cognition (Penn et al. 2008). So far, it is has rarely been used in this way, at least in the context of research

explicitly aimed at studying 'human-like' cognitive capacities; we know of only a few studies. Bryson and Leong (2007) used it to analyze transitive inference in primates, while Harrison and Trafton (2009) used it to investigate the effects of learning in the 'competitive paradigm' employed with chimpanzees.

As another example, we have used a cognitive model (van der Vaart et al. 2011, 2012) to investigate the re-caching behavior of scrub jays (Emery and Clayton 2001; Dally et al. 2006). The model consists of a kind of 'virtual bird', with a memory system based on ACT-R, a cognitive model designed to study humans (Anderson 2007). Its caching and recovery behavior has been validated against that of real birds (van der Vaart et al. 2012); its recovery errors resemble those made by Clark's nutcrackers (Balda et al. 1986; Kamil and Balda 1990), and its choice of cache is similar to that of Western scrub jays (de Kort et al. 2007). In further work, the 'virtual bird' was extended with assumptions related to re-caching (van der Vaart et al. 2012). The 'virtual bird' wanted to cache more when it was stressed, and it was stressed by the presence of conspecifics, as well as by finding caches missing. This meant that it re-cached more *while* it was watched, as real scrub jays do (Dally et al. 2005); namely, it kept burying and re-burying the few worms that were available. As an emergent pattern, this confused the 'virtual bird's' memory, which caused it to experience more failures at recovery. These failures again caused it stress, resulting in an increased desire to cache and re-cache. As a consequence, it re-cached more *after* being watched (Emery and Clayton 2001). The same process also produced more re-caching from trays most visible to competitors (Dally et al. 2006). As the 'virtual bird' cached less in such trays, they were emptier at recovery, and the odds of unsuccesfully recovering there were higher. As this is what caused the 'virtual bird' stress, this caused it to re-cache more from trays best visible to conspecifics.

Importantly, this theory of why scrub jays re-cache the way they do is completely novel, and conceived entirely thanks to our use of a model. When we first set out to simulate the social cognition of scrub jays, there was no indication that stress or memory errors might be relevant; instead, these ideas developed gradually, from an iterative cycle of reviewing the literature and exploring the 'virtual bird'. Initially, the model focused on an experiment where scrub jays could actually see a competitor steal their worms (Emery et al. 2004), and the model's central assumption was that re-caching was caused by the bird's memories of watching such pilferings. However, it became apparent that if re-caching was caused by the bird's memories of its own failed recovery attempts, then this would result in the same observable behavior; after all, the higher the number of pilferings, the higher the number of failed recovery attempts. This led to the realization that *any* increase in failed recovery attempts could potentially drive enhanced re-caching. This prompted a return to the literature, where we found evidence that scrub jays re-cache more while they are watched (Dally et al. 2005) and that many birds cache more in stressful circumstances (Wein and Stephens 2011; Pravosudov 2003; Lucas et al. 2006). Putting all this together, we developed a new, less anthropomorphic explanation for why scrub jays re-cache the way they do.

Furthermore, the fact that a computational model must actually be implemented ensures automatically its specificity, and thus, its ability to generate exact empirical predictions. For instance, our explanation predicts that, during recovery, scrub jays start re-caching *after* they start making memory errors. In sum, this type of modeling

should help to design control experiments, as well as to reduce the tendency to see 'theory of mind' where it might not be present.

7 Conclusions

In this paper, we reviewed three concerns often cited by those who remain unconvinced that nonhuman species can attribute mental states. These skeptics argue that the evidence consistent with animals having 'theory of mind' comes either from experiments that are of the wrong kind, from experiments that are insufficiently controlled, or from experiments that are interpreted in a biased fashion. We concluded that, firstly, it seems *irrefutable* that existing experiments cannot conclusively distinguish mind-readers from behavior-readers. However, it seems like any experiments that *would* be able to do so are so complex that no subject will pass them; therefore, it seems like this question is best left alone. Irrespective of whether animals are reasoning about 'unobservable inner experiences' or 'states of the physical world', a more interesting question is how *sophisticated* their reasoning really is. Secondly, we argued that while it is difficult to determine whether experiments are insufficiently controlled, one way forward is to make sure that researchers of 'theory of mind' are also intimately familiar with associative learning theory, so that alternative explanations may at least be thought of. Thirdly, we argued that a preference for interpreting results in a more 'theory-of-mind-like' fashion does seem to exist, and that the literature suggests a number of conceptual frameworks for avoiding such biases; namely, a better knowledge of the limited 'theory of mind' humans actually use, and thinking of animal cognition in an embodied, embedded way. We concluded by pointing out that computational models are a powerful tool for facilitating such embodied, embedded thinking, and that agent-based models and particularly cognitive models may make important contributions. By explicitly simulating the rules and representations that underlie animal performance on specific tasks, it becomes much easier to look past one's own biases and to see what animals might actually be doing.

