

Mentalizing capacities in chimpanzees (*Pan troglodytes*)

by

Carla Krachun

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Institute of Cognitive Science

Carleton University

Ottawa, Ontario, Canada

Copyright © Carla Krachun 2008



Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*
ISBN: 978-0-494-47480-8
Our file *Notre référence*
ISBN: 978-0-494-47480-8

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

Mentalizing capacities in chimpanzees (*Pan troglodytes*)

by

Carla Krachun

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Institute of Cognitive Science

Carleton University

Ottawa, Ontario, Canada

Copyright © Carla Krachun 2008

ABSTRACT

Mentalizing refers to the capacity to recognize and respond to mental states (e.g., perceiving, knowing and believing), in both oneself and others. Children have well-developed mentalizing capacities by about 5 years of age, best demonstrated by their ability to recognize when their own and others' mental states are false (e.g., false beliefs). To determine whether this is a uniquely human ability, I tested false belief understanding in apes, focusing on chimpanzees. Chimpanzees have performed poorly in previous false belief tests, but those tests required them to cooperate for food and understand communicative intentions, both problematic for chimpanzees. The two novel false belief tests I gave chimpanzees (and also bonobos, in one test) avoided these problems, but the apes still failed. I also investigated whether chimpanzees were capable of recognizing when they *themselves* were experiencing false mental states. First, however, because very little work has investigated chimpanzees' most basic mentalizing capacities with regard to themselves, I tested their ability to recognize what they themselves could or could not see from different visual perspectives. After establishing that chimpanzees have a good understanding of their own visual perception, I examined their ability to recognize when their own perceptions were mistaken. In the children's literature, this is often investigated with appearance-reality tests, in which children must recognize that their perception of an object need not correspond with reality (e.g., a magnified object looks big but is really small). Children pass such tests around the same age as they pass false belief tests, and many researchers theorize that the tests draw on common mentalizing capacities. My appearance-reality test for chimpanzees involved presenting them with a choice between a small and a big grape that were magnified and minimized, respectively, to make their

apparent relative sizes the reverse of reality. Some chimpanzees passed this test by most often choosing the truly bigger (apparently smaller) grape, suggesting that recognizing one's own false mental states may be easier than recognizing others', at least for chimpanzees. Alternatively, chimpanzees may find perceptually based states easier to recognize than false belief states. I ended with recommendations for future research to investigate these and other outstanding questions.

ACKNOWLEDGEMENTS

The pronoun ‘we’, when used to mean ‘I’, is jokingly referred to as the ‘Royal We’. In this thesis, I make liberal use of what I call the ‘Royal I’: I tested apes, I ran control trials, I considered alternative explanations. In reality, the people involved in making this thesis possible span two continents. In Germany, Michael Tomasello, Malinda Carpenter, and Josep Call played huge roles in helping me plan and execute my research. I am thrilled to have them as co-authors on the publications resulting from this work (see Appendix). Also in Germany, a large and constantly shifting contingent of practical students, zookeepers and research assistants provided invaluable help at every stage of testing. The Max Planck Institute is truly one of the most supportive places in the world to do research, and I consider myself extremely fortunate to have been able to work there.

The Canadian crew was smaller but no less amazing in their support and assistance. Deepthi Kamawar not only supervised my work with an extremely high level of intelligence and competence but was also just a really great person. Her pep talks were some of the best I’ve ever received from anyone, on this or any other continent. I would also like to thank my external examiners, Valerie Kuhlmeier and Mary Gick, and the other members of my thesis committee: Catherine Plowright, who taught me how to test pigeons in my pre-ape days; Andrew Brook, who persistently forced me to think hard about what things *meant*; and Robert West, who first got me interested in cognitive science by asking “Can computers be conscious?”

I am also extremely grateful to Colleen Fulton and Lianne Dubreuil, not just for providing adept administrative help but also for their warmth and encouragement.

Thanks also to my fellow students in both Canada and Germany for sharing the roller coaster ride, and to Steve Zabarylo for letting my junk and mail (and junkmail) live at his place while I lived in Germany. And most importantly, I want to express appreciation to my family and friends for being so wonderfully supportive. In particular, I thank my mother and father for their mothering and fathering (sadly, my mother died before I could finish this); my best friend Suzanne for her unwavering confidence in me; my partner David for supporting me in every possible way while I wrote this thesis; and my daughter Stacy for happily and selflessly accepting having a full-time student for a mother.

Funding was provided by the Max Planck Institute for Evolutionary Anthropology; Carleton University's Institute of Cognitive Science and Faculty of Graduate Studies; and the Ontario Graduate Scholarship program. For this, I am also very grateful.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	vi
LIST OF TABLES.....	ix
LIST OF FIGURES.....	x
LIST OF APPENDICES.....	xii
CHAPTER 1 – Introduction.....	1
Mentalizing in Humans	1
Why Look for Mentalizing in Chimpanzees?	8
Focus of This Thesis	16
CHAPTER 2 – Conceptual Issues.....	20
The Meaning of ‘Mental States’	20
Necessary Conditions for Knowing	23
The Theory of Mind Mechanism.....	28
Mindreading <i>Versus</i> Behavior Reading	29
CHAPTER 3 – Research to Date	34
Visual Perspective Taking: Understanding Seeing and Its Relation to Knowing	34
Children	35
Chimpanzees	40
False Belief: Understanding the Effect of <i>Not</i> Seeing on Belief States	55
Children	55
Chimpanzees	61
Appearance <i>Versus</i> Reality: Recognizing When Seeing Conflicts With Knowing	66
Children	66
Chimpanzees	71
Summary	75
Next Steps	75

CHAPTER 4 – False Belief	77
Study 1 – Competitive Reaching Study	77
Experiment 1 – Children	81
Methods	81
Results and Discussion.....	90
Experiment 2 – Chimpanzees and Bonobos.....	98
Methods	98
Results and Discussion.....	107
General Discussion.....	113
Study 2 – Change-of-Contents Study	119
Previous Study with Children	122
Current Study with Chimpanzees.....	124
Methods	124
Results and Discussion.....	131
General Discussion of Studies 1 and 2	135
CHAPTER 5 – Visual Perspective Taking for Self	138
Study 3 – Perspective-shifting Study	138
Experiment 1 – Perspective-shifting Test	142
Methods	142
Results and Discussion.....	152
Experiment 2 – UP <i>versus</i> OPPOSITE	160
Methods	161
Results and Discussion.....	163
Experiment 3 – UP <i>versus</i> DOWN.....	165
Methods	166
Results and Discussion.....	172
General Discussion.....	173
CHAPTER 6 – Appearance <i>versus</i> Reality	177
Study 4 – Lens Study	177
Experiment 1 – Children	182
Methods	183
Results and Discussion.....	191
Experiment 2 – Chimpanzees.....	196
Methods	196
Results and Discussion.....	209
General Discussion.....	216

CHAPTER 7 – Final Comments	221
Summary and Meta-analysis of Findings	221
Implications and Future Directions	225
Conclusion.....	234
 REFERENCES	 236
 APPENDIX – Articles Based on This Thesis	 263

LIST OF TABLES

Table 4.1: Proportion of correct True belief and False belief trials for individual chimpanzees in Study 2	133
Table 6.1: Number of correct trials for individual chimpanzees in various components of the Lens test, Study 4	210
Table 7.1: Proportion of correct trials across tests for the 15 apes participating in at least two of the four studies.....	223

LIST OF FIGURES

Figure 1.1: Evolutionary timeline	2
Figure 4.1: Experimental set-up for Study 1	82
Figure 4.2: Choice analysis for Study 1: mean proportion of correct trials for children and apes in both versions of the procedure combined and separately.....	92
Figure 4.3: Looking analysis for Study 1: mean proportion of trials in which children and apes looked toward the unchosen container before choosing.....	95
Figure 4.4: Indecisive looking: A chimpanzee looks at the container on her right just before touching the one on her left to choose it in Study 1.....	105
Figure 4.5: Mean proportion of correct trials for children in Krachun et al.'s (in preparation) True and False belief tests	123
Figure 4.6: False belief procedure for Study 2.....	126
Figure 4.7: Mean proportion of correct trials for chimpanzees in Study 2	132
Figure 5.1: (a) Photograph of testing room and (b) schematic diagram with the sides of the booth fanned outward to make the windows visible.....	144
Figure 5.2: Perspective-shifting test containers for Study 3	146
Figure 5.3: Mean proportion of trials in Study 3, Experiment 1, in which chimpanzees attempted to look into the containers	154
Figure 5.4: Mean proportion of trials in Study 3, Experiment 1, in which chimpanzees first looked first from the UP, SIDE, or OPPOSITE (OPP) perspective or made no response (NR).....	156
Figure 5.5: Mean proportion of trials in Study 3, Experiment 1, in which chimpanzees first moved UP, SIDE, or OPPOSITE (OPP) or made no response (NR)	158
Figure 5.6: Mean proportion of trials in Study 3, Experiment 2, in which chimpanzees first looked from the UP or OPPOSITE (OPP) perspective or made no response (NR)	165
Figure 5.7: Experimental set-up for the two conditions in Study 3, Experiment 3: the Three-tall condition (top) and the Three-short condition (bottom).....	167

Figure 5.8: Mean proportion of trials in Study 3, Experiment 3, in which subjects first looked from the UP or DOWN perspective	172
Figure 6.1: Containers used in the Lens test, Study 4, showing the real and apparent sizes of the grapes.....	185
Figure 6.2: Children’s median number of correct trials in the various components of the Lens test, Study 4	191
Figure 6.3: Number of children passing the various additional appearance-reality tests administered in Study 4.....	192
Figure 6.4: Containers used in the reverse contingency control tests of Study 4.....	199
Figure 6.5: Chimpanzees’ median number of correct trials in the Lens test, Study 4, as compared to the two reverse contingency control tests	213
Figure 7.1: Performance of individual chimpanzees in the Perspective-shifting test of Study 3 and the Lens test of Study 4.....	224

LIST OF APPENDICES

Appendix: Articles Based on This Thesis	263
---	-----

CHAPTER 1 – INTRODUCTION

In “A Report to an Academy,” Kafka imagines an ape who, in a desperate attempt to escape his captivity, teaches himself to be human. The story is interesting because it urges us to consider what distinguishes the human mind from that of our closest evolutionary relatives. Kafka’s ape achieved humanity in five years—our transformation has taken a little longer. The human evolutionary line diverged from the chimpanzee and bonobo line an estimated 5 million years ago, and the ancestor we share with all four great ape species existed about 14 million years ago (Figure 1.1).¹ What cognitive changes happened during those years to result in our uniquely human way of interacting with the world? What makes us alone capable of building cities and creating symphonies and giving reports at academic conferences? Comparative behavioral research provides one means of trying to answer such questions. The goal is to identify similarities and differences in the cognitive capacities of humans and apes, evident in their behavioral responses to similar kinds of problems. We can then assume with some confidence that shared capacities were present before the evolutionary divergence, whereas capacities found only in humans evolved afterwards (Byrne, 1995).

Mentalizing in Humans

One way in which some suggest that humans differ from apes is in our capacity to recognize and respond to mental states (see Povinelli & Vonk, 2003, and an edited

¹ Throughout this thesis, when referring specifically to one of these four species I use the name for that species; when referring to apes generally or to two or more species collectively I use the more inclusive term ‘apes’.

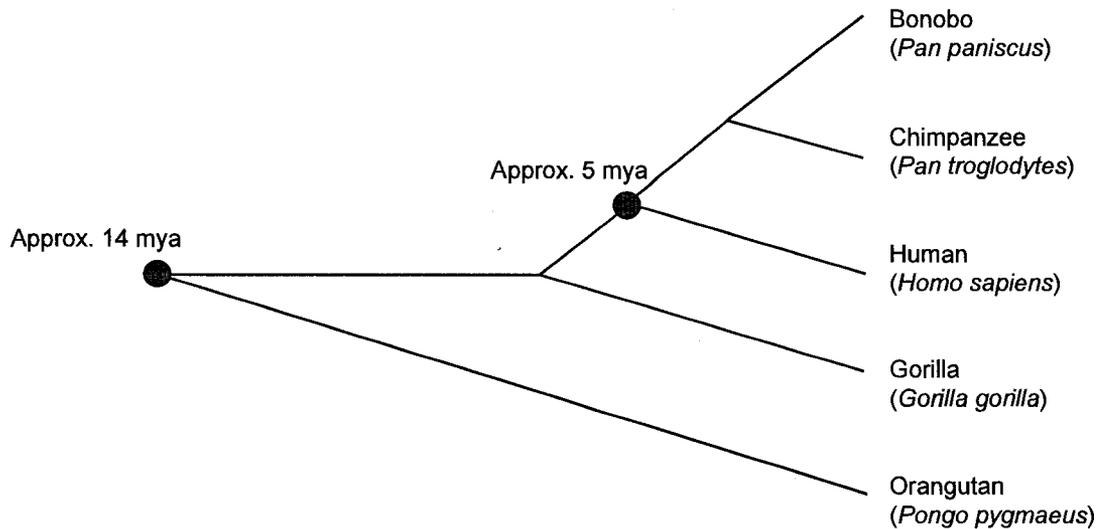


Figure 1.1. Evolutionary timeline, adapted from Whiten & Suddendorf (2007, p. 33). mya = millions of years ago

collection by Astington & Baird, 2005). We see others as not just physical and behavioral but also mental agents—as creatures capable of perceiving, thinking, knowing, believing, feeling, wanting, intending, and so on. At the same time, we appreciate that others’ mental experiences at any given moment may differ from our own—that others may perceive, know or believe something we do not, and vice versa. Related to this, we know that others’ mental states, as well as our own, may be at odds with the true state of affairs in the world, as in the case of mistaken beliefs and perceptions. Finally, we can use our inferences about mental states to predict and explain behavior.

A wide variety of terms have been used to describe these capacities, including mentalizing, mental-state attribution, mindreading, theory of mind, folk psychology, common sense psychology, belief-desire psychology, second-order representation, metarepresentation, and metacognition (although this last term often refers specifically to

the attribution of mental states to oneself). I prefer the terms mentalizing, mental-state attribution, and mindreading, which I use interchangeably throughout this thesis. While ‘theory of mind’ is also very commonly used among developmental and comparative psychologists, I avoid this term because it is somewhat misleading, for reasons that I discuss in Chapter 2.

For some years, there was a sense among researchers that the ability to recognize and respond to mental states was something one either had or did not have. In many respects, mental states of all varieties have much in common: they are all internal, personal, subjective, and representational. Intuitively, it seems that if one can recognize one type, one should be able to recognize them all. As the results of more and more studies poured in, however, it became apparent that the capacity to recognize different mental states emerges at different ontogenetic stages. While there is some variability in findings across studies depending on the methods used, the appreciation of others’ intentions and desires appears to emerge earliest in development by about 1.5 years of age (e.g., Gergely, Bekkering, & Kiraly, 2002; Meltzoff, 1995; Repacholi & Gopnik, 1997). This is followed by visual perspective taking beginning at about 2 years of age (Moll & Tomasello, 2006), and then by an understanding of both others’ and one’s own knowledge states by 3 to 4 years of age (Povinelli & DeBlois, 1992). In their everyday speech, children also begin to explain others’ actions in terms of beliefs by 3 to 4 years of age (Bartsch & Wellman, 1989). However, they do not reliably begin to pass most tests requiring them to recognize that others can entertain *false* beliefs about the world until

they are about 5 years old (Wellman, Cross, & Watson, 2001). These findings suggest that false belief understanding is a relatively advanced mentalizing capacity.

Two tests of false belief understanding have become so widely used that they have come to be known as the 'standard' false belief tests. The first is variously called the *Sally-Anne*, *Maxi*, or *Change-of-location* test (e.g., Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983). In this test, the child watches a story acted out in which one character, Sally, places an object in a particular location and then leaves the room. While Sally is gone a second character, Anne, enters and moves the object to a new location. Anne then leaves and Sally returns to retrieve the object. The child is asked to predict where Sally will look for the object, and the obvious correct answer is where she left it. A correct answer indicates that the child understands that Sally's failure to witness the moving of the object caused her to have a false belief about its location.

The other standard test is known as the *Smarties*, *Unexpected contents*, or *Change-of-contents* test (Perner, Leekam, & Wimmer, 1987). In this test, children are shown a Smarties box and asked to say what they think is inside (they usually say Smarties, or candy). The experimenter then opens the box and shows the child that there is actually something unexpected inside, such as crayons. The box is then closed and the child is asked to predict what somebody else who comes into the room and looks at the box will think is inside of it, before the box is opened. To answer correctly (by saying Smarties or candy), the child must understand that someone who has not already seen the actual contents of the box will falsely believe there are Smarties inside. This test has also been used to examine children's understanding of how their own belief states are affected

by perceptual experience. This is accomplished by asking children to state what they used to think was in the box before they saw what was in it. Children who recognize that they had a prior false belief about the contents of the box should be able to answer that they used to think there were Smarties in the box, even though they now know better. Typically, the standard false belief tests also include memory and comprehension questions to verify that children understand the task and are able to remember the prior and current locations of the treat (in the Sally-Anne test) or the prior and current contents of the box (in the Smarties test). Even when children fail the false belief test questions, they pass these control questions the vast majority of the time.

False belief tests were long considered to be the acid test of mental-state understanding. This is because a child who understands that another person can hold an incorrect belief about the world (and who behaves in accordance with that understanding) clearly recognizes that people can mentally represent the world in ways that may or may not accord with reality. False belief understanding is no longer equated with mental-state understanding more generally. As we saw above, the latter also includes understanding of desires, perceptions, intentions, and so on (see Bloom & German, 2000). It is nevertheless still considered by many to be the essential test of children's ability to recognize others' belief states.

Another common way of testing children's ability to recognize that mental states may not accord with reality are appearance-reality tests. Unlike false belief tests, however, appearance-reality tests require children to recognize that their *own* mental states (in this case, their own visual perceptions) can be at odds both with what they

know to be the case and also with what truly is the case. The most well-known appearance-reality test for children is the *Rock-sponge* test (Flavell, Flavell, & Green, 1983; Flavell, Green, & Flavell, 1986) in which children are presented with a sponge painted to realistically resemble a rock (other similarly deceptive objects are sometimes used). To pass the test, children must correctly say both what the object looks like (rock) and what it really is (sponge). In another standard test, the experimenter moves an object behind a transparent, colored filter and then questions children about the object's real and apparent color (Flavell et al., 1986; Flavell, Green, Wahl, & Flavell, 1987). Children have also been asked to comment on the real and apparent size or shape of objects distorted by various means, such as by placing them in a glass of water or behind magnifying or minimizing lenses (Braine & Shanks, 1965a, 1965b). Children typically begin to pass these tests at about 4 to 5 years of age, around the same time that they begin to pass the standard false belief tests.

Researchers have noted striking similarities in children's performance on appearance-reality and false belief tests (e.g., Flavell, 1993; Gopnik & Astington, 1988). In standard verbal versions of both types of test, there is a clear developmental progression in performance from 3 to 5 years of age (Flavell, 1986, 1993; Flavell et al., 1983; Wellman et al., 2001), and younger children tend to perform better in tests involving nonverbal responses (Carpenter, Call, & Tomasello, 2002; Matsui & Miura, 2008; Sapp, Lee, & Muir, 2000) or deceptive elements (Hala, Chandler, & Fritz, 1991; Rice, Koinis, Sullivan, Tager-Flusberg, & Winner, 1997; Sullivan & Winner, 1993). For both appearance-reality and false belief tests, there is some evidence that young

children's difficulties in standard verbal procedures may be due to misinterpreting the test questions (Deák, 2006; Hansen & Markman, 2005; Lewis & Osborne, 1990). And significant positive correlations in performance on appearance-reality and false belief tests have sometimes been observed (e.g., Andrews, Halford, Bunch, Bowden, & Jones, 2003; Frye, Zelazo, & Palfai, 1995; Gopnik & Astington, 1988; Moore, Pure, & Furrow, 1990).

Flavell (2000) has also noted positive relations between both false belief and appearance-reality understanding and another mentalizing capacity: visual perspective taking. In visual perspective taking tests, children must judge what others see (or do not see) from a perspective that differs from their own (e.g., Flavell, Shipstead, & Croft, 1978; McGuigan & Doherty, 2002; Moll & Tomasello, 2006). One common assumption is that parallels in performance across visual perspective taking, false belief and appearance-reality tests reflect the fact that they all rely on the same underlying capacity: to recognize that things can be represented in two different ways by two different people, and even by the same person (e.g., Flavell, 2000). Thus, all these tests require one to simultaneously consider two conflicting representations of the same object or state of affairs (known as 'dual representation'). In the case of false belief and appearance-reality tests, one must further recognize that representations can be at odds with what is truly the case. Thus, visual perspective taking skills are almost surely foundational to developing the more advanced forms of mental-state understanding measured by false belief and appearance-reality tests (see Baron-Cohen, 1995, Chapter 4; Gopnik, Slaughter, & Meltzoff, 1994). It is not difficult to imagine how this might be the case when one further

considers the intimate relationship between perceiving and knowing or believing. Visually perceiving objects or events leads directly to knowledge or beliefs being instilled in the mind of the perceiver. However, the accuracy of the knowledge or beliefs instilled depends on precisely what was or was not perceived, and also on whether or not there were any distorting influences. Perceptions can be inaccurate, leading in turn to inaccuracies in the knowledge and beliefs they engender. There is an important distinction to be made here between perceptions, knowledge, beliefs, and other kinds of mental states. Because they are externally driven, perceptions, knowledge, and beliefs can be wrong. This is not the case for internally driven mental states such as desires and intentions. One cannot mistakenly desire or falsely intend something, but one can clearly have mistaken perceptions and false beliefs. A number of people have pointed to such distinctions between different kinds of mental states in their writing (see, for example, Premack & Dasser, 1991). In short, mental states such as knowledge and beliefs are highly vulnerable to what information comes in and in what form, and so what one perceives and how one perceives it is closely linked up with what one knows and believes. Given all of this, we should not at all be surprised to find connections between the ability to understand that others' perceptual states may differ from one's own, that others' belief states may differ from one's own, and that both perceptual and belief states may differ from reality.

Why Look for Mentalizing in Chimpanzees?