References

Anderson, J.R. (2007). *How can the human mind occur in the physical universe?* Oxford: Oxford University Press.

Andrews, K. (2005). Chimpanzee theory of mind: Looking in all the wrong places? *Mind and Language, 20*, 521–536.

Andrews, K. (2008). It's in your nature: A pluralistic folk psychology. *Synthese, 165*, 13–29.

Apperly, I.A., & Butterfill, S.A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*, 953–970.

Balda, R.P., Kamil, A.C., & Grim, K. (1986). Revisits to emptied cache sites by Clark's nutcrackers (Nucifraga columbiana). *Animal Behaviour, 34*, 1289–1298.

Barrett, L. (2010). Too much monkey business. In G. Semin & G. Echterhoff (Eds.), *Grounding sociality*. (pp. 219–236). London: Psychology Press.

Barrett, L., Henzi, P., & Rendall, D. (2007). Social brains, simple minds: Does social complexity really require cognitive complexity? *Philosophical Transactions of the Royal Society B, 362*, 561–575.

Bolhuis, J.J., & Wynne, C.D.L. (2009). Can evolution explain how minds work? *Nature, 458*, 832–833.

Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition, 10*, 439–448.

Bryson, J.J., & Leong, J.C.S. (2007). Primate errors in transitive 'inference': A two-tier learning model. *Animal Cognition, 10*, 1–15.

Bulloch, M.J., Boysen, S.T., & Furlong, E.E. (2008). Visual attention and its relation to knowledge states in chimpanzees, Pan troglodytes. *Animal Behaviour, 76*, 1147–1155.

Byrne, R.W., & Bates, L.A. (2010). Primate social cognition: Uniquely primate, uniquely social, or just unique? *Neuron, 65*, 815–830.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*, 187–192.

Camazine, S., Deneubourg, J.-L., Franks, N. R., Sneyd, J., Theraulaz, G., & Bonabeau, E. (2001). *Self-organization in biological systems*. Princeton: Princeton University Press.

Clary, D., & Kelly, D.M. (2011). Cache protection strategies of a non-social food-caching corvid, Clark's nutcracker (Nucifrage columbiana). *Animal Cognition, 14*, 735–744.

Clayton, N.S., Dally, J.M., & Emery, N.J. (2007). Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist. *Philosophical Transactions of the Royal Society B, 362*, 505–522.

Couzin, I.D., & Krause, J. (2003). Self-organization and collective behavior in vertebrates. *Advances in the Study of Behavior, 32*, 1–75.

Dally, J.M., Emery, N.J., & Clayton, N.S. (2004). Cache protection strategies by western scrub-jays (Aphelocoma californica): Hiding food in the shade. *Proceedings of the Royal Society B, 271*, 5387–5390.

Dally, J.M., Emery, N.J., & Clayton, N.S. (2005). Cache protection strategies by western scrub-jays, Aphelocoma californica: Implications for social cognition. *Animal Behaviour, 70*, 1251–1263.

Dally, J.M., Emery, N.J., & Clayton, N.S. (2006). Food-caching western scrub jays keep track of who was watching when. *Science, 312*, 1662–1666.

de Kort, S.R., Correia, S.P.C., Alexis, D.M., Dickinson, A., & Clayton, N.S. (2007). The control of food-caching behavior by western scrub-jays (Aphelocoma californica). *Journal of Experimental Psychology: Animal Behavior Processes, 33*, 361–370.

Dickinson, A. (1980). *Contemporary animal learning theory*. Cambridge: Cambridge University Press.

Emery, N.J. (2005). The evolution of social cognition. In A. Easton & N.J. Emery (Eds.), *Cognitive neuroscience of social behaviour*. New York: Psychology Press.

Emery, N.J., & Clayton, N.S. (2001). Effects of experience and social context on prospective caching strategies by scrub jays. *Nature, 414*, 443–446.

Emery, N.J., & Clayton, N.S. (2004). The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science, 306*, 1903–1907.

Emery, N.J., & Clayton, N.S. (2008). How to build a scrub jay that reads minds. In S. Itakura & K. Fujita (Eds.), *Origins of the social mind: Evolutionary and developmental perspectives*. (pp. 65–98). Tokyo: Springer.

Emery, N.J., & Clayton, N.S. (2009). Comparative social cognition. *Annual Review of Psychology, 60*, 87–113.

Emery, N.J., Dally, J.M., & Clayton, N.S. (2004). Western scrub-jays (Aphelocoma californica) use cognitive strategies to protect their caches from thieving conspecifics. *Animal Cognition, 7*, 37–43.

Evers, E., de Vries, H., Spruijt, B.M., & Sterck, E.H.M. (2011). Better safe than sorry - Socio-spatial group structure emerges from individual variation in fleeing, avoidance or velocity in an agent-based model. *PLoS ONE, 6*, e26189. doi:26110.21371/journal.pone.0026189.