It is by no means universally agreed upon that the mentalizing capacities described above are likely to be uniquely human (e.g., Byrne, 1995; Gómez, 2004;

Terrace, 2005; Tomasello, Call, & Hare, 2003a). There are, in fact, a number of reasons to suspect that chimpanzees might possess at least some of the same capacities for mental-state attribution that we do. Along with bonobos, they are more closely related to humans evolutionarily and genetically than any other currently existing species (Wildman, Uddin, Liu, Grossman, & Goodman, 2003). And like humans, chimpanzees and other apes show differential expansion of the prefrontal cortex relative to the rest of the brain in comparison with other mammals and even other nonhuman primates, although the expansion is most pronounced in humans (Preuss, 2000). Imaging studies in which brain activity is recorded while participants carry out tasks requiring mental-state attribution suggest that, at least in humans, the prefrontal cortex is implicated in mindreading (Happé et al., 1996; Singer, 2006). Baron-Cohen (1995), among others, proposes that the human mindreading system was developed by natural selection. It is therefore possible that at least some aspects of this system may have developed before the chimpanzee line split from the human line. Some researchers have also pointed out that chimpanzees' complex social ecology—with clear dominance hierarchies, fission-fusion dynamics whereby subgroups emerge and dissolve, and shifting allegiances and alliances—would have exerted significant adaptive pressure for the development of advanced forms of social understanding (i.e., the so-called Social Intelligence Hypothesis: Humphrey, 1976; Jolly, 1966). Such understanding might include the ability to not just infer mental states but also to manipulate others' mental states and associated behaviors to one's own advantage, for example by withholding information from others or even deliberately misleading them (dubbed "Machiavellian intelligence" by Byrne &

Whiten, 1988). Observations of chimpanzees in the wild have produced many compelling anecdotes to suggest that they do indeed possess such a capacity for deception (Byrne & Whiten, 1988, 1990; Whiten & Byrne, 1988). Finally, as we saw above, some tests of mental-state understanding require the capacity to conceive of an object or a situation in two different ways simultaneously. Chimpanzees have demonstrated such a capacity in at least one non-mental domain. Namely, they are able to use a scale model of an enclosure as a source of information about where food is hidden in the enclosure (Kuhlmeier & Boysen, 2001; Kuhlmeier, Boysen, & Mukobi, 1999). This indicates that they are able to simultaneously consider the model as both an object in its own right and also as a representation of something else.

Thus, there are several good reasons why we might expect to find mental-state attribution capacities in chimpanzees. Nonetheless, we must also keep in mind that while chimpanzees may be very similar to humans genetically and neuroanatomically, they are not exactly like us (for reviews of the unique features of the human brain see Preuss, 2000, 2001). The impact of such differences on mentalizing abilities is not clear, but it is conceivable that even small disparities in neural structure could potentially be manifested in important differences in cognitive function. Povinelli and Giambrone (1999) also rightly point out that behaviors indicative of mentalizing in humans could be instantiated by different inner, unobservable processes in chimpanzees. Chimpanzees might, for example, produce all the same responses as humans by representing others' behavioral patterns rather than their mental states. Thus, while anecdotes from naturalistic observations may seem very compelling, many researchers consider them to be just

starting points for more rigorous experimental investigations that allow for the control of possible confounding factors. Povinelli and Vonk (2003; 2004) maintain that, even in laboratory settings, it will almost always be impossible to completely rule out non-mentalistic interpretations. This may be so, but we can at least be more confident, if never absolutely certain, that observed behaviors are indicative of mental-state understanding when they are expressed across a variety of situations and experimental methodologies (Heyes, 1993; Tomasello & Call, 2006), and when appropriate controls make alternative interpretations unlikely.

Besides contributing to our knowledge of human cognitive evolution, knowing what apes might understand about the mind is important for two additional reasons. First, it can help us to determine which, if any, mentalizing capacities are possible without language. Some researchers and theoreticians argue that at least the more advanced aspects of mentalizing, such as recognizing when others have false beliefs, depend largely on linguistic abilities (see Astington & Baird, 2005). As we will see in Chapter 3, however, recent studies with pre-linguistic infants suggest that they may understand more about false beliefs than was previously thought. We will also see in Chapter 3 that there is much evidence that children possess many of the simpler aspects of mental-state attribution long before they are competent language users, such as the ability to attribute perceptual states to others. One might still argue, however, that the human mind is at least primed for language, and that the cognitive capacities that support language may give even pre-linguistic humans mentalizing abilities not found in other species (e.g., LeDoux, 1996). As apes never have a complex language at any point in their development,

determining what they understand about the mind can contribute importantly to this discussion. A final reason for wanting to understand more about apes' cognitive capacities is the animal welfare aspect. There are thousands of apes in zoos, research facilities, and sanctuaries around the world. Understanding their cognitive capacities will allow us to institute housing and handling practices that address not just their physical but also their emotional and intellectual needs (Held, Mendl, Devereux, & Byrne, 2001).

Premack and Woodruff's (1978) seminal paper "Does the chimpanzee have a theory of mind?" is often cited as the official launching point of the modern experimental study of mentalizing in apes. It also catalyzed a vast number of investigations on mindreading in human children. In the past three decades, mentalizing research with children has, in fact, far outgrown the ape research in its volume and scope, with children's understanding regarding the entire range of mental states being charted in hundreds of studies (recent general reviews include Baron-Cohen, Tager-Flusberg, & Cohen, 2000; Flavell, 2000, 2004; Repacholi & Slaughter, 2003; Suddendorf & Whiten, 2001; Wellman, 2002; Wellman et al., 2001). Mentalizing research with chimpanzees and other apes has also continued over the past 30 years (see reviews by Gómez, 2005; Johnson & Karin-D'Arcy, 2006; Povinelli, 2003; Tomasello & Call, 1997), but far less has been accomplished relative to research with children for a number of reasons, many of them practical. Because there are only a handful of facilities properly equipped for carrying out behavioral research with apes, apes are more difficult to gain access to than children. Researchers working with apes typically have to make do with a much smaller number of participants. Fewer than 10 is commonplace (e.g., Boysen, Berntson, &

Mukobi, 2001; Call & Carpenter, 2001; de Waal & Pokorny, 2008; Hare, Call, & Tomasello, 2006; Hirata & Matsuzawa, 2001; Horner & Whiten, 2007; Povinelli & Eddy, 1996b), and in many cases just one or two apes have been tested (e.g., Itakura & Tanaka, 1998; Matsuzawa, 1985; Muncer, 1983; Okamoto et al., 2002; Premack & Woodruff, 1978; Shillito, Shumaker, Gallup, & Beck, 2005; Younger & Bjorklund, 2004).

Additionally, because apes can be very aggressive, interactions with them have to usually take place through metal bars or thick glass barriers, greatly restricting the kinds of exchanges that can occur. The physical configuration of the enclosure further constrains the kinds of tasks that can be carried out, and experimental apparatuses must be built to withstand apes' enormous strength. By far, however, the most difficult logistical problem to overcome is the communication barrier. Experimental tasks cannot be explained verbally for apes in the way they can be for even very young children. Apes must instead figure out what is expected of them in any given task through trial and error. This can be especially problematic when the same apes are participating concurrently in several studies, increasing the potential for confusion and interference across tasks. Training apes in the basic components of a task to prepare them for testing can thus be very time consuming, and studies with apes can take much longer than those with children. The result of all this is that, with apes, we are still very much at the 'can-they-or-can't-they' stage of questioning—still cataloguing which mentalizing capacities they share with humans and which they do not.

There is some evidence that chimpanzees have the capacity to recognize some mental states in others, including desires (Buttelmann, Call, & Tomasello, 2007),

intentions (Buttelmann, Carpenter, Call, & Tomasello, 2007), visual perceptions (Hare, Call, Agnetta, & Tomasello, 2000; Hare et al., 2006; Melis, Call, & Tomasello, 2006), and knowledge construed as what others have or have not seen in the recent past (Hare, Call, & Tomasello, 2001; Hirata & Matsuzawa, 2001; Menzel, 1974). However, there are still large gaps in our knowledge of what chimpanzees understand about both others' and their own mental states. With regard to the former, the question of whether chimpanzees can recognize others' false belief states has barely been touched upon. Do chimpanzees understand that others can hold beliefs about the world that are at odds both with reality and with what they themselves know to be true? This is an important question because passing a false belief test is still considered by many to be the best evidence that one appreciates the representational nature of mental states (Wellman et al., 2001). Prior to my research, only two published studies (Call & Tomasello, 1999; O'Connell & Dunbar, 2003) had tested apes' false belief understanding, and the apes performed poorly in both. They failed Call and Tomasello's (1999) false belief test; and they passed O'Connell and Dunbar's (2003) false belief test but failed a crucial control test, thereby invalidating the false belief results. However, there is some evidence to suggest that those tests may have been inherently difficult for apes, for reasons that will be discussed later. Thus, apes' capacity to recognize false beliefs in others is an area much in need of further investigation.

Additionally, apes' capacity to recognize their *own* false mental states, as in appearance-reality tests, has never before been directly investigated. As with false beliefs, it has been argued that the best evidence that apes (or other animals) impute mental states

to themselves would be behavior indicative of recognizing when they were experiencing false perceptions (Allen & Bekoff, 1997; Carruthers, 2008). As we saw above, false belief understanding and appearance-reality understanding may rely on the same underlying capacity to recognize how one thing (situation, object, etc.) can be represented in two different ways simultaneously, one of which is false. Testing apes' appearance-reality understanding would thus be instructive, as follows: If apes either passed both tests or failed both tests, this would provide good evidence that they were (in the former case) or were not (in the latter case) capable of understanding that mental states can be false. On the other hand, if they failed one test but passed the other, this would suggest that they may be able to recognize false mental states in others but not themselves, or vice versa, as the case may be. It is thus important to test apes' capacity to distinguish appearance from reality, as has been done with human children.

At the same time, it must also be recognized that, up to this point, precious little has been done to determine apes' most basic mentalizing capacities with regard to themselves, such as their understanding of their own simple visual perceptions and knowledge states. In research on what apes understand about others' mental states, it was established that apes were capable of recognizing states of perceiving and knowing in others before studies were undertaken to see whether they also recognized others' false beliefs. Similarly, in research on what apes understand about their own minds, it is important to establish their basic capacities before moving on to test their more advanced ability to recognize their own false mental states. In the only published study on the subject previous to my own research, Call and Carpenter (2001) found that chimpanzees

(and also 2-year-old children) could discriminate between their own states of knowing or not knowing as a function of what they had or had not seen in the recent past. As we will see later, however, that study was limited in a number of respects that need to be addressed. Further tests of apes' basic capacities regarding their own mental states are in order.

Focus of This Thesis

As we saw earlier, there are positive relations among children's visual perspective taking skills, false belief understanding, and ability to distinguish appearance from reality. This is likely due to the element of dual representation common to all of these, and also to the intimate relation between perceiving, knowing and believing discussed earlier. It is also likely due, at least in part, to the fact that all of these mental states have one important feature in common: they can be wrong. I therefore chose to focus on apes' capacities regarding these mental states in this thesis, to the exclusion of other, qualitatively different kinds of mental states such as desires and intentions.

Much is known about children's mentalizing capacities regarding all of these mental states; less is known about apes'. For example, very little has been done to investigate apes' capacity to recognize when others have false beliefs, despite the fact that this is widely considered to be the most convincing test of mental-state attribution. Apes' capacities with regard to their own, as opposed to others', mental states have also received little attention. As yet, no one has directly tested whether apes recognize when they themselves are experiencing a false perception, as in appearance-reality tests. And while a fair amount of research has examined apes' visual perspective taking skills with

respect to others, little has been done to examine how much they understand about their own visual experiences.

In this thesis, I strove to fill these gaps in our knowledge and improve our understanding of how apes' mentalizing abilities differ from those of humans. My main goal was to test apes' capacity to recognize when mental states, both others' and their own, are mistaken or false. The ability to pass false belief and appearance-reality tests emerges relatively late in childhood and is therefore considered to involve the more advanced components of mindreading. As almost no research has examined apes' more basic capacities with regard to their own mental states, my secondary goal was to do this. Specifically, I focused on apes' understanding of their own visual perceptual experiences.

I began with two false belief studies: the *Competitive reaching* study (Study 1) and the *Change-of-contents* study (Study 2), both of which are reported in Chapter 4. In each of these studies, apes had to recognize that an experimenter had a false belief because she failed to witness one important event in a sequence of events, and they had to predict her behavior on that basis. In Study 1, the experimenter's false belief was about the location of a hidden food item; in Study 2 it was about the hidden contents of a box. Both studies addressed potential problems inherent in previous false belief tests that might have hindered apes' performance. If apes passed the false belief tests, this would indicate that their previous failures were likely due to these task factors; if they failed, this would provide further evidence that apes are incapable of recognizing others' false beliefs. In the remainder of the thesis I investigated apes' capacities regarding their own mental states. As almost no research had been carried out so far in this area, I started in Chapter

5 with the basics by examining apes' visual perspective taking abilities with respect to themselves in the *Perspective-shifting* study (Study 3). I did this by engaging the apes in a food-finding task in which they had to recognize which of several different visual perspectives they needed to adopt under different conditions in order to see the hidden food. Follow-up experiments, run immediately after the *Perspective-shifting* test and also reported in Chapter 5, ruled out the possibility that they were using simpler mechanisms to solve the task. Given that apes have demonstrated good visual perspective taking skills with respect to others in previous studies, I expected them to have little difficulty in these tasks (although, it is conceivable that apes could do well in tests requiring them to judge others' visual perspective but poorly in tests requiring them to judge their own, for reasons I come back to later). Nevertheless, it was important to establish apes' basic mentalizing capacities with regard to themselves before moving on to look at their more advanced skills in this respect, which I did in the final study, reported in Chapter 6. In the *Lens* study (Study 4), I examined apes' understanding of their own false mental states with an appearance-reality test. In this test, I presented apes with a situation in which what they knew to be the case was in conflict with what currently appeared to be the case. Specifically, what looked like a big grape was actually a small one, and vice versa. Apes had to recognize this in order to choose the truly bigger and more desirable grape. Given that this was, as far as I am aware, the first time anyone has directly tested apes' capacity to distinguish appearance from reality, it was difficult to predict how they might perform. However, the fact that they have had some success in liquid conservation tasks requiring them to understand how the volume of a liquid remains constant across changes in the

shape of its container (Call & Rochat, 1996, 1997; Muncer, 1983; Suda & Call, 2004; Woodruff, Premack, & Kennel, 1978) suggested they might also evidence some degree of success in my appearance-reality test.

Throughout my research, I focused on chimpanzees because of their close evolutionary relatedness to humans, and I also included bonobos in one study.² In addition, for comparative purposes, and also to validate what I thought would be the most challenging tasks for the apes, I gave human children adapted versions of my false belief and appearance-reality tests. Before presenting my studies, I first discuss in Chapter 2 a number of conceptual challenges involved in carrying out mentalizing research with apes. In Chapter 3, I then review the most relevant research on mentalizing in children and chimpanzees, allowing me to better situate my own experiments within the larger context of the work carried out to date in this area.

² It was unfortunately not practical to include bonobos in the remaining studies.

CHAPTER 2 – CONCEPTUAL ISSUES

In addition to some of the practical difficulties mentioned in Chapter 1, the study of mentalizing in apes involves a number of conceptual challenges. Some of these are inherited from long-standing philosophical debates about the nature of the human mind and human mental experience. Others are specific to the project of attempting to understand the unobservable inner experiences of nonlinguistic species. In this chapter, I outline the issues and discuss their implications for the study of ape mentalizing. My main purpose in doing so is to raise issues that might have a bearing on the interpretation of my research findings. I also describe how I have dealt with each of these issues within the context of this thesis.

The Meaning of ‘Mental States’

The first issue to be dealt with is the widespread disagreement about the meaning of the very entities being studied in mentalizing research. Mental states are not at all well understood, even in humans. In fact, there are some who would argue that mental states do not exist but are completely illusory. Such ‘eliminativists’ claim that what we think of as beliefs, knowledge, and so on are in no way correlated to anything real going on the brain and have no causal effect on behavior (e.g., Churchland, 1981, 1988). If this is the case, then research on mental-state attribution, in both apes and humans, is entirely misguided. Fortunately, most people do not adhere to this extreme eliminativist stance on mental states.

Nevertheless, even among those who do accept the existence of mental states, there is much disagreement about what, exactly, mental states are: nonphysical

substances or properties? physical states of the brain? epiphenomena? dispositions to behave in certain ways under certain circumstances? multiply realized entities with particular causal roles? (see Ravenscroft, 2005 for a general introduction to the various theories). This inability to get a grasp on the very form and function of mental states would seem to present a major obstacle to studying them experimentally. Avramides (2001) refers to similar difficulties regarding the well-known philosophical problem of other minds:

Perhaps the question we ought to be asking is not, how do I know whether others (other human animals, non-human animals or machines) have minds, but rather, what is a mind such that we can understand that others as well as myself can be said to have a mind? (p. 3)

Extending this question to the study of mentalizing in apes, we might ask, what are minds such that we can understand whether apes recognize that they and others have minds? Thus, it seems that we need to know what constitutes mental phenomena before we can go about making them the object of study.

The philosopher Franz Brentano (1995 [1874]) argued that what distinguished the mental from the non-mental was that the former exhibited a characteristic called 'intentionality'. The philosophical meaning of this term should not be confused with the common use of the word *intention*, which is synonymous with purpose, goal, or objective. Rather, what makes mental states intentional in the philosophical sense is that they refer to, or have as an object, something outside of themselves. One desires, perceives, or knows *something*. Mental states are also often described as being intentional

in the sense that, unlike non-mental entities, they have representational content, or are *about* things other than themselves (Brook & Stainton, 2000). The province of Newfoundland is not about anything, for example, but one can have beliefs about Newfoundland (e.g., that people there eat cod tongues, that they have a unique dialect, that the former causes the latter, etc.). As these examples show, mental states may be identified by the fact that they can be expressed in the form of ‘propositional attitudes’, or words that introduce ‘that’ clauses. For example, he *believes that* the Newfoundland dialect is caused by eating cod tongues, but she *knows that* cod tongues do not affect speech. Also apparent in these examples is an important consequence of the representational nature of mental states: they can be false. This last feature is especially relevant for this thesis, given that three of my four studies are designed to test apes’ ability to recognize false mental states in others or themselves.

A complicating factor in all of this, however, is that not all the states that many label as mental possess the above features. Pain and other bodily sensations, for example, do not seem to be about anything; and desires and intentions cannot falsely represent the world in the way that perceptions and beliefs can. Also problematic is that some of the above features are also possessed by non-mental entities. Words and other symbols refer to or represent things outside of themselves, but they are not mental states. Similarly, one can cook, kill, and eat something, but few would describe killing, cooking, and eating as mental phenomena.

Taking on the challenge of discovering the precise nature of mental states (how they relate to the physical, the means by which they affect behavior, and so on) is far

beyond the scope of this thesis. Such ontological questions need not, in any case, overly concern us here. There clearly do exist particular types of internal experience that we label as mental states. The ones that I have chosen to focus on in this thesis (perceptions, knowledge and beliefs) are perhaps the least controversially mental. Few people today would dispute the claim that apes also have such internal experiences, even if they do not label them (e.g., see Gómez, 1996). My question is simply thus: Do apes have the capacity to recognize and respond to these particular kinds of internal experiences in both themselves and others?

Necessary Conditions for Knowing

In addition to disagreement regarding the nature of mental states, what it means to know (or understand or recognize) something is far from widely agreed upon, leading to potential difficulties in deciding what apes know about their own and others' mental states. A customary way of defining knowledge in philosophy is as justified true belief (e.g., Gettier, 1963): one knows something when (1) one believes it, (2) one is justified in believing it (because there is good evidence to support it, for example), and (3) it is objectively true. Thus, on this account, there can be no false knowledge, only false beliefs.

While this widely accepted definition is useful for distinguishing knowledge from belief, it is unfortunately not enough to dispel all confusion about what might or might not constitute knowledge. One source of disagreement is the importance of conscious experience in knowing. Can we claim that apes attribute mental states to themselves or others in the absence of any awareness of doing so on their part? Some people (with

whom I agree) would answer yes, pointing out that many stimuli are perceived and acted upon unconsciously (e.g., Schulkin, 2000). Others, however, claim—or at least imply in their writing—that mental-state attribution must involve conscious processing. Povinelli and Giambrone (2001), for example, question apes' ability to “reason about” mental states; and Flavell (1986), in discussing the relations among appearance-reality understanding, false belief understanding, and visual perspective taking, clearly identifies these as conscious processes:

Knowledge about the appearance-reality distinction is but one instance of our more general knowledge that the selfsame object or event can be represented (apprehended, experienced, etc.) in different ways by the same person and by different people. In this analysis, then, its development is worth studying because it is part of the larger development of our conscious knowledge about our own and other minds. (p. 419)

The effect of all this is to make experimental results overly subject to interpretation depending on one's theoretical leanings. If one accepts that knowing may be unconscious, for example, then the range of states one is willing to classify as knowing will be far greater than if one maintains that only conscious knowledge truly constitutes knowing.

In the experimental psychology literature on mindreading, the consciousness issue appears most often in the form of discussions about whether knowledge of mental states is explicit or implicit, with quite a bit of ensuing confusion. Explicit knowledge is often equated with conscious knowledge. This in turn is unfortunately often mistakenly equated with declarative knowledge, meaning knowledge that can be explained or talked about

(e.g., I know that Newfoundland is a province of Canada and I can verbally express this knowledge). Implicit knowledge, in contrast, is often equated with unconscious knowledge, which in turn is often mistakenly equated with procedural knowledge, such as knowing how to ride a bike but not being able to articulate the means by which we are able to do so. Following is one example of how explicit and implicit knowledge have been conceptualized by experimental psychologists. Clements and Perner (1994) suggest that children who can answer test questions correctly, either verbally or gesturally (for example by pointing), are demonstrating explicit knowledge of others' mental states. For these researchers, explicit knowledge is knowledge that is available to children in making active judgments about how to respond. They further propose that implicit knowledge is not used in such active responding, but may nevertheless show up in participants' spontaneous behaviors, such as in their looking responses. For example, even while actively choosing the wrong location, children might look toward the location where someone with a false belief would be expected to return and search for an object (called *anticipatory* looking; see Garnham & Ruffman, 2001).

There are a number of problems with these conceptions of explicit and implicit knowledge. First, apparently implicit knowledge may still be capable of driving active responses. In a study by Povinelli and DeBlois (1992), 3-year-old children responded correctly in a task requiring them to judge others' states of knowledge and ignorance but were unable to say how they knew what to do. They thus appeared to be using implicit knowledge to make active judgments (or, alternatively, they had explicit knowledge but were not able to articulate it, which is also problematic for the conceptions of knowledge

given above). Similarly, Gómez (1996) suggests that while apes may not have explicit knowledge of others' mental lives, they may have implicit know-how about them that allows them to choose the appropriate response in particular situations (i.e., a kind of working knowledge of mental states). Carpenter et al. (2002) suggest that perhaps active behavioral responses that do not involve explicit reporting but are also not just spontaneous reactions should be considered to fall somewhere in between implicit and explicit.

A further source of confusion is that some researchers claim that spontaneous behaviors could also be indicative of explicit, not just implicit, knowledge. Southgate, Senju, and Csibra (2007) maintain that anticipatory looking demonstrates explicit knowledge because it involves making a prediction, and that children who look at the correct choice but then respond incorrectly may do so because of other task demands. Once again, in this case, explicit knowledge would not necessarily be reportable. Further, Ruffman, Garnham, Import, and Connolly (2001) suggest that anticipatory looking might indicate explicit knowledge that participants are too uncertain about to act upon. In either case, participants would thus have explicit knowledge while appearing to have only implicit knowledge. In short, there is quite a bit of disagreement about what constitutes implicit versus explicit knowledge, much of which may stem from the failure to properly distinguish implicit from procedural knowledge and explicit from declarative knowledge. The fact is that declarative knowledge can also be implicit (most memories are) and procedural knowledge can also be explicit (we can articulate how to drive a car even after it becomes automatic). Given all these competing and conflicting accounts of implicit and

explicit knowledge, it is not surprising that some researchers (e.g., Call & Tomasello, 1999) admit they do not know quite how to characterize the difference between knowledge that drives active responses and knowledge that is manifested only in spontaneous behaviors. Call and Tomasello (1999) suggest that these reflect different “levels” of understanding, but what these levels correspond to is unclear. They could signify knowledge that is explicit versus implicit, but also knowledge that is certain versus uncertain, deeper versus shallower, or stronger versus weaker, to name a few possibilities.

Because of the widespread uncertainty over what might constitute implicit versus explicit knowledge, in this thesis I adopt Call and Tomasello’s (1999) notion of different levels of understanding. Like determining the nature of mental states, deciding what constitutes implicit versus explicit knowledge—or whether there can even be true knowledge without consciousness—is beyond the scope of my research. In fact, if consciousness is to be a criterion for explicit knowledge, it may be impossible in principle to ever determine if apes have explicit knowledge of any mental states. We currently have no way of detecting consciously experienced knowledge in nonverbal creatures. And even if neuroscientists could someday pinpoint the neural correlates of conscious experience in humans and then observe the same neural activity in apes, we could still never know with absolute certainty that that neural activity had the same effect in the ape’s cognitive system that it had in our own. Thus, in this thesis, I consider any positive results I find to be evidence of knowledge of mental states on some level, but I

make no impossible claims about whether that knowledge is explicit or implicit, or about whether apes consciously reflect upon it.

The Theory of Mind Mechanism

Besides uncertainty regarding the nature of mental states, and disagreement over what constitutes implicit versus explicit knowledge, the term ‘theory of mind’ has caused a fair amount of conceptual muddle. As noted in Chapter 1, in developmental and comparative psychology the term is currently used (synonymously with the terms mindreading, mental-state attribution, and so on) to refer to the ability to recognize and respond to mental states. However, in its traditional philosophical usage, it is meant to designate one hypothesized *mechanism* by which we do so: namely, by using an implicit folk theory of mental states. This folk theory is thought to be analogous to a scientific theory in that it includes rules and principles about which kinds of things are expected to occur under which kinds of conditions. In the case of theory of mind, the rules and principles would constitute a body of knowledge, perhaps innate (e.g., Leslie, 1994), about how people tend to act in certain kinds of situations depending on their beliefs, desires, and other mental states. Thus, according to this ‘Theory theory’ of mind (Carruthers, 1996), a mindreader examines the available evidence (observable behavioral/physical cues together with social/environmental context) to infer the content of others’ mental states and then uses those inferences to predict and explain their actions. However, it is not at all clear that this is the best explanation for how we go about inferring mental states and using them to predict behavior. Many believe it more likely involves a process of simulation, in which we imagine ourselves in another

person's situation and take our own decision-making system off-line to predict how the other person will respond—called the 'Simulation theory' of mind (e.g., Gordon, 1996). And, of course, there are numerous variations on both theories, as well as hybrid theories that include elements of both (for examples, see edited collections by Carruthers & Smith, 1996; Davies & Stone, 1995a, 1995b).

In short, the term 'theory of mind' as it is currently used refers to a particular capacity (to infer mental states) that could be instantiated by different hypothesized mechanisms. However, because it contains the word *theory* it seems to presuppose that the mechanism involved is theory-like rather than simulation-like. My research does not permit me to discriminate between these, or other, possibilities, so I have chosen to avoid confusion on the matter by steering clear of this term whenever possible. I should point out that uncertainty with regard to mechanism is not an obstacle to investigation. It is still possible to seek to determine whether or not apes can attribute particular types of mental states to themselves and others without specifying the means by which they are able to do so.

Mindreading *Versus* Behavior Reading

The final conceptual issue I will discuss here is specific to carrying out mentalizing research with nonhuman animals. How can we be sure that they are not using others' behavioral patterns to predict their actions without making any inferences about the contents of others' minds? I touched on this issue in Chapter 1 when I suggested that anecdotal observations seemingly indicative of mental-state attribution in apes were best

followed up with well-controlled experimental studies. In the latter, measures can be taken to minimize the problem, as I will discuss below.

At the heart of the issue is that many inner, unobservable mental states have outwardly visible physical and behavioral markers. For example, particular emotional states are reliably correlated with distinctive facial expressions and bodily postures. Similarly, the motivational state of wanting, for example, is apparent in the behaviors of approaching and trying to retrieve the desired object. And the perceptual state of seeing is reliably associated with the behaviors of orienting and gazing toward an object. It is thus possible that apes could use the physical and/or behavioral markers associated with particular mental states to predict others' behavior, without additionally inferring the contents of their minds (Heyes, 1993, 1998; Povinelli, 1994a, 1994b; Povinelli & Giambrone, 2001). This possibility makes it extremely difficult—some would say virtually impossible—to distinguish between responses based solely on behavior reading from those based on mindreading. I call this 'Povinelli's problem' because he and his colleagues have argued the issue most vehemently, claiming that apes have not needed any mentalizing capacities to achieve any of the behavioral feats they have so far performed in any research to date on the topic (Penn & Povinelli, 2007; Povinelli & Giambrone, 1999, 2001; Povinelli & Vonk, 2003, 2004). They could just as easily have responded, they argue, purely on the basis of learned contingencies between particular behaviors and particular responses (called the *low-level* model; additionally representing mental states is called the *high-level* model). Povinelli and his colleagues do not reject outright the idea that behavioral responses might tell us something about what is going on

in apes' minds, but they do clearly hold apes to a far stricter standard than humans. Behaviors considered to be indicative of mentalizing in humans are rejected as evidence when observed in apes. The reason, they claim, is that the argument by analogy is flawed: similar responses across species could be driven by different inner mechanisms, and so apes could conceivably produce all the same responses as humans using behavioral representation alone (Povinelli, Bering, & Giambrone, 2000; Povinelli & Giambrone, 1999).

Povinelli's problem is a particular version of the philosophical problem of other minds: just as we cannot know that others have minds by observing their behavior, we also cannot know that others *know that others* have minds by observing their behavior. Behavior is simply inadequate evidence of mentality. At least in the case of humans, we can know that they have minds because they can tell us (Dennett, 1996), and we can know that they know about other minds for the same reason. This is unfortunately not the case with apes. Povinelli's problem is admittedly a significant one to be contended with, not just in observational research but also in experimental studies. However, in the studies I present in this thesis, I take measures to minimize the possibility that any positive results can be easily attributable to non-mentalistic, low-level mechanisms. I include control manipulations to rule out alternative explanations. In some cases, I also limit the number of trials administered to apes to make contingency learning within the context of the experiments unlikely. When a larger number of trials is desirable, I also examine participants' performance in initial trials to see if the pattern of responding in those trials is consistent with the overall pattern. Further, I use novel situations and

stimuli that the apes have not experienced before, to minimize the effects of prior learning outside the context of the study. Povinelli and Vonk (2003) have argued that such an approach is more likely to be successful in detecting any mentalizing capacities apes might have than approaches using highly familiar stimuli and situations. This is because it forces the apes to think about how to respond rather than just responding automatically or relying on what has worked in the past.

I should also note that the particular mentalizing capacities I investigate here are not as vulnerable to explanation in terms of behavior reading as those studied in many earlier studies. As mentioned above, states such as wanting and seeing have clear behavioral markers associated with them—but this is not so for false beliefs. Indeed, this is part of what makes false belief tests, such as the ones I present in Studies 1 and 2 of this thesis, such convincing tests of *mind* reading. Furthermore, it is difficult to explain how apes might use behavioral cues to solve tasks requiring an understanding of their own mental states, which I investigate in the second part of my thesis. In Study 3, for example, I point out that participants cannot rely upon the gaze cues of others to solve the task, as they could in previous visual perspective taking tasks. Some people argue that when nonhuman animals respond as if attributing knowledge and ignorance states to themselves (e.g., Foote & Crystal, 2007; Hampton, 2001; Smith et al., 1995), they could just be responding to their own bodily cues associated with certainty or uncertainty (Carruthers, 2008). However, this explanation is more difficult to apply to situations in which an individual's false perceptions are pitted against his or her own knowledge, as we will see in Study 4 of this thesis.

Finally, like many investigators (e.g., LeDoux, 1996; Panskepp, 1998), I have decided not to overly concern myself with the issue of absolute (hence, unattainable) philosophical certainty. Almost all of scientific investigation involves some degree of ambiguity, but that is no reason to throw our hands up in defeat; we can still look for evidence highly suggestive of mental-state attribution in apes. In the words of Shettleworth and Sutton (2006):

The best we can do in such investigations is to define rigorously the behaviour accompanied by a given mental process and see if the animals show it. If the process is one usually assessed by verbal report, we will never be able to have the same kind of evidence for it in other species as in humans. (p. 235)

Thus, even if we cannot know with complete certainty that the same underlying mental mechanisms are at work, we can at least determine if there is functional similarity in behavior across species—i.e., if humans and apes respond similarly under similar conditions (Hampton, 2001). If they do, and obvious low level explanations are effectively ruled out, then we can be reasonably confident that some degree of mental state understanding is present. In any case, even if functionally similar responding in apes does not reflect true mental-state attribution, it is nevertheless interesting to demonstrate because it could constitute the basis of later evolutionary developments leading to mindreading proper. Now let us move on to Chapter 3 and see what the past 30 years of mentalizing research have (and have not yet) told us.

CHAPTER 3 – RESEARCH TO DATE

An extensive amount of research has been carried out on human children's mentalizing capacities. This research has inspired many of the methods devised for studying mentalizing in apes. Because of the sheer number of child studies, I review here a selection that is both representative of the larger body of experimental work and most pertinent to my own research. The experimental literature on mental-state attribution in apes is far less extensive and also more directly relevant to my own work, so I review that in somewhat more detail. Within each section, the literature on children is presented first followed by the literature on apes. I focus specifically on research investigating children's and apes' understanding of the perceptual state of seeing and how it relates to knowing and believing. The bulk of the existing literature, especially for apes, deals with their basic visual perspective taking capacities, and so I begin there. These capacities are likely foundational in developing the more advanced capacities I review next: the ability to recognize false beliefs in others and to distinguish appearance from reality. My goal in presenting this review is to pinpoint the areas in which our knowledge of apes' capacities is lagging far behind what we know about human children. My own experiments, presented in Chapters 4 through 6, begin to close these gaps.

Visual Perspective Taking: Understanding Seeing and Its Relation to Knowing

One of the most basic forms of mental-state understanding is recognizing that others, like oneself, visually perceive objects and events in the world around them, and recognizing the conditions under which they are, or are not, able to do so. Researchers have investigated a variety of abilities in attempting to discover the precise nature and

extent of visual perspective taking capacities in both children and chimpanzees. These abilities, which I review below, include gaze following, simple perspective taking, and recognizing the role of seeing in knowledge formation.

Children

Gaze Following and Simple Perspective Taking

At the most basic level is gaze following or “looking where someone else is looking” (Butterworth, 1991, p. 223). Human infants less than a year old have been observed to reflexively reorient themselves in the direction of another’s gaze (D’Entremont, Hains, & Muir, 1997). Later in development, however, gaze following behavior becomes elaborated in ways that indicate an emerging understanding of its role in visual perception. By 18 months of age, human infants follow gaze past distractor objects and to locations outside of their own visual field (Butterworth & Jarrett, 1991), and they are significantly less likely to follow the gaze of somebody whose eyes are closed or covered (Brooks & Meltzoff, 2002; Caron, Butler, & Brooks, 2002). They are also less likely to follow the gaze of an experimenter when an opaque (but not a transparent) barrier is in her line of vision to an object (Dunphy-Lelii & Wellman, 2004), and they will move around an opaque barrier to see what an experimenter is looking at when their own view is blocked (Moll & Tomasello, 2004).

These latter findings begin to cross into the realm of what Flavell (1977; 1992) called Level 1 perspective taking: recognizing what others are, or are not, able to see given their particular view on a scene. Level 1 perspective taking includes understanding

how opaque occluders can block visual access to objects, so that others may not be able to see things that one can see oneself, and vice versa. Several lines of evidence indicate that Level 1 perspective taking abilities are well developed by early preschool age (Flavell, 1992). For example, 2.5-year-olds could place an object behind an opaque barrier so that a person seated opposite them was not able to see the object (Flavell et al., 1978; McGuigan & Doherty, 2002). In another study by Moll and Tomasello (2006), 2-year-olds inferred that an experimenter who was looking for an object must have been looking for the one hidden from her view (but not from the child's view) by an opaque occluder, rather than for an object that was out in the open and thus visible to both.

The Role of Seeing in Knowing

Moving on from simple visual perspective taking, we can ask how much young children understand about the role of visual perception in knowledge formation. Do they recognize that seeing something in the past leads to knowing about it in the present, even after the object is no longer visually accessible? And do they understand that others who do not see what they themselves see will, as a consequence, not know what they themselves know? A common paradigm for testing such capacities has been to allow, or not allow, someone to peek into a container at a hidden object and to then ask children if that person knows what is inside the container. Children are also sometimes asked to judge their own knowledge states after peeking or not peeking inside the container. Pillow (1989) and Pratt and Bryant (1990) found that 3- to 4-year-old children were successful using this paradigm; however, Wimmer, Hogrefe, and Perner (1988) found that 5-year-olds, but not 3- to 4-year-olds, could succeed on this task. Pratt and Bryant

(1990) hypothesized that the discrepancy in findings was probably due to the younger children in Wimmer et al. (1988) being confused by the “double-barreled” nature of the questioning used in that study (e.g., Does she know what is in the box or does she not know?). Such possibilities underline the importance of developing nonverbal tests not just for apes but also for young children.

Another common paradigm involves hiding an object while someone either observes or does not observe, and then seeing whether children can accurately judge that person’s knowledge/ignorance of the object’s location. O’Neill (1996) found that 2-year-olds gestured less often to a hidden toy when their mother had witnessed the hiding event than when she had not because she was out of the room or had her eyes covered. However, when Dunham, Dunham and O’Keefe (2000) included a sham condition in which the child’s mother covered her eyes at some point in the trial but not during the actual hiding event, 2-year-olds continued to respond as if their mother did not know where the toy was hidden. At 2.5 years of age, children showed a better understanding of their mother’s knowledge state, gesturing more in the unseen-hiding condition than in both the seen-hiding and sham conditions. Povinelli and DeBlois (1992) tested 3- to 4-year-olds’ ability to infer others’ knowledge states using their well-known *Guess-er-know-er* paradigm. One experimenter left the room while another experimenter hid a treat among an array of containers (and children could not see where she hid it). The second experimenter then returned and each experimenter pointed to a different container. The older children, but not the younger ones, more often chose the container indicated by the knowledgeable experimenter.

At younger ages, there may be some disjunction between children's ability to do the tasks correctly and their ability to verbally report on their knowledge. For example, in the study by Pillow (1989) described above in which children peeked into a box at a hidden object, 3-year-old children had difficulty saying *how* they knew what was in the box. Similarly, Povinelli and DeBlois (1992) found that when an experimenter hid a treat under one of two cups and then pointed to the correct cup, both 3- and 4-year-old children searched under the correct cup but 3-year-olds could not explain how they knew where to look, even when they witnessed the hiding. Such findings may indicate different levels of knowledge in younger children than in older children, as discussed in Chapter 2.

Children's understanding that visual information has to be not just present but also adequate to engender knowledge appears to develop a little later than the ability to judge others' knowledge states based on past visual access. Such knowledge would include, for instance, knowing that a person cannot be sure of an object's identity if only a portion of the object is visible (Flavell, 2004). Taylor (1988) found that 4- to 6-year-olds believed that somebody who had access to only part of a drawing knew the identity of the animal depicted in the drawing—although the children realized that other details would not be known, such as the animal's name or what activity it was engaged in. Children older than 6 years of age understood that the partial view was not enough to identify the objects depicted.

Primary and Modality-specific Knowledge Assessment

O'Neill, Astington, and Flavell (1992) make an important distinction between primary knowledge assessment and modality-specific knowledge assessment. Primary

knowledge assessment is the ability to determine whether or not others know about the presence or location of objects depending on whether or not they have had perceptual access to them at some point in the past. Modality-specific knowledge assessment involves determining whether or not others know some specific quality of an object because they have had perceptual access to it in the appropriate modality (for example, knowing that an object is soft because one touched it rather than smelled it). Thus, we can ask whether children, beyond understanding that one must perceive to know and that the information perceived must be adequate, also understand that different types of sensory experiences support different kinds of knowledge. O'Neill and Chong (2001) asked 3- to 4-year-olds to perform the correct action for learning about some property of an object, such as its color or scent. Children were also asked to indicate on a Mr. Potato Head doll what sensory organ he would need in order to learn about the property in question. Finally, the children were asked to select which of several dolls had the correct sensory organ for discovering various properties of objects (each doll had only one sensory organ). Four-year-olds performed significantly better than 3-year-olds on all the tasks, suggesting that a marked improvement in modality-specific knowledge assessment occurs between these ages.

In summary, children's understanding of visual perception follows a clear developmental progression within the first few years of life. Reflexive gaze following in infancy gives way to more elaborated forms of gaze following after age 1 and to simple perspective taking by about 2 to 2.5 years of age. In the next year or two, children begin to understand the relation between seeing and knowing, and by about 4 years of age they

recognize how vision and the other sensory modalities engender different types of knowledge. Let us now examine what our nearest primate relatives understand about visual perception.

Chimpanzees

Gaze Following and Simple Perspective Taking

The evidence for even the simplest visual perspective taking capacities in chimpanzees has been somewhat mixed. However, the overall picture suggests that mature chimpanzees understand about at least as much as 3-year-old children. As with nonhuman primate research in general, most of the research on gaze following and simple perspective taking in chimpanzees has focused mainly on adults, making it difficult to comment on the developmental course of their abilities. The only two studies that I am aware of on gaze following in infant chimpanzees have produced contradictory results. Okamoto et al. (2002) found that a 13-month-old chimpanzee was capable of following an experimenter's gaze left or right to one of two objects. In contrast, Tomasello, Hare, and Fogleman (2001) found that three infant chimpanzees (1.5 to 2.5 years of age) did not spontaneously follow an experimenter's gaze upward into the sky, although older infants (3- to 4-year-olds) did. Possible reasons for these discrepant findings include that in the Okamoto et al. (2002) study (1) the chimpanzee was trained to follow gaze over several months, and (2) the experimenter looked at an actual object in the chimpanzee's field of view. In any case, the very small number of infants tested in each study makes it difficult to draw any conclusions about chimpanzees' gaze following abilities at this age.

By the time they reach adulthood, however, chimpanzees display the same sophisticated gaze following abilities observed in 2-year-old children. They follow the gaze of both humans and conspecifics (i.e., members of their own species) past distractor objects to locations outside of their own visual field, check back with the gazer when they find nothing interesting at the target location, and discontinue following the gaze of an individual who repeatedly gazes at nothing (Bräuer, Call, & Tomasello, 2005; Call, Hare, & Tomasello, 1998; Itakura, 1996; Povinelli & Eddy, 1996a, 1997; Tomasello, Hare, & Agnetta, 1999; Tomasello et al., 2001). These behaviors suggest that chimpanzees expect others' gaze to have some content, or external referent.

Findings from experimental studies on gaze following in the presence of visual occluders also provide evidence that apes are capable of simple visual perspective taking. In a study by Povinelli and Eddy (1996a), when an experimenter looked toward an opaque partition, chimpanzees rarely followed her gaze to the back wall of the room where it would have landed if the partition were absent. More often, they moved around to look at the other side of the partition, in apparent recognition that the experimenter's sight could not project through and beyond it. Tomasello and colleagues (1999) obtained similar results in a task involving four visual barriers, each requiring chimpanzees to move to a different location in an enclosure to see what an experimenter was gazing at. When the experimenter gazed into a gutter in the ground, for example, chimpanzees had to approach the gutter, bend down, and peer into it from above. A board positioned just outside the enclosure required chimpanzees to move to the front and far side of the enclosure to see around the board, or to climb upward to see over it. Two different walls

in separate parts of the enclosure served as the other two barriers. The chimpanzees performed the action required to see around the barriers significantly more often in experimental trials than in control trials, in which the experimenter gazed at a top corner of the cage. Bräuer et al. (2005) conducted a similar study with all four great ape species (chimpanzees, bonobos, gorillas and orangutans) and obtained comparable results. Gaze following has also been heavily studied in other primate and non-primate species, with mixed findings depending on the species and the particular skills tested (see reviews in Emery, 2000; Gómez, 2005; Johnson & Karin-D'Arcy, 2006).

Other studies have examined whether chimpanzees understand when others can or cannot see *them* (as opposed to some other object). Positive evidence has been found in observational studies. In a study of 20 juvenile chimpanzees (Tomasello et al., 1997), for example, the apes produced both visual- and auditory-based gestures significantly more often when recipients were looking at them than when they were not. These “audience effects” were also observed by Liebal, Call, and Tomasello (2004), who noted that chimpanzees often moved in front of a recipient before producing a visually based gesture. Pika, Liebal, Call, and Tomasello (2005) extended these findings to all four great ape species and one lesser ape species, siamangs (*Symphalangus syndactylus*). In that study, 80–90% of the time that visual gestures were produced, the recipient was oriented toward the gesturer, whereas 10–20% of the time the recipient’s back was turned.

In summary, observational studies suggest that chimpanzees have a good appreciation of when others can or cannot see them. However, Povinelli and colleagues (Povinelli, Bierschwale, & Cech, 1999; Povinelli & Eddy, 1996a; Reaux, Theall, &

Povinelli, 1999) found conflicting results using a *Begging* paradigm that took advantage of chimpanzees' natural begging gesture (i.e., arm extended with palm upward). In a series of studies, they allowed chimpanzees to choose between two experimenters when begging for food, one who could see them and one who could not (because she was turned away, had her eyes closed, was wearing a blindfold, etc.). The chimpanzees more often chose to approach the experimenter who had her body oriented toward them rather than the one who was oriented away. However, they had to learn over trials to approach the experimenter whose face and eyes were directed toward rather than away from them and were uncovered rather than covered. The researchers concluded that the chimpanzees showed no evidence of understanding which experimenter could see them and which could not. They suggested that even when the chimpanzees did eventually become successful at the task, they were using stimulus-based rules they had learned throughout the experiment, such as *approach the experimenter whose eyes are visible*. This low-level model was distinguished from the high-level model, which would include the additional element of representing the experimenters' perceptual state—in this case, seeing. As a final test of the low-level model, the chimpanzees were again brought into the room, this time with one of the experimenters facing forward but looking over her shoulder away from the ape, and with the other experimenter facing backward with her head turned over her shoulder to look directly at the ape. As predicted by the low-level model, the chimpanzees approached the forward-facing experimenter (who could not see them) more often than the backward-facing experimenter (who could). Povinelli and colleagues argued this was strong evidence *against* visual perspective taking in the chimpanzees.