Flombaum, J.I., & Santos, L.R. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology, 15*, 447–452.

Hare, B. (2001). Can competitive paradigms increase the validity of experiments on primate social cognition?. *Animal Cognition, 4*, 269–280.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour, 59*, 771–785.

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour, 61*, 139–151.

Harrison, A., & Trafton, J.G. (2009). Gaze-following and awareness of visual perspective in chimpanzees. In: Howes A., Peebles D., Cooper R. (Eds.), *Proceedings of the Ninth International Conference on Cognitive Modeling*, (pp. 292–297), Manchester.

Hemelrijk C. K. (1996) Reciprocation in great apes: From complex cognition to self-structuring. In: W.C. McGrew, M. L.F., & T. Nishida (Eds.), Great Ape societies. Cambridge: Cambridge University Press.

Hemelrijk, C.K. (2002). Understanding of social behaviour with the help of complexity science. *Ethology, 108*, 655–671.

Hemelrijk, C.K., & Bolhuis, J.J. (2011). A minimalist approach to comparative psychology. *Trends in Cognitive Science, 15*, 185–186.

Hemelrijk, C.K., & Hildenbrandt, H. (2011). Some causes of the variable shape of flocks of birds. *PLoS ONE, 6*, e22479. doi:22410.21371/journal.pone.0022479.

Heyes, C.M. (1993). Anecdotes, training, trapping and triangulating: Do animals attribute mental states?. *Animal Behaviour, 46*, 177–188.

Heyes, C.M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences, 21*, 101–148.

Heyes, C.M. (2012). A qualified defense of associative learning. *Philosophical Transactions of the Royal Society B*, 367, 2695–2703.

Hofstetter, A.B., Russell, J.L., Freeman, H., & Hopkins, W.D. (2007). Now you see me, now you don't: Evidence that chimpanzees understand the role of the eyes in attention. *Animal Cognition, 10*, 55–62.

Horner, V., Carter, J.D., Suchak, M., & de Waal, F.B.M. (2011). Spontaneous prosocial choice by chimpanzees. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 13867–13851.

Hurley, S., & Nudds, M. (2006). The question of animal rationality. In S. Hurley & M. Nudds (Eds.), *Rational animals?* (pp. 1–83). Oxford: Oxford University Press.

Kamil, A.C., & Balda, R.P. (1990). Differential memory for different cache sites by Clark's nutcrackers (Nucifraga columbiana). *Journal of Experimental Psychology: Animal Behavior Processes, 16*, 162–168.

Kaminski, J., Call, J., & Tomasello, M. (2004). Body orientation and face orientation: Two factors controlling apes' begging behavior from humans. *Animal Cognition, 7*, 216–223.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition, 109*, 224–234.

Karin-D'Arcy, M.R., & Povinelli, D.J. (2002). Do chimpanzees know what each other see? A closer look. *International Journal of Comparative Psychology, 110*, 21–54.

Keysar, B., Barr, D.J., Balin, J.A., & Brauner, J.S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Pyschological Sciences, 11*, 32–38.

King, A.J., Sueur, C., Huchard, E., & Cowlishaw, G. (2011). A rule-of-thumb based on social affiliation explains collective movement in desert baboons. *Animal Behaviour, 82*, 1337–1345.

Krachun, C., Call, J., & Tomasello, M. (2009). Can chimpanzees (Pan troglodytes) discriminate appearance from reality? *Cognition, 112*, 435–450.

Kummer, H., Dasser, V., & Hoyningen-Huene, P. (1990). Exploring primate social cognition: Some critical remarks. *Behaviour, 112*, 84–98.

Lucas, J.R., Freeberg, T.M., Egbert, J., & Schwabl, H. (2006). Fecal corticosterone, body mass, and caching rates of Carolina chickadees (Poecili carolinensis) from disturbed and undisturbed sites. *Hormones and Behavior, 49*, 634–643.

Lurz, R. (2009). If chimpanzees are mindreaders, could behavioral science tell? Towards a solution of the logical problem. *Philosophical Psychology, 22*, 305–328.

Lurz, R. (2011). Belief attribution in animals: On how to move forward conceptually and empirically. *Review of Philosophy and Psychology, 22*, 19–59.

Lurz, R., & Krachun, C. (2011). How could we know whether nonhuman primates understand other's internal goals and intentions? Solving Povinelli's problem. *Review of Philosophy and Psychology, 2*, 449–481.

Lyons, D.E., & Santos, L.R. (2006). Ecology, domain specificity, and the origins of theory of mind: Is competition the catalyst? *Philosophy Compass, 1*, 481–492.

Onishi, K.H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*, 255–258.