Behavior reading, they argued, was not only just as good an explanation of the chimpanzees' behavior, but it was a better one, given their pattern of responses.

In short, Povinelli and colleagues have found mostly negative results with regard to chimpanzees' understanding of what others can or cannot see. Their conclusions have been criticized, however. First, chimpanzees' apparent inability to understand the role of the *eyes* in seeing does not mean that chimpanzees do not infer seeing. Eyes may have little meaning for chimpanzees as behavioral cues for visual attention, because they are not as salient a feature in the chimpanzee face as they are in the human face. The sclera is brown instead of white, and so they do not stand out clearly in contrast against the chimpanzee's brown skin and fur (Kobayashi & Kohshima, 2001). For this reason, chimpanzees may not take advantage of gaze cues to the extent that humans do, although it is clear that they can learn to pay attention to them if necessary. Thus, Povinelli and colleagues' experiments do not rule out the possibility that the chimpanzees were inferring the perceptual state of seeing; they may have simply been inferring it based on what we would consider to be a very crude cue (body orientation), and ignoring the more subtle behavioral cues that humans use to infer seeing. This interpretation is consistent with chimpanzees recognizing what other apes and experimenters can or cannot see when they are oriented in a particular direction, as in the observational and gaze-following studies described earlier. It is also consistent with findings by Kaminski, Call, and Tomasello (2004), who found that chimpanzees gestured to an experimenter more often when she was bodily oriented toward them rather than away from them and also (in the former case) when she had her face turned toward them rather than away. The

chimpanzees in that study also showed little sensitivity to whether the experimenter's eyes were open or closed.

A second criticism of Povinelli and colleagues' Begging tasks is that they required chimpanzees to interact with and infer the perceptual states of humans rather than other chimpanzees. It seems unlikely that this was the problem, given that chimpanzees and other apes have performed well in gaze-following paradigms even when the gazer was human (Bräuer et al., 2005; Povinelli & Eddy, 1996a; Tomasello et al., 1999). Nevertheless, to test chimpanzees' skills at inferring visual perceptual access in conspecifics, rather than humans, Hare et al. (2000) developed a new paradigm that involved pitting chimpanzees against one another in a competition for food. In what I will call the *Competition paradigm*, a dominant and a subordinate chimpanzee were placed in opposite cages, with a third cage between them. The cage in the middle contained two desirable food items, one out in the open where both chimpanzees could see it, and another placed behind a small opaque occluder such that only the subordinate could see it. The question was would the subordinate chimpanzee understand that the dominant chimpanzee had perceptual access only to the food out in the open? If so, then when both chimpanzees were released into the middle cage, the subordinate should avoid the item visible to the dominant and instead attempt to retrieve the item hidden behind the occluder. In most cases, this is precisely what subordinates did. Further manipulations ruled out the possibilities that subordinates were avoiding the food pursued, or gazed at, by the dominant, or that the subordinate was simply attracted to food near physical barriers. These manipulations included giving subordinates a head start into the center

cage, blocking the subordinate's view of the dominant before the subordinate was released, and putting the food behind a transparent rather than opaque occluder.³

Thus, the overall picture that emerges from past research is that chimpanzees do indeed have good Level 1 perspective taking abilities. The next question is whether they additionally understand the role of seeing in knowledge formation.

The Role of Seeing in Knowing

As with research into gaze following and simple perspective taking, research into what chimpanzees understand about the relation between seeing and knowing has produced contradictory findings. Using the same Guesser-knower procedure Povinelli and DeBlois (1992) used to test human children, Povinelli, Nelson, and Boysen (1990) failed to obtain convincing evidence that chimpanzees recognize what others know based on past visual access. As in the children's study, chimpanzees had to choose which of two containers held a treat by using the pointing gestures of two human experimenters. One experimenter witnessed the hiding event; the other did not because she was out of the room or had her head covered when it happened. If the chimpanzees understood that witnessing the hiding event led to knowing about the location of the treat, they should have chosen the container indicated by the knowledgeable experimenter. Although chimpanzees did eventually learn to succeed in the task, they did not do well in initial trials. This makes it impossible to conclude whether they had any true understanding of

³ Karin-D'Arcy and Povinelli (2002) failed to replicate this result, but Bräuer, Call & Tomasello (2007) suggest that important methodological differences (mainly the distance of the two pieces of food from one another) accounted for this.

the experimenters' knowledge states, or whether they instead had learned that paying attention to the gesture of the experimenter who was present with her head uncovered during the trial was more likely to result in a reward.

Of course, the same criticism applies to the Guesser-knower task as to the Begging task described earlier: chimpanzees were required to interact with and infer the mental states of humans, not conspecifics. Thus, Hare et al. (2001) extended the competition procedure to test whether chimpanzees showed evidence of understanding what other chimpanzees knew or did not know based on what they had or had not seen in the recent past. Dominant and subordinate chimpanzees once again competed for hidden food in a central cage between their two cages. The subordinate witnessed all the hiding events and therefore knew where all food items were located. The dominant, in contrast, was only allowed to witness the hiding of one piece of food but not a second piece. (In another condition the dominant watched as only one piece of food was hidden but then did not witness the experimenter subsequently move the food to another location). Subordinates more often approached food that dominants had not seen hidden or moved in the recent past and could therefore not know about in the present. The authors interpreted their results as being the first clear experimental evidence that chimpanzees represent states of knowing in others.

It is worth noting here that the manipulation in which the food was moved in the dominant's absence was strictly a knowledge-ignorance test and not a false belief test. To solve the task, subordinates only needed to recognize that the dominant, not having seen the experimenter place the food behind the second occluder, had no idea it was there.

They did not need to recognize that the dominant mistakenly believed the food to be in the other location. In other words, chimpanzees could have solved the task by attributing lack of perceptual access to a hiding event rather than by false belief attribution (see also Gómez, 2004, for a similar interpretation of a study by Gómez & Teixidor, 1992).

Others have also used chimpanzees' natural tendency to compete for food to test what they understand about their conspecifics' knowledge states. Hirata and Matsuzawa (2001), working with pairs of chimpanzees, allowed one chimpanzee to witness a food item being hidden in one of several containers in an outdoor enclosure. The second chimpanzee of the pair could not see where the food was being hidden, but could see that the first chimpanzee witnessed the hiding. In several cases, displaying the kind of Machiavellian intelligence Byrne and Whiten (1988) described, the knowledgeable chimpanzee misled the naïve one by taking a route to an empty container. Additionally, naïve chimpanzees sometimes developed strategies to forestall knowledgeable ones, such as following and intimidating the knowledgeable chimpanzee so that she would not retrieve the food, and then searching for the food themselves. Menzel (1974) described similar tactics and counter-tactics in a study carried out years earlier with a different group of chimpanzees.

Conspecifics or Competition?

Can we conclude, then, that chimpanzees are capable of inferring the perceptual and knowledge states of conspecifics but not humans? Given how much their performance improved when interacting with conspecifics, this would seem to be a reasonable conclusion. Alternatively, perhaps the crucial element in the competitive

paradigms described above was not the conspecific, but the competition itself (Hare, 2001). That is, earlier studies required chimpanzees to not just interact with humans, but to also respond in ways that would be seem logical and rational to humans. Chimpanzees forage competitively, not cooperatively, and they rarely share information about the location of monopolizable food sources. Further, they do poorly in even very simple tasks requiring them to make use of intentionally communicative cues such as pointing (Call, Agnetta, & Tomasello, 2000; Povinelli, Reaux, Bierschwale, Allain, & Simon, 1997; Tomasello, Call, & Gluckman, 1997). Having a cooperative experimenter voluntarily indicate the location of a desired, hidden food item, as in the Guesser-knower task, may be outside their realm of experience or understanding.

To explore this possibility, Hare and Tomasello (2004) directly tested the effects of using a cooperative versus a competitive paradigm (for convenience, I will call their test the *Pointing-vs.-reaching* test). They found that chimpanzees performing an object-choice task could not consistently choose the correct of two containers (i.e., the one in which food was hidden) when a cooperative experimenter pointed to it. However, they were able to locate the food when a competitive experimenter reached for the correct container with the apparent intention of taking the food. The two experimenters also acted differently toward the chimpanzees to further establish the cooperative or competitive context. The cooperator was encouraging and acted pleased when chimpanzees found the food, whereas the competitor acted angry and frustrated. The gestures used by the experimenters were almost identical in both cases: arm fully

extended with index finger jutting outward. The only variable was the context (cooperative or competitive), and it made a clear difference.

Hare et al. (2006) thus ran another paradigm involving competition, again with a human competitor. They had an experimenter sit in a booth between two windows, one on his left and one on his right. Chimpanzees could choose to approach either window to take a piece of food placed there. Because the experimenter was a competitor, however, if he saw the chimpanzee approaching he would quickly take the food away before the chimpanzee could retrieve it. Chimpanzees selectively avoided the window the experimenter had his face and eyes turned toward (even when he had his body turned the other way—i.e., he was looking over his shoulder). When the experimenter was facing forward, chimpanzees also selectively approached the window he could not potentially see through because it was blocked by an opaque barrier. Furthermore, they sometimes made an indirect approach to the occluded window, but only in conditions in which it would be effective (namely, when the experimenter could see them leaving the area but could not see them returning from another angle to retrieve the food). Most interesting was that chimpanzees appeared to be trying to hide their attempt to hide. The authors suggested that such “active concealment” indicated a deliberate attempt to block the competitor’s informational access to their activities.

Outstanding Questions

There is ample evidence to suggest that chimpanzees, at least in situations that make sense to them, possess good Level 1 visual perspective taking skills. Far less is known about what chimpanzees understand about their own visual perceptual

experiences. In the only published study on the topic prior to my own research, Call and Carpenter (2001) manipulated the visual experiences of chimpanzees, orangutans, and 2-year-old children by either allowing or not allowing them to watch as a reward was hidden among an array of three opaque tubes. Participants were then given the opportunity to look into the tubes before choosing. All three species responded differentially across conditions by looking into the tubes significantly more often when they had not witnessed the hiding than when they had, thereby demonstrating that they discriminated between their own states of knowledge and ignorance about the location of the reward. Call (2005) replicated this finding and extended it to bonobos and gorillas. In Call and Carpenter (2001), when apes did look into the tubes, they used efficient search strategies more often than inefficient ones: they looked into each tube only once; stopped looking as soon as they had found the reward; and in many cases they chose the third tube without first looking into it, if they had already looked into the other two tubes and found nothing. Interestingly, children engaged in more unnecessary searching than the apes did by often continuing to search even after they had already spotted the reward.

Call and Carpenter (2001) claimed that their results were “consistent with the hypothesis that apes knew when they did not know where the reward was, and that, when necessary, they could act to obtain this knowledge” (p. 218). While acknowledging that there could be lower-level explanations for their findings, they suggested that, for various reasons, these explanations were unlikely. For example, one might argue that the apes acquired the looking response through trial and error learning; however, most participants began looking into the tubes from the earliest trials. Furthermore, participants who began

looking only in later trials showed a response pattern more indicative of insight learning, shifting suddenly from not looking at all to looking in every trial. Associative learning explanations are also conceivable for even the more complex patterns of responses participants displayed (such as choosing the third tube without looking into it, if they had already looked into both other tubes and found nothing). One could argue, for example, that apes simply learned the following rule: *if the first two tubes are empty then choose the third tube to get a food reward*. The authors point out, however, that such an approach would be difficult for apes because it would require them not only to learn compound cues (first tube empty plus second tube empty) but also to make a spatially incompatible response (choose the tube *not* associated with the cue). Finally, one could argue that apes did not understand what they did or did not know, but only what they had or had not seen in the recent past. Once again, however, this explanation is not consistent with the finding that apes chose tubes they had not looked into when they could infer from exclusion that the food was there. Given the overall pattern of results, the authors concluded that their task “may involve understanding of at least one’s own perception, and perhaps one’s own knowledge” (p. 219).

Nevertheless, Call and Carpenter’s (2001) study was limited in a number of other respects. One is that it is difficult to know whether the response measure they used (crouching to look into tubes) reflected deliberate, calculated attempts on the part of the apes to gather visual information. Carruthers (2008) points out that simple information gathering behaviors such as pausing, peering more closely, and moving one’s head from side to side are common in animals during moments of uncertainty. He argues, however,

that these behaviors most likely reflect automatic mechanisms that evolved because they were adaptive. A demonstration of greater complexity and flexibility in apes' responses during their attempts to visually locate hidden food would thus be more convincing.

Additionally, Call and Carpenter's (2001) study did not test apes' understanding of the importance of *particular* visual perspectives, because the food in their study could be viewed from only one vantage point: through the front of the tube. Beyond having a rough sense that occluders block vision, good visual perspective taking would also require one to understand specifically where occluders have to stand in spatial relation to objects to be effective. For example, occluders positioned to the side, above, or below an object would not block the view of someone seated directly opposite the object. It is not possible to tell from the research done so far whether chimpanzees understand this. For example, Karin-D'Arcy and Povinelli (2002, Experiment 6) found that subordinate chimpanzees, when put into competition for food with dominants, preferred food positioned to the side of an occluder (and therefore not blocked from the dominant's view) to food that was not located near any occluders. Subordinate chimpanzees may prefer to forage near structures because they have learned from experience that it is safer than foraging in open spaces (Karin-D'Arcy and Povinelli call this the "peripheral feeding" hypothesis). However, it is also conceivable that the chimpanzees thought the dominant could not see the food because there was a barrier positioned directly next to it. Other studies investigating chimpanzees' visual perspective taking abilities with regard to others described earlier also cannot tell us whether they understand the relation between particular visual perspectives and seeing. For example, in Hare et al.'s (2000)

Competition paradigm, subordinate chimpanzees had to understand very little about specific angles of view in order to judge whether dominants could or could not see the occluded food. They only needed to understand that dominants could not see the food if a solid, opaque object intervened somewhere in the general area between the dominant and the food. Further, in the studies in which apes needed to move around various visual barriers to see what an experimenter was gazing at (Bräuer et al., 2005; Tomasello et al., 1999), the apes were judged as having made the correct response if they moved to the correct position at any point within a trial lasting 1 minute, irrespective of where they had moved previously. It is therefore impossible to tell if they anticipated the correct angle before moving. Additionally, in those studies there was a social cue available for the apes to use: the gaze of the experimenter. Apes could have thus extrapolated the experimenter's line of sight from her face direction and used this to help them determine where to move. In fact, all of the visual perspective taking studies involving social partners necessarily had gaze cues available that apes could use to solve the tasks, without truly understanding the importance of particular visual perspectives in a more general way. I address all of the above limitations in Chapter 5 when I investigate chimpanzees' visual perspective taking capacities using a task in which they must judge what they themselves would or would not be able to see from various perspectives.

I noted earlier that visual perspective taking abilities are positively associated with, and likely foundational to, more advanced mentalizing capacities. These include the ability to recognize false mental states in both oneself and others, as measured by false belief and appearance-reality tests. To reiterate, the positive relations between

understanding of visual perspective, false belief, and appearance versus reality are likely due to the fact that all of these require an understanding that one thing can be simultaneously represented in two different ways. However, false belief and appearance-reality understanding are more advanced because they additionally require the understanding that one of those two representations can conflict with what is truly the case. Further, an understanding of the relation between seeing and knowing may lead to recognizing how *not* seeing some important event can result in false beliefs. And recognizing that visual perceptions may sometimes be at odds with what one knows to be the case (e.g., visual illusions) would seem to require an advanced understanding of visual perception and its relation to knowing.

In the remainder of this chapter, I review the literature on the two advanced mentalizing capacities investigated in this thesis, beginning with false belief understanding and then moving on to the appearance-reality research. As with research on visual perspective taking, a great number of studies have examined human children's false belief and appearance-reality understanding. As we will see, however, far less has been done in these areas with apes.

False Belief: Understanding the Effect of *Not* Seeing on Belief States

Children

Standard Verbal Tasks

There are many ways to acquire a false belief. One way is to fail to witness an important event that causes some state of the world as one knows it to change. This the

logic behind the standard Change-of-location (Sally Anne) test described in Chapter 1, in which an item is displaced to a new location in someone's absence (Baron-Cohen et al., 1985). Another way that one might come to have a false belief is by viewing an object that is perceptually misleading. This describes the standard Unexpected contents (Smarties) test in which someone encounters a candy box that actually contains crayons (Perner et al., 1987). Over the past quarter of a century, a large body of research has established that normally developing children begin to pass standard tests of false belief understanding between about 4 and 5 years of age, depending on the population studied and some specific features of the tasks used. Most notably, these features include whether children are allowed to actively participate in the action leading to the false belief, and whether the task involves explicitly deceptive elements (Wellman et al., 2001). These manipulations are discussed in more detail below.

Non-standard Manipulations for Younger Children

While children younger than about 4 years of age consistently have trouble with standard verbal false belief tests, they have performed well in some non-standard manipulations. Adding an element of deception appears to make a big difference. For example, Sullivan and Winner (1993) found that older 3-year-olds who failed a Smarties test nevertheless passed it when it was turned into a game of trickery in which the children participated by switching the contents of the box. In two other studies, some 3-year-olds were capable of intentionally deceiving someone about the location of a treasure even when they had to figure out how to do so themselves. For example, they wiped away footprints leading to the actual location of the treasure and added footprints

leading to another, incorrect location (Chandler, Fritz, & Hala, 1989; Hala et al., 1991; but see also Sodian, Taylor, Harris, & Perner, 1991).

Younger children have also performed better in false belief tests involving active helping responses (Carpenter et al., 2002; Matsui & Miura, 2008). For example, in Carpenter et al.'s (2002) task for 3-year-olds, one experimenter (Experimenter 1) placed Item A into a box and left the room, whereupon a second experimenter (Experimenter 2) surreptitiously switched it for Item B. Experimenter 1 then returned and struggled briefly, and unsuccessfully, to open the box in an attempt to retrieve Item A, which she believed was still inside. She then became distracted and absorbed in some other activity. While Experimenter 1 was not attending, Experimenter 2 removed Item B from the box and placed it, along with Item A, on a chair across the room. A short time later, Experimenter 1 noticed both objects on the chair and asked the child to help her "get it." Between one- and two-thirds of children responded in way that suggested they took account of Experimenter 1's earlier false belief about the box's contents, bringing her Item A, which she had originally placed into the box.

A number of researchers have also addressed the possibility that younger children fail the standard verbal tasks not for lack of understanding, but because their grasp of language is not well developed enough to deal with the high linguistic demands of standard tasks. This was one of the motivations behind Carpenter et al.'s (2002) helping task described above. Three-year-olds also performed better in a false belief test by Carlson, Wong, Lemke, and Cosser (2005) that focused on gestures rather than verbal responses. It has also been suggested that younger children may have difficulty with

standard tasks because of a prepotent response bias (Moses, 2005; Wellman et al., 2001). That is, they may know the correct response but are unable to inhibit a strong competing response tendency. In the Change-of-location test, for example, children might be compelled to indicate what they know to be the true location of the hidden item, even while understanding that the story character thinks the item is in the other location. Manipulations in which this ‘pull of the real’ have been removed (for example by having someone eat the hidden treat) have elicited a greater number of correct responses from younger children (e.g., Koós, Gergely, Csibra, & Biro, 1997).

Thus, while it is a robust finding that children do not typically begin to pass standard verbal tasks until about 4 to 5 years of age (Wellman et al., 2001), non-standard manipulations have uncovered apparent false belief understanding in children as young as 3 years old. In the following section, I discuss evidence suggesting that children even younger than this may be capable of recognizing false beliefs on some level.

Spontaneous Looking Measures

Researchers have also looked for evidence of false belief understanding in children younger than 4 years old by examining their spontaneous behaviors while performing a task. As we saw in Chapter 2, spontaneous looking has become an increasingly popular measure because of its potential to tap into understanding not evident in active, elicited responses. To reiterate, younger children may have some inkling of what the correct response in a false belief task is but be too uncertain in their knowledge to act on it. It is also possible that such children could, on some level, detect the false belief without any awareness of having done so. Such understanding might not

be strong enough to drive children's active responses but could nevertheless show up in spontaneous behaviors, such as looking at the correct response. While there is little consensus about the extent to which such spontaneous behaviors (and indeed even active responses) reflect implicit versus explicit knowledge, many consider them to be evidence of understanding false beliefs on some level. The looking studies briefly mentioned in Chapter 2 are thus described in more detail below.

Clements and Perner (1994) and Garnham and Ruffman (2001) presented children with a standard Change-of-location false belief task in which an object is moved in a story character's absence and the character later returns to retrieve it. In addition to asking children to indicate (by stating or pointing) where the story character would look for the object, the experimenters also coded anticipatory looking as the character was about to return to the scene. That is, just before the character returned to retrieve the object, the experimenter wondered aloud which location the character would go to. Many 3-year-olds correctly looked toward the original location of the object (where a character with a false belief would be expected to go), even though they went on to wrongly indicate its current location in their active, elicited response. Thus, children's anticipatory looking responses betrayed knowledge of the character's false belief that was not evident in their active choices. Ruffman et al. (2001) also found that both 3- and 4-year-olds showed correct anticipatory looking in a similar false belief test in which they made incorrect choices. Interestingly, the 4-year-olds were less confident about their incorrect choices than the 3-year-olds, suggesting that the 4-year-olds were moving into another level of understanding of false beliefs but were still too uncertain in their knowledge to

act on it. The authors suggested that this transition was complete by 5 years of age, as those children responded correctly and also showed high levels of confidence in their choices.

Recently, several researchers have used spontaneous looking measures to test for false belief understanding in even preverbal infants. To avoid the need for any degree of linguistic understanding in their participants, they devised fully nonverbal versions of the Change-of-location procedure in which infants simply watched a sequence of events unfold. Southgate et al. (2007) found that children as young as 2 years old showed anticipatory looking toward where an experimenter would be expected to search for a toy moved in her absence. The authors argued that their results show that 2-year-old children possess a level of false belief understanding equivalent to that of older children, with performance limitations leading to failures in standard tasks.

Violation-of-expectancy (VOE) looking-time methods provide another way of using spontaneous looking behavior to investigate the mentalizing capacities of preverbal infants. The VOE method involves allowing children to witness two different events, one that meets normal expectations about what should occur in that situation and another that violates normal expectations. The amount of time infants spend looking at each event is measured. If they do indeed understand what is to be normally expected in the events taking place, they should look significantly longer at the expectancy-violating event than at the other event. Onishi and Baillargeon (2005) used the VOE method to investigate false belief understanding in 15-month-old infants. An experimenter hid an object in one of two locations as the infant observed. The object was then moved to the other location either (1)

while the experimenter was not looking (the *False belief* condition) or (2) while she observed (the *True belief* control condition). In both cases, the experimenter then searched for the object in its actual location. Infants looked significantly longer at the experimenter's searching in the False belief condition than in the True belief condition, presumably because they understood on some level that the experimenter should not have known the object was there. Using a similar test, Surian, Caldi, and Sperber (2007) found evidence that even 13-month-olds may recognize others' false beliefs on some level.

To summarize, based on findings from hundreds of false belief studies using standard verbal tasks, it has long been accepted that children do not begin to recognize false beliefs until about 4 to 5 years of age. However, non-standard manipulations have pushed the age for actively responding correctly in false belief tests down to 3 years old; and recent studies using spontaneous looking measures suggest that even 1- to 2-year-old infants may recognize false beliefs on some level. Let us now look at what our closest primate relatives, at any age, may or may not understand about false beliefs.

Chimpanzees

Nonverbal Adaptations of Standard Verbal Tasks

While we have learned much about false belief understanding in children, it is still an open question whether it is unique to humans or is also present in apes. The main stumbling block has been the difficulty of adapting the highly verbal standard tests for non-linguistic participants. And because the goal has been to directly compare children and apes on the same task, it has been a challenge to devise procedures appropriate for

both species. Before this thesis, just two published studies, both of them comparative, had investigated false belief understanding in apes.

Call and Tomasello (1999) developed a nonverbal variation on the standard Change-of-location procedure to test chimpanzees, orangutans and 4- to 5-year-old children. In initial training trials, one experimenter (the “hider”) hid a reward in one of two identical containers while a second experimenter (the “communicator”) observed. Participants themselves could not see where the hider was placing the reward. The communicator then indicated the location of the reward for participants by briefly marking the correct container with a wooden block for 1 to 2 seconds. Once participants had learned to use the placement of the marker to find the hidden reward at greater than chance levels,⁴ the False belief trials were run. In these trials, immediately after the reward was hidden the communicator left the area, during which time the hider switched the positions of the containers. The communicator, who now had a false belief about the location of the reward, returned and marked the incorrect container. Five-year-old children indicated that they recognized the communicator’s false belief by choosing the unmarked container, but younger children and both ape species failed the test. A number of control conditions (including True belief control trials) confirmed that they did not fail for reasons unrelated to the false belief element of the task. The results were argued to be strong evidence that apes did not have any understanding of false belief.

⁴ Note that the apes took between 36 and 90 trials to reach this criterion. In contrast, children almost immediately recognized that the block was meant to indicate the location of the reward.

O'Connell and Dunbar (2003) used a similar procedure to test for false belief understanding in chimpanzees, 3- to 6-year-old children, and autistic adults.⁵ During the training portion of their task, an experimenter positioned a wooden peg above one of four drawers in an apparatus and then moved behind the apparatus to place a reward in that drawer through an opening at the back. Participants had to choose the correct drawer to win the reward.⁶ In subsequent False belief trials, a second person switched the position of the peg after the experimenter had moved behind the apparatus, leading her to have a false belief about the peg's position. In True belief control trials, the experimenter saw that the peg had been moved before she moved behind the apparatus to hide the reward. Results for the humans were that children of all ages and autistic adults performed better than chance in the True belief test but only 6-year-old children also passed the False belief test. (Note that the fact that 5-year-old children did not pass this test indicates that it was more difficult than standard false belief tests.) The four chimpanzee participants also exceeded chance performance in the False belief test but, importantly, they failed the True belief control test, calling the False belief results into serious question. The chimpanzees may have learned, for example, to choose the drawer associated with the initial position of the peg and then applied this rule in both the True and False belief

⁵ Many believe that mindreading is impaired in autistic individuals (e.g., Baron-Cohen, 1995).

⁶ As in Call and Tomasello (1999), chimpanzees took much longer than humans to learn this. For children, the passing criterion was set at 4 consecutive trials correct, and it took them from 4 to 8 trials to reach this criterion. For apes and autistic adults (because of their slower rate of learning), the criterion was set at 6 or more trials correct in a block of 10 trials. Chimpanzees took between 42 and 117 trials to reach this criterion, while autistic individuals took 20 trials (the authors did not provide a range in this case).

trials. To convincingly demonstrate false belief understanding, participants would have to show evidence of switching strategies across the True and False belief conditions, as the 6-year-old children did. Thus, to date, there has been no convincing evidence that chimpanzees are capable of recognizing others' false belief states.

Outstanding Questions

It is important to note that both Call and Tomasello's (1999) and O'Connell and Dunbar's (2003) studies used cooperative-communicative paradigms. In the review of the visual perspective taking literature, we saw that such paradigms may be very problematic for chimpanzees because they normally compete for food and have difficulties understanding cooperative-communicative intentions (Hare, 2001; Hare & Tomasello, 2004). Additionally, both Call and Tomasello (1999) and O'Connell and Dunbar (2003) measured only active, elicited responses: the participant's choice of container or drawer, respectively. However, spontaneous behaviors may indicate some level of false belief understanding not apparent in active responses, as suggested by the anticipatory looking responses of the 3-year-old children who failed Clements and Perner's (1994) task, for example. There are other possible reasons why active responses might not tell the whole story. First, if a competing response tendency is strong enough (i.e., a prepotent response), it may be difficult for participants to overcome it. There is some evidence that inhibitory issues of this sort can influence apes' responding in experimental tasks (Boysen & Berntson, 1995; Boysen, Berntson, Hannan, & Cacioppo, 1996; Boysen et al., 2001), just as they can young children's (e.g., Carlson & Moses, 2001; Moses, 2005; Wellman et al., 2001). Second, participants may be too uncertain in their understanding

of false beliefs to act on it, as suggested by Ruffman et al.'s (2001) finding that 4-year-olds who failed a false belief task were less confident about their choices than 3-year-olds who failed the same task. Similarly, Suda and Call (2006) compared apes who were moderately successful in a liquid conservation task with those who were either highly successful or unsuccessful. The moderately successful apes displayed the most hesitancy in responding, for example by changing their choice or trying to choose both alternatives simultaneously. Other researchers have also noted such "ancillary behaviors" possibly indicative of uncertainty in animals faced with difficult decisions (Smith, Shields, & Washburn, 2003).

Thus, we cannot conclude on the basis of extant research that apes do not understand false beliefs. It is important to examine their false belief understanding with tests that are more species relevant for them, in the sense that they do not require them to cooperate for food and to understand communicative intentions. It is also important to measure apes' spontaneous behaviors during false belief tasks, to see if they might possess some level of understanding that is not showing up in their active responses. I do both of these in Chapter 4 with two new nonverbal false belief tests.

I next review the appearance-reality literature. Like false belief understanding, the ability to distinguish appearance from reality requires the capacity to recognize that mental states can be false. While most false belief tests have been designed mainly to investigate whether participants recognize others' false mental states, appearance-reality tests typically measure whether individuals recognize when their own mental states are mistaken. As with visual perspective taking and false belief understanding, there is a

large body of research on children's ability to distinguish appearance from reality. As before, I begin below by discussing this literature. However, while visual perspective taking and false belief understanding have also been studied in chimpanzees (although minimally, in the latter case), there have as yet been no direct tests of appearance-reality understanding in chimpanzees. Thus, for the chimpanzees, I discuss the somewhat relevant research on their ability to understand liquid conservation.

Appearance *Versus* Reality: Recognizing When Seeing Conflicts With Knowing

Children

Standard Verbal Tasks

As with recognizing false beliefs, distinguishing appearance from reality requires one to recognize that mental states can be wrong. In the case of false beliefs, this is someone else's mistaken belief (although children are sometimes also asked to comment on their own prior false beliefs). In the case of appearance-reality understanding, it is one's own mistaken perception. There is a rich literature on human children's understanding of the appearance-reality distinction, thanks in large part to the work of Flavell and colleagues throughout the 1980s (e.g., Flavell, 1986; Flavell et al., 1983; Flavell et al., 1986; Flavell, Green et al., 1987; Taylor & Flavell, 1984). As described in Chapter 1, the standard verbal tests involve presenting children with a perceptually misleading object (e.g., a sponge that looks like a rock, a piece of paper behind a color filter, or an object behind a magnifying lens) and then asking them to state the true and apparent identity or properties of the object (Braine & Shanks, 1965a, 1965b; Flavell et

al., 1983; Flavell et al., 1986; Flavell, Green et al., 1987). As with false belief tests, children do poorly on these standard tests of appearance-reality understanding until about 4 to 5 years of age. However, the picture is somewhat more complex than that, as the kinds of errors that younger children make depend on what kinds of questions they are asked about the objects, as we will see below.

Phenomenism and Intellectual Realism

In some earlier publications, Flavell (1963; 1977) suggested that young children who did not have a well-developed grasp of the appearance-reality distinction would be prone to focus on what was more perceptually striking—on aspects of surface appearance rather than on underlying reality. This would lead them to err in appearance-reality tests by giving appearance answers to both questions, for example saying a rock-sponge looked like a rock and also really was a rock. Such “phenomenism,” as Flavell called it, was documented in several early studies (Braine & Shanks, 1965a, 1965b; Elkind, 1966; Langer & Strauss, 1972; Murray, 1968). For example, in the two studies by Braine and Shanks (1965a; 1965b) children were asked questions about the apparent and real size or shape of items that were distorted by various means (e.g., by placing them in a glass of water or behind a magnifying or minimizing lens). Up to the age of 7 or 8, children tended to spontaneously report apparent size or shape as if it was real. When given corrective feedback, however, children 5 years of age and older could quickly learn to respond appropriately, suggesting that their failures before receiving feedback may have been due to misinterpreting the experimenter’s questions.

Other studies have, however, observed the opposite of phenomenism in young children, as they report reality when appearance is requested (called “intellectual realism”). Flavell et al. (1983) cite a number of studies, for example, in which children report, draw, or choose a photo depicting details of objects or scenes they know to be there, even if the children cannot see the details because they are currently blocked from their view. This occurs even when experimenters explicitly ask children to pay attention only to exactly what they see. Flavell et al. (1983) also observed intellectual realism errors in 3-year-olds given a standard appearance-reality test that involved a rock-sponge, a rubber eraser that looked like a pencil, a candle that looked like an apple, a doll that looked like a ghost when a handkerchief was placed over it, and a white index card that looked pink when placed behind tinted transparent plastic. Three-year-olds were able to answer correctly for the ghost-doll and pink-white card (perhaps because they could witness the transformation of the doll into a ghost and the white card into a pink one), but they had trouble with the other items. Four-year-old children performed well with all items, and 5-year-old children were almost without error. The errors typically made by the 3-year-olds included saying that the object looked like what it really was (e.g., saying that the rock-sponge looked like a sponge).

Further experiments by Flavell et al. (1983) and Taylor and Flavell (1984) revealed that realism errors are more common when children are asked about the real/apparent *identity* of an object, but phenomenism errors are more common when children are asked about the real/apparent *physical qualities* of objects, such as their size, color, or shape. For example, Flavell et al. (1983) changed the apparent size or color of

various objects (such as a rock painted to look like an egg) by displaying them behind tinted plastic or size-altering lenses. Children were then asked to state their real/apparent identity or their real/apparent size and color. Children made realism errors more often when asked about an object's identity, and they made phenomenism errors more often when asked about its physical qualities. Flavell (1993) concluded that 3-year-olds were more prone to both types of error than older children and that their difficulties with appearance-reality tasks were a robust finding over many studies. He and others (e.g., Gopnik & Astington, 1988; Taylor & Hort, 1990) have argued that younger children's difficulties with appearance-reality tests reflect a deep-seated representational inflexibility that makes them incapable of holding two conflicting representations of the same object in mind at the same time. Not everyone shares this view, however, as we will see in the next section.

Non-standard Manipulations for Younger Children

Some researchers have found the difficulty that 3-year-old children have with appearance-reality tasks to be surprising, given that children of this age and younger often incorporate pretense into play and can describe familiar objects as having pretend identities, such as describing a banana as being a 'pretend' telephone (e.g., Abelev & Markman, 2006; Lillard, 1993). In fact, as with false belief tests, 3-year-olds have succeeded in some appearance-reality tests when more natural questioning and nonverbal responses were used (Sapp et al., 2000), and also when information-processing requirements were reduced or deceptive elements were added (Rice et al., 1997). And even when 3-year-olds have performed poorly overall in a test, they have sometimes done

well in some conditions (see, for example, Study 1 in both of Flavell et al., 1983; Flavell, Green et al., 1987).

Sapp et al. (2000) hypothesized that young children have trouble with appearance-reality tasks because when faced with an unfamiliar object, they may not feel that they can apply two different labels to it, called the 'mutual exclusivity constraint'. To test whether some such linguistic factor might be causing problems for young children, these researchers developed nonverbal appearance-reality tasks that used the same objects employed in earlier studies (e.g., the rock-sponge), and they gave these to children along with the standard verbal tasks. Their nonverbal tasks, unlike previous verbal ones, did not require children to label or describe the objects, but only to respond appropriately to an experimenter's request. For example, the experimenter told children she wanted to take a picture of a teddy bear with something that looked like a rock, and she asked children to hand her something she could use from among an array of items. In another trial, she spilled water and asked children to give her something she could wipe it up with. In both cases the correct response was the rock-sponge. Three-year-old children did very well in the task, but only when they had not received the standard verbal task first. On the other hand, doing the nonverbal task first (and doing well on it) did not improve performance in the verbal task. The authors concluded that the verbal task was especially difficult for children, probably due to the mutual exclusivity constraint, and that it even confused them enough to interfere with their performance on a nonverbal version of the task that they otherwise performed well on. Sapp et al.'s (2000) study suggests that 3-year-olds understand more about the distinction between appearance and reality than earlier studies

using verbal paradigms suggest. Thus, as with false belief understanding, nonverbal tests of children's capacity to distinguish appearance from reality may better capture underlying competence that may not show up in verbal tasks due to non-relevant task demands.

Chimpanzees

Related Research

It is also important to test chimpanzees' understanding of the appearance-reality distinction, in order to gain a more complete picture of how their mentalizing capacities might or might not differ from humans'. However, up until now, chimpanzees' appearance-reality understanding has not been directly investigated. The closest thing to appearance-reality tasks carried out with apes to date are the Piagetian liquid-conservation studies by Woodruff and colleagues (1978), Muncer (1983), Call and Rochat (1996; 1997) and Suda and Call (2004) referred to briefly in Chapter 1. Liquid conservation tasks test participants' ability to recognize that the quantity of a liquid remains constant when transferred to a new container with different proportions. In the standard liquid-conservation paradigm used with children, an experimenter shows the child two identical glasses, each containing the same amount of liquid. As the child observes, the experimenter pours the contents of one glass into a relatively taller or shorter glass. The child is then asked if the glasses contain the same amount of liquid or if one contains more than the other. Children usually do not reliably pass this test until about 7 or 8 years of age (Suda & Call, 2004).

Woodruff et al. (1978) used a variation on this standard paradigm to test their one chimpanzee participant, Sarah, who had previously been trained to use tokens to signify 'same' and 'different'. As Sarah observed, an experimenter transferred liquid from one clear container into a second clear container of different proportions. Sarah was then required to indicate, using her tokens, whether the amount had changed. She succeeded in the task, and the fact that she could not make accurate judgments when simply shown two differently-shaped containers of liquid (i.e., without seeing the transformation) showed that she was not just estimating quantity by visual inspection. Muncer (1983), also testing only one chimpanzee, used a variation on the standard task in which the larger of two quantities of liquid was sometimes transferred into the smaller of two containers and sometimes into the larger of the two. In the latter case, this reversed the relation between the original and final levels of liquid, such that the container with the originally higher level had the lower level after the transfer was complete, and vice versa. The chimpanzee was successful in choosing the larger of the two rewards in both cases, and like Woodruff et al.'s (1978) Sarah, the chimpanzee did not succeed without visual access to the transformation.

Using a similar procedure as Muncer (1983), Call and Rochat (1996) found that three of their four orangutan participants could choose the larger amount of liquid after the transfer, but they determined that the orangutans were probably using the perceptual strategy of choosing whichever container had the higher level of liquid after the transfer. In most cases, using this strategy allowed participants to make the correct choice. In a follow-up study, Call and Rochat (1997) looked further into the three possible strategies

orangutans could be using to choose the larger reward: (1) visually estimating the quantities of liquid after the transfer, (2) watching how much liquid was being poured into each container during the transfer, or (3) noting which of the original quantities was larger and then tracking the contents of that container through the transfer. They concluded that the orangutans used mainly the perceptual strategy of visual estimation after the fact, which they called “pseudoconservation” rather than true conservation.

Suda and Call (2004) sought to resolve the discrepancy between the findings with chimpanzees and orangutans by testing both species (and also bonobos) using identical procedures. Their study, which included 14 apes, looked more closely than previous ones at how perceptual features interacted with identity to influence apes' choices. They varied whether the containers into which the liquid was transferred were the same or different in shape, and also whether participants could or could not see the end state of the liquid after it had been transferred (i.e., target containers were either clear or opaque). Apes did better when they could see the end state than when they could not, suggesting that they used after-the-fact perceptual estimation of the quantities to some extent. In addition, the apes were influenced by the shape of the containers, preferring the short, wide container to the tall, skinny one, in many cases even when the larger quantity of liquid was transferred into the latter. Thus, the apes were largely influenced by the appearance of the target containers and the liquid inside of them. However, Suda and Call (2004) point out, this does not mean that their apes made no use of logical strategies involving the identity of the liquid. If this were the case they could not have achieved higher-than-chance performance in most conditions, which they did. In addition, there were large individual

differences in ability, with the choice patterns of some participants in some conditions suggesting that they took identity information into account.

Outstanding Questions

While some chimpanzees had some degree of success in the liquid conservation tasks, it is difficult to know what to make of the findings. Woodruff et al. (1978) and Muncer (1983) tested only one chimpanzee each, making their positive results difficult to generalize. Suda and Call's (2004) study, which included 14 apes comprising three different species, was better in terms of participant numbers. However, in that study, differences in the shapes and sizes of the containers were confounded with the appearance of the reward, making it difficult to tell to what extent container preferences influenced apes' choices, independent of how the liquid looked inside the container. In addition, it is unknown whether the apes actually perceived the tall, narrow container as containing more liquid than the short, wide one, especially given that they tended to choose the latter. Perhaps apes are more influenced by width than height when judging the quantity of a substance. Alternatively, apes may have preferred the short containers because they were more similar to the original cups used to hold the liquid before the transfer event. In short, the extent to which apes appreciate the distinction between appearance and reality is still far from clear. In Chapter 6, I address this question with the first-ever appearance-reality test for apes.

Summary

Apes have displayed capacities equivalent to humans, at least on some basic levels, in tasks related to visual perception and its relation to knowledge. However, we still know little about apes' capacities regarding the more complex aspects of the relation between seeing, knowing and believing. Only two published studies to date (Call & Tomasello, 1999; O'Connell & Dunbar, 2003) have looked at apes' ability to recognize that not seeing a crucial event can result in a false belief state. And apes' capacity to deal with situations in which there is a disjunction between what is seen and what is known has never been directly investigated. Additionally, while there is a good amount of positive evidence to suggest that apes recognize what others can or cannot see and what others do or do not know based on past visual access, their abilities with regard to themselves in these respects is still largely unknown.

Next Steps

In the remainder of this thesis, I present the original research I undertook to address these outstanding questions about apes' mentalizing abilities. The studies are presented in the order in which they were carried out because I feel this best reflects the progression in my thinking about ape mentalizing and how to go about investigating it. Past research has already shown us that chimpanzees are capable of recognizing both perceptual and knowledge states in others. I therefore began by examining whether chimpanzees (and also bonobos) are further capable of recognizing others' false belief states, which is widely considered to be a relatively advanced mental-state attribution capacity. The apes, however, failed the novel tests I gave them, despite the fact that these

tests were simpler and more species-relevant than previous ones (see Chapter 4). These negative results suggest that false belief understanding, at least on a level that can be translated into adaptive action, may exceed apes' mentalizing capacities. My next step was to examine chimpanzees' capacities with regard to their own mental states, about which very little is currently known. I first confirmed in Study 3 that chimpanzees have good visual perspective taking abilities with regard to themselves, just as they have demonstrated with regard to others. My next question was, given that chimpanzees have difficulty conceiving of false beliefs in others, would they have similar difficulties recognizing when they themselves were experiencing a false perceptual state? As we will see, the results from Study 4 suggest that while some chimpanzees do have such trouble, other individuals seem capable of recognizing when what they see is at odds with what they know. I now present these studies in the following three chapters.⁷

⁷ Note that all of the experiments undertaken for this thesis were in full compliance with the laws of the country in which they were conducted (Germany) concerning the ethical treatment of children and animals in scientific research.

CHAPTER 4 – FALSE BELIEF

STUDY 1 – COMPETITIVE REACHING STUDY

In Chapter 3, we saw that apes have thus far not shown any convincing signs of understanding false beliefs (e.g., Call & Tomasello, 1999; O’Connell & Dunbar, 2003). We also saw, however, that they may have been unfairly disadvantaged in those studies because the tasks were cooperative and required participants to understand communicative intentions. Support for the importance of competitive paradigms came from Hare and Tomasello’s (2004) finding that chimpanzees could not locate a hidden food reward when a cooperative experimenter pointed to it, but they could find the reward when she reached competitively for it. And in tests of visual perspective taking, chimpanzees have performed poorly in cooperative-communicative paradigms (Call et al., 2000; Povinelli & Eddy, 1996b; Povinelli, Rulf, & Bierschwale, 1994; Reaux et al., 1999) but they have performed well in competitive paradigms (Hare et al., 2000; Hare et al., 2001, 2006; Hirata & Matsuzawa, 2001; Menzel, 1974). If competition is crucial, as these findings suggest, this raises the following question: Did apes fail earlier false belief tasks because they had no understanding of false beliefs or because the tasks involved a cooperative experimenter who attempted to communicate to them the location of the hidden reward? Given the latter possibility, my goal was to give apes a nonverbal false belief test that was competitive and did not require participants to understand communicative intentions. I tested both chimpanzees and a small group of bonobos. Although bonobos are as closely related to humans as are chimpanzees (see Figure 1.1),

they had never before been tested for false belief understanding, so it was important to test them here. Further, there is evidence that bonobos are less competitive than chimpanzees (de Waal & Lanting, 1997; Hare, Melis, Woods, Hastings, & Wrangham, 2007), so I was also interested in knowing if they would perform more poorly on the competitive task than chimpanzees. I further wanted a test that could be easily administered to human children. Testing children would allow me to determine whether the novel competitive test was relatively more or less difficult than standard verbal tests, and whether or not it correlated with standard tests. Children's performance would thus provide a useful baseline against which to judge the apes' performance. I tested young 5-year-olds because children typically begin to exceed chance levels in standard verbal tests by this age.

The Competitive reaching false belief test used in this study combined elements of Call and Tomasello's (1999) nonverbal false belief test and Hare and Tomasello's (2004) Pointing-vs.-reaching test. An experimenter hid a reward in one of two identical containers while a competitor observed (and participants, who could not see the hiding themselves, saw that she observed). Unbeknownst to the competitor, the experimenter then switched the locations of the containers, leading the competitor to falsely believe the reward to be in its original location. As the competitor then reached with effort for the *incorrect* container, the experimenter slid the containers within participants' reach so that they could choose a container. Participants who recognized the competitor's false belief should choose the container she was not reaching for. For comparison, a True belief

control test was also included in which the competitor witnessed the switch and thus reached for the correct container.