Pearce, J.M. (2008). *Animal learning and cognition: An introduction*. London: Psychology Press.

Penn, D.C., Holyoak, K.J., & Povinelli, D.J. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds. *Behavioral and Brain Sciences, 31*, 109–178.

Penn, D.C., & Povinelli, D.J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philosophical Transactions of the Royal Society B, 362*, 731–744.

Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. Cambridge: MIT Press.

Povinelli D.J., & Eddy T.J. (1996) What young chimpanzees know about seeing (Vol. 61, No. 3, Monographs of the Society for Research in Child Development). Chicago: University of Chicago Press.

Povinelli, D.J., & Vonk, J. (2003). Chimpanzee minds: Suspiciously human? *Trends in Cognitive Sciences, 7*, 157–160.

Povinelli, D.J., & Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind. *Mind and Language, 19*, 1–28.

Pravosudov, V.V. (2003). Long-term moderate elevation of corticosterone facilitates avian food-caching behaviour and enhances spatial memory. *Proceedings of the Royal Society B, 270*, 2599–2604.

Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *Proceedings of the National Academy of Sciences of the USA, 104*, 13861–13867.

Puga-Gonzalez, I., Hildenbrandt, H., & Hemelrijk, C.K. (2009). Emergent patterns of social affiliation in primates: A model. *PLoS Computational Biology, 5*, e1000630. doi:1000610.1001371/journal.pcbi.1000630.

Reaux, J.E., Theall, L.A., & Povinelli, D.J. (1999). A longitudinal investigation of chimpanzees' understanding of visual perception. *Child Development, 70*, 275–290.

Salvucci, D.D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors, 48*, 362–380.

Seed, A.M., & Tomasello, M. (2010). Primate cognition. *Topics in Cognitive Science, 2*, 407–419.

Shettleworth, S.J. (2010). Clever animals and killjoy explanations in comparative pyschology. *Trends in Cognitive Sciences, 14*, 477–481.

Shettleworth, S.J. (2010). *Cognition, evolution and behavior*. New York: Oxford University Press.

Singh, D., Dixson, B.J., Jessop, T.S., Morgan, B., & Dixson, A.F. (2010). Cross-cultural consensus for waist-hip ratio and women's attractiveness. *Evolution and Human Behavior, 31*, 176–181.

Sueur, C., & Deneubourg, J.-L. (2011). Self-organization in primates: Understanding the rules underlying collective movements. *International Journal of Primatology, 32*, 1413–1432.

Sun, R. (Ed.). (2008). *The Cambridge handbook of computational psychology*. Cambridge: Cambridge University Press.

Taatgen, N.A., & Anderson, J.R. (2010). The past, present, and future of cognitive architectures. *Topics in Cognitive Science, 2*, 693–704.

Tempelmann, S., Kaminski, J., & Liebal, K. (2011). Focus on the essential: All great apes know when others are being attentive. *Animal Cognition, 14*, 433–439.

Tomasello, M., & Call, J. (2006). Do chimpanzees know what others see—or only what they are looking at? In S. Hurley & M. Nudds (Eds.), *Rational animals* (pp. 371–384). Oxford: Oxford University Press.

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees understand psychological states—the question is which ones and to what extent. *Trends in Cognitive Sciences, 7*, 153–156.

Tomasello, M., Call, J., & Hare, B. (2003). Chimpanzees versus humans: It's not that simple. *Trends in Cognitive Sciences, 7*, 239–240.

van der Vaart, E., Verbrugge, R., & Hemelrijk, C.K. (2011). Corvid caching: Insights from a cognitive model. *Journal of Experimental Psychology: Animal Behavior Processes, 37*, 330–340.

van der Vaart, E., Verbrugge, R., & Hemelrijk, C.K. (2012). Corvid re-caching without 'theory of mind'. *PLoS ONE, 7*, e32904. doi:32910.31371/journal.pone.0032904.

van Maanen, L., van Rijn, H., & Taatgen, N.A. (2012). RACE/A: An architectural account of the interactions between learning, task control, and retrieval dynamics. *Cognitive Science, 26*, 62–101.

Vonk, J., & Povinelli, D.J. (2011). Preliminary investigations of cognitive plasticity: Social and physical causality in home-reared chimpanzees. In N. Eilan, H. Lerman, & J. Roessler (Eds.), *Perception, causation and objectivity* (pp. 342–367). Oxford: Oxford University Press.

Vonk, J., & Shackelford, T.K. (2012). Toward bridging gaps: Finding commonality between evolutionary and comparative psychology. In J. Vonk & T.K. Shackelford (Eds.), *Oxford handbook of comparative evolutionary psychology.* (pp. 3–16). New York: Oxford University Press.

Wein, J.M., & Stephens, D.W. (2011). Caching economics: Jays cache more when handling times are short and habitats are poor. *Animal Behaviour, 82,* 579–585.