I ran two slight variations on the procedure, the main difference being whether or not the competitor was out of the room (Version A) or had her back turned (Version B) during the switching of the containers. Version B was added because the children seemed distracted by the competitor leaving the room in Version A, and their performance in that version was weaker than expected. I thought that both children and apes might have more success in the task if this distracting element was removed. I also considered that the competitor's failure to witness the switch might be more salient for participants if they could plainly see her attending elsewhere while the switch happened behind her back. All apes participated in both versions of the test but a different group of children was used for each version.

Given that we are still at the stage of trying to determine whether apes have any capacity whatsoever to attribute false beliefs to others, I looked for evidence of false belief understanding on any level. Thus, in addition to the main active *choice* measure, I measured *looking*. In previous chapters, we saw that both anticipatory looking and violation-of-expectancy (VOE) looking time responses have been taken as evidence that children who fail false belief tests may nevertheless possess some level of false belief understanding (Clements & Perner, 1994; Onishi & Baillargeon, 2005; Ruffman et al., 2001; Southgate et al., 2007; Surian et al., 2007, but see also Perner & Ruffman, 2005). We also saw, however, that there is much debate about how to characterize the knowledge revealed by looking responses. While VOE looking time measures are usually

described as tapping into implicit knowledge (Onishi & Baillargeon, 2005; Southgate et al., 2007; Surian et al., 2007), anticipatory looking has been variously described as indicating implicit (Clements & Perner, 1994), explicit (Southgate et al., 2007) and uncertain (Ruffman et al., 2001) knowledge.

In the current study, I measured neither anticipatory looking nor VOE looking time.⁸ The measure I used is best described as ‘indecisive’ looking: in the moments before participants settled on their choice of container, I noted whether or not they looked at the other container. The fact that I measured yet a third type of looking meant that, even if there were agreement on what type of knowledge anticipatory looking and VOE looking time measures tapped into, I would still have no precedent for how to characterize the knowledge measured by my looking measure. Such indecisiveness or hesitancy would seem to suggest uncertainty, but this is just an intuition. Therefore, while I would consider any positive looking results I found here to be suggestive evidence of some (as yet undefined) level of understanding, the more convincing evidence would be success on the active choice measure.

To be consistent with the order in which I reviewed the extant research in Chapter 3, I begin with the children’s study.

⁸ Anticipatory looking was not an appropriate measure because, as apes did not know where the reward was, there was no way for them to know where the competitor *should* reach if she had a false belief versus a true belief. They could only solve the task by recognizing that when the experimenter had not witnessed the switching of the containers, wherever she reached would be wrong. VOE looking time was also not an appropriate measure because the competitor’s reaching toward one of the containers may have led to stimulus enhancement and biased the results.

Experiment 1 – Children

Methods

Participants

Forty children recruited from kindergartens in Leipzig, Germany were tested, 20 with Version A of the procedure and 20 with Version B.

Version A. Participants were 8 males and 12 females, 59–62 months old (mean = 61 months, SD = 1.2). Seven further children were tested but excluded from analyses, five because of procedural error and two for cheating (e.g., standing to see over the screen blocking their view, or trying to look inside the containers). Four additional children failed to meet the pretest criterion (see below).

Version B. Participants were 11 males and 9 females, 54–61 months old (mean = 58 months, SD = 2.1). Four more children were excluded from analyses because of procedural error, and two children's test sessions were abandoned due to uncooperativeness.

Experimental Set-up

Children sat facing an adult competitor across a table (100 x 60 cm) with a sliding platform that could be moved back and forth between the child and competitor (see Figure 4.1). Two small opaque containers were used to conceal a sticker reward. An opaque screen (32 cm high, 59 cm wide) obstructed the child's view of the containers when necessary. Plexiglas panels (52 cm high, 59 cm wide) blocked the child's and the

competitor's physical access to the containers except through two armholes (39 cm apart from the center of each hole).

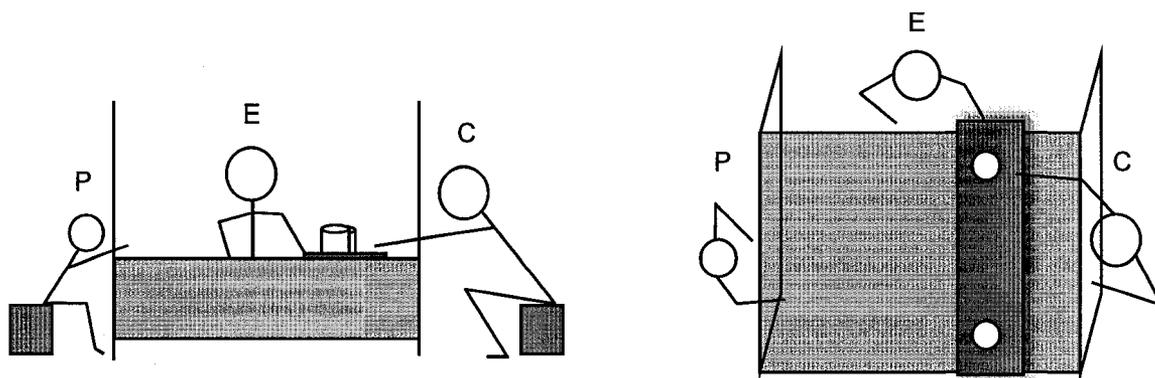


Figure 4.1. Experimental set-up for Study 1: front view on left and top view on right. P = Participant, E = Experimenter, C = Competitor

Design

All children were first given a warm-up and pretest. Half of participants then received four True belief trials followed by four False belief trials and half received the opposite order. A standard verbal false belief test (the Sally-Anne test, Baron-Cohen et al., 1985) was also administered. In Version A, all children received the Sally-Anne test at the end of the session. In Version B, half the children received it at the beginning and half at the end. Version B also included four control trials administered at the beginning of the session, immediately before the test trials. Location of the sticker was randomly determined with the constraints that it had to be on the left or right an equal number of times and could not be in the same container for more than two consecutive trials. All trials were administered in one session that lasted approximately 20–25 minutes.

Procedure

Warm-up. Children were tested in a quiet room in their kindergartens by two female experimenters, one in the role of the competitor. Both were native German-speaking research assistants. Children were first given two warm-up trials to introduce them to the general procedure and establish the competitive context. The experimenter (E) put two containers onto the sliding platform, one on the left and one on the right. She explained that they were going to play a game in which the child had to try to get a sticker hidden in one of the containers and that the competitor would also try to get the sticker. E then placed the sticker in one container while both the child and competitor watched. E began sliding the platform first in the direction of the competitor, who reached with effort but unsuccessfully for the container with the sticker inside (hereafter called the ‘correct’ container). E then slid the platform over to the child, who could then reach for and choose a container. To establish the competitive context, in the second trial the competitor managed to reach the container and take the sticker before children could do so. The competitor acted pleased when she got the sticker and disappointed when the child got it, and she often made competitive comments such as “Okay, this time *I’m* going to get the sticker!”

Pretest. Following the warm-up, a pretest was given to verify that participants could use the competitor’s reach as a cue to the reward’s location. In these trials, before hiding the sticker E positioned an opaque screen to block the child’s (but not the competitor’s) view of the containers. Children could see the competitor’s face over the screen, and they observed as the competitor watched E hide the sticker. The competitor

witnessed the hiding with clear interest, leaning forward, nodding her head, and making sounds such as “ah hah” to show that she was paying attention. So that participants could not use the competitor’s gaze to infer where the reward was being hidden, the competitor directed her gaze straight ahead during the hiding, occasionally looking rapidly back and forth between the containers. She did not track the reward and never focused her gaze on one container. When E removed the screen the competitor reached unsuccessfully for the correct container. E then slid the containers toward children so they could choose a container. Children who understood that the competitor knew the location of the sticker because she had witnessed the hiding should choose the same container she did.

In Version A, children received 12 pretest trials and needed at least 9 correct to pass. Because the pretest proved to be unnecessarily long (most children earned perfect scores), for Version B the criterion was reduced to three trials in a row correct.

False belief test. As in the pretest, E blocked the child’s view of the containers with a screen and hid the sticker as the competitor observed (and the child saw that she observed). E removed the screen and the competitor immediately either left the room (Version A) or turned around in her seat (Version B) before she had a chance to reach. She gave some excuse for doing so (e.g., to make a phone call or blow her nose). E then got the child’s attention and switched the positions of the containers without revealing the location of the sticker, smiling mischievously and glancing occasionally at the door (Version A) or the competitor’s back (Version B) to make sure the competitor did not see the switch. In Version B, the competitor muttered to herself and became highly absorbed in her task while her back was turned, clearly not attending to E’s actions. E secretly

signaled the competitor by coughing inconspicuously when she was finished, and the competitor returned to her position facing the child. E then slid the platform first in the direction of the competitor, who reached with effort but unsuccessfully for the incorrect container. E next slid the platform over to the child while the competitor continued to reach. Children who recognized the competitor's false belief should choose the container the competitor was *not* reaching for. Children who chose the correct container were allowed to keep the sticker; otherwise E slid the platform back over to the competitor, who took the sticker.

The competitor showed obvious surprise when the location of the reward was revealed. And although the task was essentially nonverbal, the experimenter and competitor chatted naturally with each other and with children throughout, while being careful to avoid references to the competitor's belief states. For example, when E displayed the sticker before hiding it, the competitor said excitedly, "Oh, a smiley/dolphin/etc.! I'm going to get it!" If the child chose the incorrect container E made statements such as, "Let's try again."

True belief test. The procedure for True belief trials was the same as for False belief trials, with the crucial difference that the competitor witnessed the switching of the containers. As before, the competitor left the room or turned her back without reaching after E hid the sticker. But in this case E just sat and waited, glancing occasionally at her watch, the child, and the door or the competitor's back. When the competitor resumed her position at the table E switched the positions of the containers in full view of both the

child and the competitor. The competitor then reached for the correct container, and the proper response for children was to also choose that container.

Control trials. In Version B, I also included control trials immediately preceding the test trials. These trials were added after results were obtained for Version A in order to rule out alternative reasons why participants might have difficulty with the False belief trials. The control trials were designed after ones used by Call and Tomasello (1999), and they evaluated two non-mentalistic component skills required to do well in the task, as follows:

Two *Ignore-competitor* trials tested participants' capacity to disregard the competitor's reach when they knew it was wrong. The procedure was as for the False belief trials except that while the competitor's back was turned E removed the sticker from the container in full view of the child and moved it to the other container. When the competitor faced forward again she reached for the incorrect container. Children who could ignore the competitor's incorrect reach should choose the other container.

Two *Invisible displacement* trials tested participants' knowledge that the reward moved from one location to the other when the containers were switched. While the competitor watched, E hid the sticker and then removed the screen blocking the child's view. The competitor immediately started reaching for the correct container but then became distracted, removed her arm, and turned around in her seat. While her back was turned, E switched the locations of the containers. When the competitor faced forward again she did *not* resume reaching as E slid the containers toward the child. If children understood that the competitor's reach indicated the initial location of the sticker, but that

the sticker had since been moved to the other location along with its container, they should look for the sticker in the new location. The two types of trials were administered in alternating order, always beginning with an Invisible displacement trial.

In the case that participants failed the False belief test, then their performance on the control trials would be informative: If they also did poorly in the control trials, then they could have failed the False belief test *not* because they did not understand false beliefs, but because they did not possess the necessary non-mentalistic component skills.

Sally-Anne test. The Sally-Anne test was administered to confirm that the children in the sample were typical for their age regarding performance in a standard verbal test. One of the experimenters acted out the test for children using dolls named Max and Hannah, a plastic cookie, a small metal pot with lid, and an opaque plastic jar. According to the standard procedure (Baron-Cohen et al., 1985), Max put the cookie in the pot and left briefly, during which time Hannah entered and moved the cookie to the jar. Hannah then left, Max returned, and the experimenter asked children the customary test question and two control questions in a fixed order: (1) “Where will Max look for the cookie first?” (2) “Where is the cookie now?” and (3) “Where was the cookie before Max left?” Children passed only if they answered all three questions correctly, either verbally or by pointing.

Coding and Analyses

Choice. The main measure was active choice, defined as the container children had settled on by the time E finished sliding the platform over to them. Children chose by reaching an arm through a hole in the Plexiglas barrier and touching a container.

Occasionally they began to reach through both holes at once, in which case E asked them to put only one arm through at a time. Choice was coded live by one of the experimenters and was double-checked later by myself from videotapes of the experimental sessions.

Proportion scores were used because two individual trials (out of the total 320 True and False belief trials) had to be excluded from the analyses: one True belief trial was excluded because of procedural error (the competitor began reaching before the switch occurred), and one False belief trial was excluded because the child refused to make a choice. I calculated choice proportion scores for each participant in both the True and False belief conditions by dividing the number of correct trials by the total number of trials. Average scores were compared to the proportion of trials expected to be correct by chance (0.50). However, I considered that a more meaningful comparison for False belief performance in this task might be the proportion of incorrect True belief trials, rather than chance (see Carpenter et al., 2002; Lohmann, Carpenter, & Call, 2005 for a similar argument regarding other tasks). That is, because of the high salience of the competitor's reaching cue, if participants had no understanding of false beliefs they would be likely to choose the container the competitor reached for in both conditions rather than choosing randomly (i.e., at chance levels as a group). If participants chose that container perfectly consistently they would achieve 100% success in True belief trials and 0% success in False belief trials. But even simple strategies can break down on occasion, for example because of momentary distraction. This would result in some proportion of failures in True belief trials and a corresponding proportion of accidental successes in False belief trials. It is thus informative to compare the proportion of successes in False belief trials to

the proportion of failures in True belief trials (1 minus the True belief score, or the *True belief complement*). If these two proportions do not differ it suggests participants are using the same strategy across conditions. However, if the False belief proportion is significantly higher than the True belief complement this suggests that participants are discriminating between conditions and thus have some understanding of the competitor's false beliefs. Note, however, that it cannot be known for certain whether the most appropriate comparison is to chance or to the True belief complement, so I also provide the more conservative comparison to chance.

Looking. The second measure was whether participants looked, at least once, at the container the competitor was not reaching for during the couple of seconds it took E to slide the platform toward them. This measure was coded from the videotapes of the experimental sessions.

The looking subset of trials. The trials included in the looking analysis were limited to those trials in which participants chose the same container as the competitor, for three reasons: (1) I was only interested in False belief trials in which participants displayed no understanding of false belief in their active choices; (2) trials in which participants chose the other container confounded choosing with looking, and including them could therefore artificially inflate the False belief results; and (3) by coding only the subset, I could directly compare True and False belief trials on an equal footing, because participants chose the container the competitor reached for in both.

Because I used within-group comparisons, only children who chose the same container as the competitor at least once in each condition were included in the looking

analysis. There were 13 such children in each of Versions A and B. After excluding the other children, the following percentage of all trials remained for the looking analysis: 55% of True belief trials and 41% of False belief trials from Version A; and 49% of True belief trials and 26% of False belief trials from Version B. My question was thus: Before choosing the same container as the competitor, did children look at the other container in a greater proportion of False belief trials than True belief trials? For each participant, I calculated a looking proportion score in each condition by dividing the number of trials in which the child looked by the number of trials retained for that child in that condition. All analyses were non-parametric and all reported p values are exact and two-tailed.

Reliability

To assess inter-observer reliability, coders naïve to the hypotheses of the study independently coded 100% of children's experimental trials for choice and 35% (Version A) and 30% (Version B) of the subset trials for looking. These trials were chosen randomly. Excellent levels of agreement were achieved: Cohen's kappas were 1.00 for choice in both versions and 0.92 (Version A) and 0.83 (Version B) for looking ($p \leq 0.002$ in both cases).

Results and Discussion

Preliminary analyses revealed no significant sex or age differences in False belief choice scores for either Version A or B ($p \geq 0.43$ in all cases). I therefore collapsed across these variables for all subsequent analyses.

Pretest

Children demonstrated from the outset a clear ability to use the competitor's reach to find the hidden reward. In Version A, the criterion to pass was at least 9/12 trials correct, and children's average number correct was 11.6 (97%) (range = 9–12, SE = 0.21). In Version B, in which the criterion was three consecutive trials correct, children's average number of trials to reach criterion was 4.3 (range = 3–11, SE = 0.49).

Choice

Order of presentation of the True belief (TB) and False belief (FB) tests had no effect on performance in either test in either Version A or B (Mann-Whitney U tests: $U \geq 40.50$, $N = 20$, $p \geq 0.47$ in all cases). There were also no statistically significant differences in children's True belief or False belief performance across versions, but the difference did approach significance for True belief (TB: $U = 136.00$, $N = 40$, $p = 0.062$; FB: $U = 153.00$, $N = 40$, $p = 0.20$). Additionally, TB-FB difference scores (calculated by subtracting the False belief from the True belief proportion score for each child) were significantly different across versions ($U = 124.50$, $N = 40$, $p = 0.038$), indicating that children's degree of success in the True belief as compared to the False belief condition differed across versions. I therefore present the results for Versions A and B separately as well as combined.

Figure 4.2a shows the mean proportion of correct choices by children in the True belief and False belief conditions, both combined across versions and separately. In the combined analysis, children performed significantly better than chance in the True belief test, verifying that they did not have any basic problems with the general procedure

(Wilcoxon test: $T^+ = 582.00$, $N = 34$ [6 ties], $p < 0.001$). They also performed better than chance in the False belief test ($T^+ = 387.00$, $N = 31$ [9 ties], $p = 0.004$), indicating that they recognized the competitor's false belief.

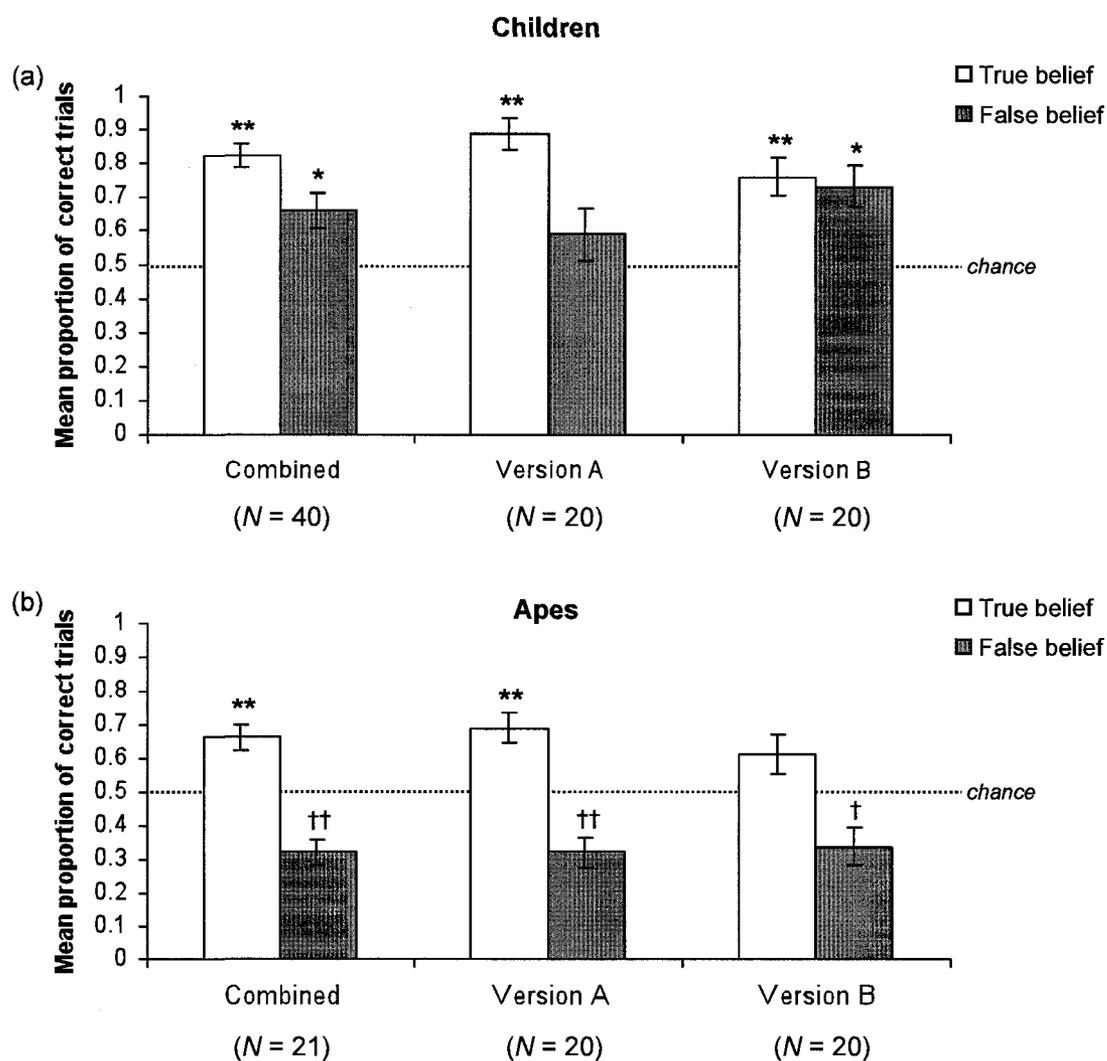


Figure 4.2. Choice analysis for Study 1: mean proportion of correct trials (\pm SE) for (a) children and (b) apes in both versions of the procedure combined and separately. * $p < 0.01$, ** $p < 0.001$, significantly better than chance; † $p < 0.01$, †† $p < 0.001$, significantly worse than chance

In Versions A and B separately, children's True belief scores still exceeded chance (Version A: $T^+ = 187.00$, $N = 19$ [1 tie], $p < 0.001$; Version B: $T^+ = 116.00$, $N = 15$ [5 ties], $p < 0.001$). In Version B, children's False belief scores were also significantly greater than chance ($T^+ = 134.00$, $N = 17$ [3 ties], $p = 0.004$), as in the combined analysis. In Version A alone, however, children's False belief performance was numerically but not significantly better than chance ($T^+ = 73.00$, $N = 14$ [6 ties], $p = 0.19$), suggesting that children found Version A more difficult than Version B. However, children's False belief scores were significantly higher than the True belief complement in Version A ($T^+ = 133.00$, $N = 16$ [4 ties], $p < 0.001$), indicating that they did respond differentially across conditions (this was also true for Version B alone [$T^+ = 136.00$, $N = 16$ (4 ties), $p < 0.001$] and for the combined analysis [$T^+ = 523.50$, $N = 32$ (8 ties), $p < 0.001$]). Thus, if the True belief complement, rather than chance, is used as the comparison with False belief, children's performance in Version A suggests some recognition of the competitor's false belief state.

I did two final analyses to get a clearer sense of how children's performance varied across Versions A and B. First, I set the passing criterion for a test at three or more trials correct out of four (75% success). According to this criterion, for the False belief test, 40% of children passed in Version A and 70% passed in Version B. For the True belief test, 90% of children passed in Version A and 70% passed in Version B. Second, I examined children's performance across versions using each child's very first trial only. In Version A, 60% of the children whose first trial was a False belief trial got that trial

correct, compared to 70% in Version B. For True belief, the percentages were again 90% for Version A and 70% for Version B.

Looking

The looking measure was included to see if participants might still show some degree of false belief understanding, even if they did not perform well in their active choices. Because children's False belief performance was somewhat weak in Version A, looking was of most interest in that version. Nevertheless, I analyzed both versions for comparison. Figure 4.3a shows that in Version A, looking occurred in a numerically greater average proportion of False belief than True belief trials (FB = 0.32, TB = 0.19), but this difference was not statistically significant ($T^+ = 21.00$, $N = 7$ [6 ties], $p = 0.30$). There was no difference in Version B (TB = 0.31, FB = 0.35; $T^+ = 33.00$, $N = 11$ [2 ties], $p = 1.00$), which was to be expected given that children did so well on the active choice measure in that version.

Sally-Anne Test

All children who answered the Sally-Anne test question correctly also answered both control questions correctly. There were no significant differences in children's Sally-Anne performance across versions ($U = 190.00$, $N = 40$, $p = 1.00$). There were also no observable order effects in Version B, in which the Sally-Anne test and the Competitive reaching test were counterbalanced for order (for both tests, $U \geq 33.00$, $N_{\text{first}} = N_{\text{last}} = 10$, $p \geq 0.20$). I therefore report the results collapsed across both versions and order.

Confirming that the children in the sample were age-typical regarding their performance on a standard verbal false belief test, the proportion of children passing the Sally-Anne test was significantly greater than the chance proportion of 0.50 (Binomial test: $N = 40$, $p = 0.04$). As a group, children's performance in the Sally-Anne test closely

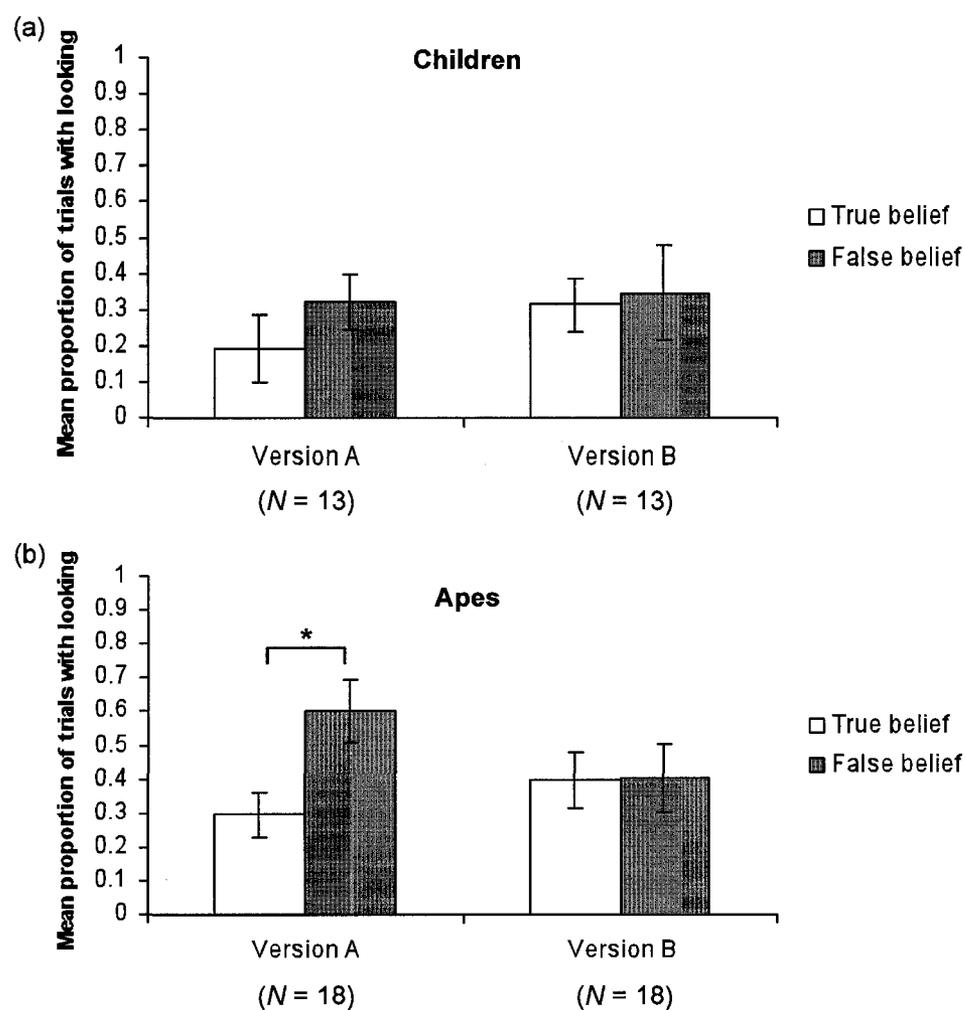


Figure 4.3. Looking analysis for Study 1: mean proportion of trials (\pm SE) in which (a) children and (b) apes looked toward the unchosen container before choosing. Note that proportions are based only on trials in which participants chose the same container as the competitor, and for apes in Version A the measure was face rather than eye orientation.

* $p < 0.05$

matched their performance in the Competitive reaching test: proportion scores (0.68 and 0.66, respectively) were not significantly different ($T^+ = 237.50$, $N = 30$ [10 ties], $p = 0.92$). This level of success may seem somewhat low for 5-year-old children, given that children of this age are widely considered to have a well-developed understanding of false belief. However, it must be kept in mind that although young 5-year-olds usually perform significantly better than chance as a group in false belief tests, their performance is typically not at ceiling levels until later in development. In Wellman et al.'s (2001) meta-analysis of 178 false belief studies, at 60 months old, children had an approximately 0.70 probability of being correct. Further, Wellman et al. (2001) also observed that the country in which a false belief study was conducted partially predicted the age at which children succeeded. Germany was unfortunately not represented in the analysis of effects of country on performance. However, children from English-speaking countries (Canada, Australia, the United States, and the United Kingdom) performed better at younger ages than children from Austria, another German-speaking country. It is thus possible that something about the German wording of the instructions or test questions makes them slightly more difficult for children than the English wording.

Because the Sally-Anne test included only one trial, I also looked at children's performance in their first False belief trial only of the Competitive reaching test (in both versions combined). The proportion of children responding correctly on that trial was 0.60, also not significantly different from the Sally-Anne results ($T^+ = 110.00$, $N = 19$ [21 ties], $p = 0.65$). These findings verify that, as a group, the children were equally challenged by the Sally-Anne test and the Competitive reaching test. I was also curious to

know how individual children's scores matched up across the Sally-Anne test and the Competitive reaching test, given that the two procedures were so different. I therefore compared children's first False belief trial in the Competitive reaching test with their Sally-Anne test trial. Twenty-one of the 40 children (53%) were consistent across tests, either failing or passing both. A test of agreement yielded a non-significant kappa of -0.02 ($p = 1.00$). Furthermore, when I included all four nonverbal False belief trials and split the children into two groups, those who passed the Sally-Anne test and those who failed it, the former did no better as a group on the Competitive reaching test than the latter ($U = 170.00$, $N_{\text{pass}} = 27$, $N_{\text{fail}} = 13$, $p = 0.88$). This pattern of results was also found for each version of the procedure analyzed separately (Version A: $U = 35.50$, $N_{\text{pass}} = 13$, $N_{\text{fail}} = 7$, $p = 0.47$; Version B: $U = 34.00$, $N_{\text{pass}} = 14$, $N_{\text{fail}} = 6$, $p = 0.53$). In short, performance was highly consistent across tests at the group level but was inconsistent at the individual level, probably because of differing task features unrelated to the false belief element of the tests.

Control Trials

Children performed well in the control trials, run in Version B only. All children earned a perfect score in the two Ignore-competitor trials. In the Invisible displacement trials, the mean proportion of trials correct (0.80, $SE = 0.06$) was far above chance ($T^+ = 78.00$, $N = 12$ [8 ties], $p < 0.001$). Given that children did well in the False belief trials in Version B (and also did well in the control trials), comparisons between the two types of trials are not so informative. However, since I report the comparison between Invisible displacement and False belief performance for apes below (where it is more informative),

I provide it here as well. Children's performance on Invisible displacement trials was not significantly different from their performance on False belief trials ($T^+ = 71.00$, $N = 15$ [5 ties], $p = 0.58$). There was no significant correlation between performance on Invisible displacement trials and False belief trials (Spearman's rank correlation: $r_s = -0.20$, $N = 20$, $p = 0.42$), which is not surprising given that many children were at ceiling levels with regard to Invisible displacement.

Overall, the results for Study 1 (at least in Version B) were typical for children in the age group I tested. By choosing the container the competitor reached for more often in the True belief trials and the one she did not reach for more often in the False belief trials, children demonstrated that they recognized the competitor's false belief about the location of the reward. The next question was whether apes would also show such insight. I believed that the competitive nature of the task would give them a better chance of succeeding than they had had in previous tasks using cooperative paradigms.

Experiment 2 – Chimpanzees and Bonobos

Methods

Participants

Sixteen chimpanzees (*Pan troglodytes*) and 5 bonobos (*Pan paniscus*) housed in social groups at the Wolfgang Köhler Primate Research Center (WKPRC) in Leipzig, Germany, participated in both Versions A and B of the procedure. An additional chimpanzee was tested but then dropped from the study because he would not make clear choices. Participants included 6 males and 15 females, all captive-born, 9 mother-reared

and 12 nursery-reared. Their ages ranged from 4 to 28.5 years when testing began (mean = 14.8 years, SD = 8.8). All apes had previously taken part in a variety of social and physical cognition studies and were familiar with the object-choice paradigm. Eleven chimpanzees had participated several years earlier in a study in which they used a human competitor's reach to find hidden food (Hare & Tomasello, 2004). Apes were not food-deprived and water was provided *ad libitum* throughout testing. One chimpanzee's Version A data were excluded from analyses because she repeatedly tried to look into the containers before choosing. Another chimpanzee's Version B data were excluded because of procedural error. The final number of apes in each version was thus 20.

Experimental Set-up

The set-up was the same as the children's except for the following: the table (92 x 80 cm) was placed in a three-sided windowed testing booth between two adjoining enclosures. The ape and the competitor (myself) each sat in a different enclosure. The experimenter (a student assistant) sat just outside the booth by the side of the table (see Figure 4.1). Rewards were grapes or banana slices. The Plexiglas windows in front of the ape and competitor each measured 49 cm high and 69 cm wide and had holes in them spaced 59 cm apart from center to center. For the competitor, the holes were large enough to fit her arms through but for safety reasons the apes' holes were just large enough for their fingers. The opaque screen used to block apes' view was 28 cm high and 64 cm wide.

Design

All apes received Version A first and then B. In both versions apes were first given warm-up and pretest/training trials. In Version A, they then received eight True belief and eight False belief trials (four of each per session for two sessions). At the start of each new test session apes were given six further pretest trials to refresh their memory of the procedure. In Version B, apes received four True belief and four False belief trials in a single testing session. Each ape thus received a total of 12 True belief and 12 False belief trials across Versions A and B. In both versions, the True and False belief trials were administered in blocks of four trials, counterbalanced for order across apes. In Version B, there were also four control trials immediately preceding the test trials. Finally, in a later session, two warm-up trials and an additional four post-test Invisible displacement control trials were also administered. The location of the reward was randomized with the constraints that it had to be on the left or right an equal number of times and could not be in the same location for more than two consecutive trials.

Procedure

Warm-up. The warm-up was the same as for children except that, as I could not verbally explain the task to apes, I gave them three warm-up trials at the start of Version A instead of two (and then two more at the start of Version B, as for the children). The experimenter (E) put two containers onto the sliding platform, one on the left and one on the right. She placed the food reward in one container in full view of both the ape and competitor. E then began sliding the platform first in the direction of the competitor, who reached with effort but unsuccessfully for the correct container (i.e., with the food

inside). E then slid the platform over to the ape, who could then reach for and choose a container. To establish the competitive context, the competitor successfully took the food in the second or third warm-up trial, before the ape could do so. The competitor also adopted a demeanor appropriate for competing with apes. While watching E hide the food, she performed behaviors indicative of attention and excitement in chimpanzees, such as head bobbing and grunting. When the competitor won the reward she gave the impression of consuming it greedily,⁹ and when she lost the reward she reacted angrily by banging on the windows or walls and shouting.

Pretest/training. As with children, the purpose of these trials was to ensure that apes could reliably use the reaching cue provided by the competitor to find the hidden food. Before hiding the food, E positioned an opaque screen to block the ape's, but not the competitor's, view of the containers. Apes could see the competitor's face over the screen, and they observed as the competitor intently watched E hide the food. As with the children, the competitor avoided visually tracking the reward or focusing her gaze on either container during this time, so that apes could not use such cues to find the food. When E removed the screen the competitor reached unsuccessfully for the correct container. E then slid the containers toward the ape so the ape could choose a container. If apes understood that the competitor knew the location of the food because she had witnessed the hiding, they should choose the same container she reached for.

⁹ For hygienic reasons, the competitor could not actually eat the food. However, so that the apes would think she ate it, she made the movement of putting it into her mouth (while actually keeping it in her closed hand) and pretended to chew and swallow.

In Version A apes received 18 trials in a session and needed to get at least 13 trials correct to pass. They were given repeated sessions as needed until they reached criterion. The purpose of the pretest in Version B was to make sure that apes had not forgotten, in the time since they completed Version A, to use the competitor's reach as a cue to the location of the hidden food. The procedure was the same as for Version A, but the criterion was three trials in a row correct. Thus, as soon as apes choose the correct container in three consecutive trials, they advanced to testing.

False belief test. As in the pretest, E blocked the ape's view of the containers with a screen and hid the food as the competitor observed (and the ape saw that she observed). E then removed the screen and the competitor immediately either left the room (Version A) or turned around in her seat (Version B) before she had a chance to reach. E then got the ape's attention and switched the positions of the containers without revealing the location of the food, smiling mischievously and glancing occasionally at the door (Version A) or the competitor's back (Version B). As with children, in Version B, the competitor muttered to herself and became highly absorbed in her task (e.g., tying her shoe or scratching her leg) while her back was turned to make it clear that she was not attending to E's actions. E secretly signaled the competitor by coughing when she was finished and the competitor returned to her position facing the ape. E then slid the platform first in the direction of the competitor, who reached with effort but unsuccessfully for the incorrect container. While the competitor continued to try to reach for the container, the experimenter slid the platform toward the ape and allowed the ape to make a choice. Apes who recognized the competitor's false belief should choose the

container the competitor was *not* reaching for. Apes who chose the correct container were allowed to keep the food; otherwise E slid the platform back over to the competitor, who took the food and pretended to eat it.

True belief test. The procedure for the True belief control trials was as for the False belief trials except that the competitor witnessed the switching of the containers. Immediately after E hid the food, the competitor either left the room (Version A) or turned her back and became absorbed in some task (Version B). The experimenter sat quietly and waited during this time, looking occasionally at the door or the competitor's back. After approximately the same amount of time had passed as in False belief trials (about 10 seconds), E coughed and the competitor returned to her position opposite the ape. Only then did the experimenter switch the positions of the containers, in full view of both the ape and the competitor. The competitor then attempted to reach, unsuccessfully, for the *correct* container as the experimenter slid the platform toward the ape.

Control trials. Immediately preceding the test trials in Version B, apes received two Ignore-competitor and two Invisible displacement trials (alternating, as for the children). However, for the Invisible displacement trials, because Call and Tomasello (1999) found that apes did better when the marker indicating the reward's location was present both during the switch and while apes made their choice, the competitor left her arm resting in the hole while her back was turned. When she faced forward again after the switch, she resumed reaching (now toward the incorrect container) as the ape chose a container.

Post-test Invisible displacement control trials. In retrospect, I realized that apes' modified Invisible displacement control trials administered before the test contained an element of false belief: the competitor's reach toward the incorrect container while the ape chose. Therefore, after testing was complete I gave apes four additional Invisible displacement trials. In two *Back-turned* trials (more similar to the children's), the competitor left her arm in the hole while her back was turned and the switch happened, as before, but when she faced forward again she removed her arm and did not resume reaching while the ape chose. In two *Facing-forward* trials, the competitor did not turn her back during the switch, and she removed her arm from the hole just as E began to switch the containers. These trials also alternated, always beginning with a Back-turned trial.

Coding and Analyses

Choice. The container chosen was defined as the one apes were orienting toward and actively poking at when E had finished sliding the platform over to them. Occasionally, an ape tried to choose both containers, in which case E waited until the ape had clearly chosen one container. If necessary, she slid the containers away from the ape and back again. Choice was coded live and double-checked later from the videotapes.

A number of trials were excluded from analyses because of procedural error (three trials), the ape did not clearly choose a container (three trials), or the ape refused to participate (eight trials, all from the same individual). I therefore used proportion scores, which I calculated for each participant, in both the True and False belief conditions, by dividing the number of correct trials by the total number of trials.

Average scores were compared to the proportion of trials expected to be correct by chance (0.50). However, as with the children, I thought that the proportion of incorrect True belief trials (the True belief complement) might provide a more meaningful comparison for False belief performance than chance in this task, and so I also did that comparison (see Experiment 1, Coding and Analyses, for details).

Looking. As for children, looking was coded when apes looked *at least once* at the container the competitor was not reaching for during the couple of seconds it took E to slide the platform toward them (see Figure 4.4). For Version A, however, I was unable to

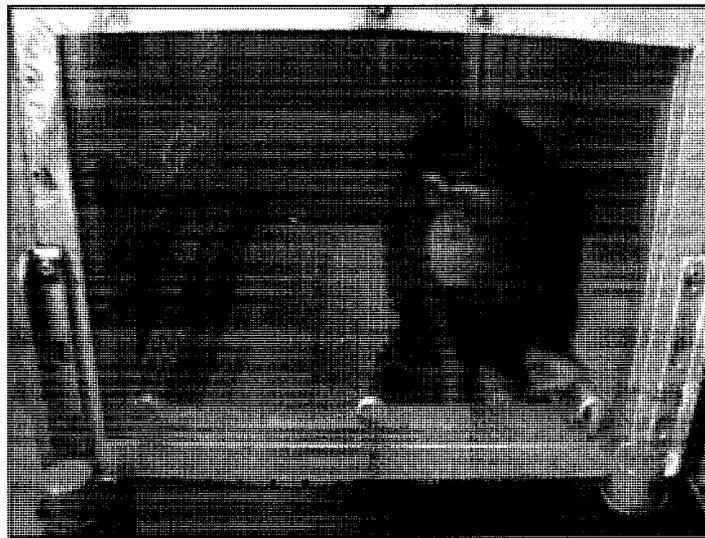


Figure 4.4. Indecisive looking: A chimpanzee looks at the container on her right just before touching the one on her left to choose it in Study 1.

obtain adequate inter-observer reliability for eye direction because the apes' eyes were difficult to see on the videos due to the poor quality of the image. I therefore used face orientation as the looking measure. The criterion for coding was that apes had to turn not just their head, but also their face, directly toward a container. For example, instances in

which apes turned their head toward a container but had their face directed downwards toward the floor or upwards toward the ceiling were not counted. Other researchers have used head orientation as a measure of focus of attention (e.g., Bethell, 2003; Tomasello, Call, & Hare, 1998); and Emery (2000) suggests that “body, head, and eye gaze are in congruent directions for the majority of the time” (p. 583). Nevertheless, the more precise measure is eye direction (Bethell, Vick, & Bard, 2007), and so for Version B, I improved the lighting and camera angles in order to make coding of eye direction possible.

The looking subset of trials. As with children, the trials included in the looking analysis were limited to a subset of the total number of trials. Trials were chosen for the subset in four ways. First, for Version A, I excluded the apes’ second session because the difficulty of the coding made the task extremely time consuming, and also because children only received one session. Further, responses in earlier trials are considered to be more meaningful because later trials may be affected by earlier experiences. Second, I coded only the subset of trials in which participants chose the same container as the competitor, for the same reasons as outlined in Experiment 1. Third, I had to exclude eight trials from the looking analysis because they were not recorded or the ape’s face was off screen during the relevant part of the trial. Fourth, because I used within-groups comparisons, only participants who chose the same container as the competitor at least once in each condition could be included in analyses. In each version, 18 apes met this criterion.

The percentage of all trials included in the looking analysis was thus as follows: 58% of True belief and 59% of False belief trials from Version A (session 1), and 59% of

True belief and 60% of False belief trials from Version B. For each of the 18 apes included in the looking subset of trials, I calculated a looking proportion score in each condition by dividing the number of trials in which the ape looked by the number of trials retained for that ape in that condition. Unless otherwise noted, all analyses were non-parametric and all reported p values are exact and two-tailed.

Reliability

Coders who were naive to the experimental hypotheses and blind to condition coded 40% (Version A) and 35% (Version B) of test trials for choice. Both kappas were 1.00. They also coded 55% of subset trials for looking in Version A and 31% in Version B, with respective kappas of 0.81 and 0.91 ($p < 0.001$ in both cases). A greater number of Version A trials were coded because the coding was more difficult and I wanted to be more confident of the results. All trials were chosen randomly.

Results and Discussion

Preliminary analyses revealed no significant sex, age, rearing or species differences in False belief performance in either version of the procedure ($p \geq 0.18$ in all cases). Thus, chimpanzees performed no better than their less competitive cousins, bonobos, in this competitive task. For the remainder of the analyses I collapsed across species, as well as across sex and rearing history.

Pretest

In Version A, apes' average number of pretest sessions to reach the criterion of 13/18 trials correct was 1.9 (range = 1–4, SE = 0.23) and the average passing score was

14.5 (80.6%) (range = 13–18, SE = 0.34). In Version B, apes' average number of trials to reach the criterion of three consecutive trials correct was 5.6 (range = 3–28, SE = 1.30).

Choice

There were no order effects ($U \geq 35.50$, $N = 20$, $p \geq 0.28$ in all cases) and no differences in True belief or False belief performance across Versions A and B (TB: $T^+ = 63.50$, $N = 14$ [5 ties], $p = 0.51$; FB: $T^+ = 55.50$, $N = 14$ [5 ties], $p = 0.87$). There was also no significant difference in the TB-FB difference scores across versions, indicating that apes' pattern of responding was similar in each version ($T^+ = 75.00$, $N = 15$ [4 ties], $p = 0.41$). I nevertheless report the results for Versions A and B combined and separately for direct comparison with the children, for whom separate analyses were important.

In False belief trials, apes' performance contrasted sharply with children's (see Figure 4.2b). Apes were significantly *worse* than chance in both the combined and separate analyses (combined: $T^+ = 183.00$, $N = 19$ [2 ties], $p < 0.001$; Version A: $T^+ = 116.00$, $N = 15$ [5 ties], $p < 0.001$; Version B: $T^+ = 70.00$, $N = 12$ [8 ties], $p = 0.014$). False belief performance was also not significantly different from the True belief complement (combined: $T^+ = 86.00$, $N = 18$ [3 ties], $p = 1.00$; Version A: $T^+ = 55.50$, $N = 14$ [6 ties], $p = 0.87$; Version B: $T^+ = 79.00$, $N = 16$ [4 ties], $p = 0.59$). In their active choice responses, apes therefore gave no sign that they recognized the competitor's false beliefs.

In contrast, in the True belief control condition, apes performed significantly better than chance in both the combined analysis and in Version A alone (combined: $T^+ = 117.00$, $N = 15$ [6 ties], $p < 0.001$; Version A: $T^+ = 89.00$, $N = 13$ [7 ties], $p = 0.001$),

suggesting they had no major problems with the basic structure of the task. Their Version B performance analyzed separately only approached significance in a one-tailed analysis ($T^+ = 69.50$, $N = 13$ [7 ties], $p_{\text{two-tailed}} = 0.11$, $p_{\text{one-tailed}} = 0.054$), although it was numerically higher than chance and significantly better than False belief ($T^+ = 136.50$, $N = 17$ [3 ties], $p = 0.003$). Apes' True belief performance may have deteriorated by the time they participated in Version B because they had by then repeatedly used the strategy of choosing the container the competitor reached for, only to have it fail in the False belief trials. This could have prompted them to choose less systematically in later trials, bringing their performance down closer to chance levels.

To be consistent with the children's analysis, I also examined apes' True belief and False belief performance in each version using a pass criterion of 75% of trials correct (at least 6/8 trials correct in Version A and at least 3/4 correct in Version B). For the False belief test, 0% of apes passed in Version A and 10% passed in Version B. For the True belief test, 50% of apes passed in both versions. And analyzing just the very first trial only for each ape, in Version A, 11% of the apes whose first trial was a False belief trial got that trial correct, compared to 40% in Version B. For True belief, the percentages were 44% for Version A and 80% for Version B. Because the same apes participated in both versions, however, comparing first-trial performance across versions is less informative for the apes than for the children.

An important issue is how the apes' previous experience using a competitor's reach to find hidden food may have affected their performance in the False belief test. Apes with more experience using the reaching cue successfully in the past might have

had greater difficulties choosing contrary to the reach in the False belief test. I found no evidence of this when comparing the False belief performance of the 11 apes who took part in Hare and Tomasello's (2004) Pointing-*vs.*-reaching test with those who did not (for each version separately and combined, $U \geq 31.00$, $N_{\text{separate}} = 20$, $N_{\text{combined}} = 21$, $p \geq 0.15$). Additionally, looking at differential experience with the reaching cue within this study, there was no significant correlation between the number of pretest trials to reach criterion and False belief performance (separately and combined, $|r_s| \leq 0.16$, $N_{\text{separate}} = 20$, $N_{\text{combined}} = 21$, $p \geq 0.49$). Previous experience thus did not affect the apes' performance.

Looking

In contrast to the choice results, the apes' looking results were slightly more suggestive of some level of false belief understanding (see Figure 4.3b). The pattern of looking responses was significantly different across the two versions of the procedure, justifying a separate analysis of each ($T^+ = 47.00$, $N = 10$ [5 ties], $p = 0.045$). In Version A, apes looked at (i.e., oriented their face toward) the unchosen container in a significantly greater proportion of False belief than True belief trials ($T^+ = 95.50$, $N = 15$ [3 ties], $p = 0.041$). These findings are consistent with the interpretation that apes may have recognized the competitor's false belief on some level. However, this result was not replicated in Version B, in which apes looked toward the unchosen container equally often in both conditions ($T^+ = 46.50$, $N = 13$ [5 ties], $p = 0.96$). I considered a possible alternative explanation for the positive looking result in Version A: apes who chose the container the competitor reached for in a False belief trial would not receive the reward. This failure could make them unsure of this strategy in subsequent trials and cause them

to be more uncertain about their choices. To test this possibility, for each ape in the Version A looking analysis, I selected out the first True belief subset trial and first False belief subset trial (in which responding could not be influenced by earlier failures) and compared these. Looking occurred in 33% of True belief trials and 61% of False belief trials. Although this difference was not statistically significant ($T^+ = 63.00$, $N = 13$ [5 ties], $p = 0.27$), there was a tendency for apes to look at the unchosen container more often in the False belief condition than in the True belief condition, even from the earliest trials.

Control Trials

Because apes failed the False belief test, it was important to examine whether their failure could have been due to some non-mentalistic component of the task. With the exception of one chimpanzee (who got one trial wrong), all apes earned perfect scores in the two Ignore-competitor trials (mean proportion correct = 0.98, SE = 0.03). For the pretest Invisible displacement trials, the mean proportion correct (0.40, SE = 0.08) was not significantly different from the chance proportion of 0.50 ($T^+ = 38.50$, $N = 10$ [10 ties], $p = 0.34$). However, the false belief element of the pretest Invisible displacement trials (see Procedure) may have negatively affected apes' performance. The results of the post-test Invisible displacement trials support this idea, as apes' performance improved in those trials. The mean proportion of trials correct (0.60, SE = 0.05) was numerically higher than chance, and this difference approached significance ($T^+ = 61.50$, $N = 12$ [8 ties], $p = 0.087$). It should be noted that the apes had received many trials by this point, and being tired with the whole affair may have weakened their performance in the post-

test trials. This is consistent with the deterioration in True belief performance and the loss of differential looking across conditions in Version B.

I also directly compared apes' performance on Invisible displacement trials and False belief trials. Apes' pretest Invisible displacement scores were not significantly better than their False belief scores ($T^{\#} = 52.00$, $N = 13$ [7 ties], $p = 0.66$), likely because of the false belief element in the pretest Invisible displacement trials, as noted above. However, apes' performance in the post-test Invisible displacement trials was significantly better than their False belief performance, whether I collapsed across the two different types of trials or analyzed them separately (in all cases, $T^{\#} \geq 82.00$, $N \geq 13$ [ties ≤ 7], $p \leq 0.017$).

In addition, I compared False belief proportion scores in apes who passed the post-test Invisible displacement trials and apes who failed them (using a pass criterion for the Invisible displacement trials of at least three out of four trials correct). If difficulty with invisible displacement explained apes' poor False belief performance in general, then apes who failed these control trials should have lower False belief scores than those who passed. Mean False belief proportion scores for the two groups were 0.39 (SE = 0.09) and 0.30 (SE = 0.07) respectively, which were not significantly different from one another ($U = 39.50$, $N_{\text{passed}} = 9$, $N_{\text{failed}} = 11$, $p = 0.50$). There was also no significant correlation between post-test Invisible displacement scores and False belief scores ($r_s = 0.22$, $N = 20$, $p = 0.37$). In short, any difficulties apes might have had with invisible displacement were not solely responsible for their poor performance in the False belief

test. The false belief element of the task, rather than its non-mentalistic components, appears to have been the limiting factor on apes' performance.

General Discussion

In Study 1, I directly compared the performance of children and apes in a new nonverbal false belief test. The test was more species-relevant for the apes (especially for the chimpanzees) than previous ones (Call & Tomasello, 1999; O'Connell & Dunbar, 2003) because it was competitive rather than cooperative and did not require participants to understand communicative intentions. Nevertheless, the picture that emerged was consistent with earlier findings: whereas young 5-year-old children passed the False belief test, both ape species failed. Children responded differentially across True belief and False belief trials in both versions of the procedure, and in Version B their choices were unambiguously consistent with recognition of the competitor's false beliefs. They most often chose the container the competitor reached for in True belief trials but the other container in False belief trials. In contrast, apes most often chose the container the competitor reached for in both True and False belief trials, across both versions of the procedure. This strategy served them well in the pretest and True belief trials, in which the competitor always reached for the correct container; but it was disastrous in False belief trials, in which their performance was worse than if they had simply guessed.

The question that arises at this point is, did apes' poor performance in this test reflect a true lack of understanding of false beliefs, or could other factors have interfered with their performance? One might argue that the test was too difficult for apes for reasons unrelated to false belief understanding. It is possible, for example, that apes

found the competitor's reach so compelling that they could not inhibit a prepotent tendency to choose the container she reached for, even when they suspected she was wrong (Boysen & Berntson, 1995; Boysen et al., 1996). For two reasons, I do not think this was a significant problem in the current study. First, apes had no trouble ignoring the competitor's reach when they knew it was wrong in the Ignore-competitor control trials. Second, they had a couple of seconds to make a choice after the competitor began reaching, while the containers were being slid within their reach. Even if their initial impulse was to reach for the same container as the competitor, they had the opportunity to reverse their decision and choose the other container. They rarely did so, and certainly no more than children, who also rarely changed their choice. One might also argue that competing with humans, rather than conspecifics, was unnatural for the apes. This seems an unlikely explanation, given that, as we saw in Chapter 3, apes have succeeded in other tasks requiring some degree of mental-state understanding, even when interacting with humans rather than conspecifics (Call, Hare, Carpenter, & Tomasello, 2004; Hare et al., 2006; Hare & Tomasello, 2004; Melis et al., 2006). Other nonhuman primates have also succeeded in tasks in which they needed to infer humans' perceptual states in competitive situations (Flombaum & Santos, 2005; Vick & Anderson, 2003).

Along with active choice, I measured looking in the subset of trials in which participants chose the same container as the competitor. The positive results for apes in Version A of the procedure gave some indication that they may not have been entirely oblivious to the competitor's false belief states. In that version, apes looked at (i.e., oriented their face toward) the container the competitor was *not* reaching for in a greater

proportion of False belief trials than True belief trials. Precisely what or how much apes understood is still unclear. They appeared to recognize that something was going on in the False belief trials that should make them unsure of how to respond. It is possible that their looking responses were indicative of some level of false belief understanding. It is also possible that the deceptive attitude of the experimenter as she performed the switch made them suspect something underhanded was happening, without knowing exactly what that something might be. Or, granting the apes a little more understanding, they may have recognized that the competitor had to be ignorant of the reward's location because she had not witnessed the switch, yet she reached confidently as if she knew its location. The apes' uncertainty about what to do in these circumstances may have caused them to vacillate between options and look at both containers.

Another important issue is that the differential looking found in Version A was not replicated in Version B, in which the competitor had her back turned during the switch. Conceivably, the competitor's leaving the room in Version A could have brought the apes closer to understanding that she could not have possibly known about the location switch and therefore must believe the food to still be in its original location. Yet, it is curious then that children had an easier time recognizing the competitor's false belief in Version B than in Version A. Another possibility is that apes, by the time they received Version B of the procedure, had become tired or bored with the task, or confused by the repeated alternating back and forth between blocks of True belief and False belief trials.

The looking results for the apes are admittedly fragile and open to a number of interpretations, although it should be noted that the latter issue also applies to many

looking studies with children (e.g., Perner & Ruffman, 2005). Given that they were based on a reduced data set, on face orientation rather than actual eye direction, and were not replicated in Version B, a cautious interpretation is recommended. Nevertheless, they are important in pointing to a possible fruitful direction for future research into apes' understanding of the mind. It would be premature at this stage to do more than speculate on what the differential looking results might mean. It is important to first see if the effect replicates across different experimental procedures.

Apes' performance aside, some unexpected results for the children warrant discussion. For example, why was children's False belief performance in Version A of the procedure weaker than in Version B? One possibility is that children found the competitor's false belief more salient in Version B, in which they could see the deceptive act carried out, literally, behind the competitor's back. It may also have been easier for children to stay focused on the task and the switching of the containers without the distraction of the competitor leaving the room, as she did in Version A. Additionally, children received far more pretest trials in Version A than in Version B, and so they may have been losing interest in the task by the time the experimenter administered the test trials. Informal observations of children's behavior during testing suggested that these factors might have very well been an issue in Version A. Another possibility is that in Version B, but not Version A, participants received pretest Invisible displacement control trials, and these may have primed children to pay more attention to the deception during the test trials. Finally, I should note that while competitive tasks might be optimal for apes this does not necessarily make them optimal for children (which was not, in any

case, the goal for this study). Nevertheless, children did well in Version B, so the competitive nature of the task is an unlikely explanation for their poor performance in Version A.

Another issue is that while children as a group performed similarly in both the Competitive reaching test and the standard verbal Sally-Anne test, there was little correspondence in individual performance across the two tests. The fact that some individual children found the Competitive reaching test easier than the Sally-Anne test, and vice versa, is not so surprising, as the tests differed greatly on many dimensions. In the Sally-Anne test, participants passively observed the sequence of events, did not compete, saw the item as it was moved (and so knew directly its location), had to predict the protagonist's future action, and were not immediately rewarded for answering correctly. In the Competitive reaching test, participants actively participated in a competition, saw only the containers but not the item being moved, had to recognize that the competitor's current actions were incorrect, and won a reward if they responded correctly. Some combination of these factors could have very conceivably led to differences in individual performance across tests. While a number of researchers have found positive correlations in performance on different false belief tests (e.g., Call & Tomasello, 1999; Carlson et al., 2005; Hughes et al., 2000), some inconsistencies in individual performance—even across standard verbal tests or on the same test administered at two different times—are not uncommon in the literature (e.g., Charman & Campbell, 1997; Holmes, Black, & Miller, 1996; Mayes, Klin, Tercyak, Cicchetti, & Cohen, 1996; Naito, 2003). And those who have reported positive correlations have not

always controlled for the possible confounding effects of age (e.g., Call & Tomasello, 1999). Finally, in some studies in which different false belief tests have been administered to the same participants, correlational analyses were not reported (e.g., Keenan & Ellis, 2003; Ziv & Frye, 2003).

Inconsistencies in performance across tests are likely due to individual differences in children's false belief competence across different types of procedures. Some children may find verbal false belief tests easier than nonverbal ones, for example, while other children may find the nonverbal tests easier. Thus, failure to find positive correlations between different false belief procedures does not necessarily indicate that one or the other test is not measuring false belief understanding. Supporting this interpretation is the fact that even the two widely used standard tests (the Sally-Anne and Smarties tests) are sometimes inconsistent (e.g., Charman & Campbell, 1997; Holmes et al., 1996; Naito, 2003), yet both are nevertheless considered to be valid false belief tests. The issue of individual variation in performance across false belief procedures clearly demands closer attention (Müller, Miller, Michalczyk, & Karapinka, 2007), and also speaks to the importance of not relying on any one test to determine false belief competence in individuals.

To conclude, in Study 1, young 5-year-olds were capable of passing a new nonverbal, competitive false belief test that apes failed. While apes did show differential looking responses across True and False belief trials in one version of the procedure, this was not replicated in another version. Thus, there was no strong evidence that making false belief tests competitive helps apes to recognize others' false beliefs. In fact, there

may be reason to suspect that competition, and not just cooperation, could also be problematic for chimpanzees in false belief tests. I explored this possibility in Study 2.

STUDY 2 – CHANGE-OF-CONTENTS STUDY

In Study 2, I considered the possibility that explicit competition might be as problematic for chimpanzees as is explicit cooperation, although for different reasons. Keenan and Ellis (2003) point out that competition might activate evolved automatic response systems that could interfere with participants' ability to make deliberate, controlled choices. This could be especially disruptive in computationally complex tasks, such as monitoring which events in an unfolding sequence others have and have not witnessed, and then combining these separate pieces of information to judge their resulting belief states. Indeed, 3- to 4-year-old children's performance deteriorated when a false belief test was framed in terms of predator avoidance, another situation likely to activate evolved mechanisms (Keenan & Ellis, 2003). And while competition has been found to improve chimpanzees' performance in other tests of mental-state attribution (Hare et al., 2000; Hare et al., 2001, 2006) those tests may have been computationally simpler than false belief tests (Hogrefe, Wimmer, & Perner, 1986). Thus, although intended to aid performance, the competitive nature of the false belief test administered in Study 1 may have inadvertently hindered it. Chimpanzees may have performed poorly because competing is *too* natural for them. It is thus important to test them using a false belief test that is neither explicitly cooperative nor competitive. A more neutral false belief test would also increase the validity of cross-species comparisons by leveling the

playing field for chimpanzees and children. While children could have an advantage over chimpanzees in cooperative tasks, they might find competing with adults somewhat unusual or unnatural. For chimpanzees, competition could potentially help or hinder performance depending on the type of task and perhaps the social status of the participant relative to the competitor (i.e., dominant versus subordinate). A false belief test that is cast in a more neutral context and can be administered to both children and chimpanzees is therefore desirable.

There is a further problem with existing false belief tests (including the test in Study 1) that applies to research with both children and apes. Most of the nonverbal tests developed to date using active response measures are adaptations on the standard verbal Change-of-location paradigm, in which a false belief is instilled by displacing an object in someone's absence (Baron-Cohen et al., 1985). A greater variety of paradigms is needed. One possibility is a nonverbal version of the other standard verbal test, the Smarties test (Perner et al., 1987). Recall that in that test, children first discover that a candy box really contains some other item, and they must then predict what someone else will think is inside the box. With its focus on contents rather than location, the Smarties test provides a useful complement to the Sally-Anne test, as the two procedures can be administered to the same participants without seeming repetitive. As we saw earlier, however, although the tests are designed to measure the same underlying skill, they do not always yield consistent results (Charman & Campbell, 1997; Holmes et al., 1996; Naito, 2003). This is likely because each requires, in addition to false belief understanding, other skills that may vary across individuals. Thus, although apes have

performed consistently poorly in false belief tests using the Change-of-location paradigm, they might perform better with a different paradigm.

In Study 2, my goal was to test chimpanzees with a new nonverbal false belief test that was neither explicitly cooperative nor competitive and employed a Change-of-contents paradigm to instill a false belief. As in Study 1, this task also involved a change that happened unbeknownst to an experimenter, but it was a change in the *contents* of a container rather than in the location of the container. While this task required the experimenter to cooperate with the chimpanzees in terms of pushing a table forward so they could make a choice, this cooperative element was not central to the task itself. The experimenter's role was not to help the chimpanzee find the food but only to make the containers physically accessible. Chimpanzees did not need to understand communicative intent in order to solve the task.

In this study, I wanted to give chimpanzees every possible opportunity to demonstrate any false belief capacities they might have. Therefore, in addition to the *Main* test, I ran two follow-up tests aimed at making it easier for them to recognize the experimenter's false beliefs. The first was the *Duping* test, in which some of the chimpanzees were given personal experience being tricked regarding the contents of a container, as in the standard Smarties test (Perner et al., 1987). The second was the *In-room* test, in which the experimenter stayed in the room within view of the chimpanzees at all times but had her back turned during the switch (because this manipulation improved children's performance in Study 1). In addition, because I found some evidence in Study 1 (Version A) that apes might understand more about false beliefs than their

active choices suggested, I also measured chimpanzees' spontaneous looking here, again using the indecisive looking measure. Because Study 2 required extensive training over many sessions, the procedure was administered to a smaller group of chimpanzees with the plan to follow up with more individuals if there was any indication that greater statistical power might significantly alter the findings. As the number of participants was much smaller than in Study 1, I gave each participant a larger number of False belief trials, making it possible to test for statistically significant effects in individual performance.

Previous Study with Children

A similar test was previously carried out with 24 human children (Krachun, Carpenter, Call, & Tomasello, in preparation).¹⁰ In that test, an experimenter put a reward (either a sticker or a rubber stamp) into an opaque plastic Easter egg as children observed. She then put the egg into one of two chutes leading to openings in the bottom left and right sides of the box, where children could then go to retrieve the reward. During initial training trials, children learned that eggs with stickers inside always came out on the right side of the box and eggs with stamps inside always came out on the left. In the subsequent False belief trials, the experimenter put either a sticker or stamp into the egg but then left the area, whereupon another person sneakily switched the reward for the other type. When the experimenter returned, children had to predict that she would put

¹⁰ I was not directly involved in the child study except to perform some coding and data analyses and to write the manuscript (Krachun et al., in preparation) reporting both the child and chimpanzee studies.

the egg into the wrong chute, and they should thus go to the side of the box normally associated with the egg's *original* contents. In True belief control trials, the experimenter witnessed the switch and put the egg into the chute corresponding to its current contents.

The results of that study followed the typical developmental pattern (Figure 4.5).

Children performed poorly before the age of 4, very well by the age of 5, and at

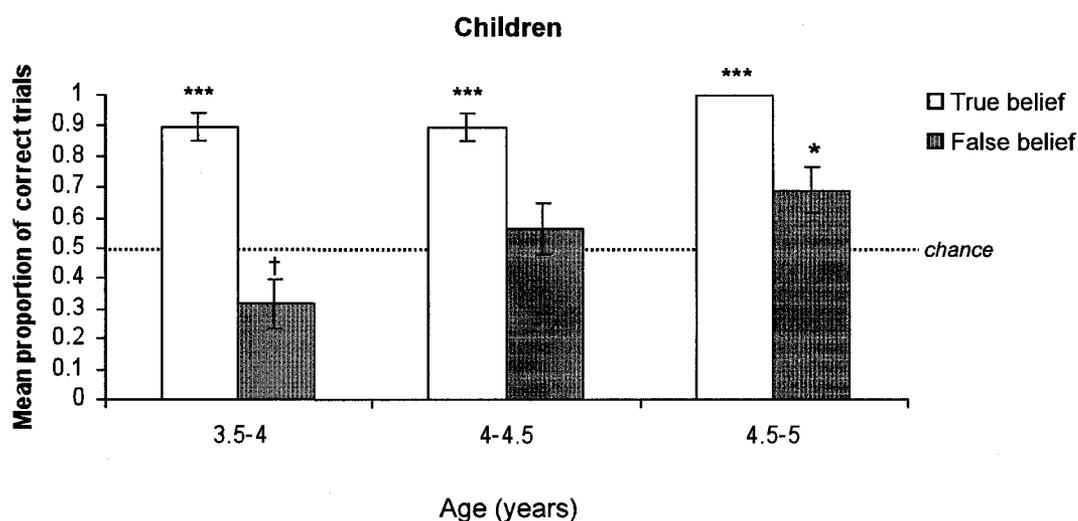


Figure 4.5. Mean proportion of correct trials (\pm SE) for children in Krachun et al.'s (in preparation) True and False belief tests. * $p < 0.05$, *** $p < .001$, significantly better than chance; † $p < 0.05$, significantly worse than chance

intermediate levels in between. However, also consistent with previous research (Carpenter et al., 2002), even the youngest children showed some signs of false belief understanding by switching strategies across the True and False belief conditions, rather than indiscriminately responding as if the experimenter always had a true belief.¹¹ The

¹¹ Note that children's performance in Krachun et al.'s (in preparation) study was positively correlated with a Sally-Anne test but it was not correlated with a Smarties test, and the Sally-Anne and Smarties tests were also not correlated with one another. Once

fact that children were able to pass that test (from which the current one was adapted) verifies that there were no fundamental problems with the basic procedure that would advise against adapting it for use with apes.

Current Study with Chimpanzees

Methods

Participants

Five chimpanzees housed together at the Wolfgang Köhler Primate Research Center (WKPRC) in Leipzig, Germany participated. The group included one male and four females with an average age of 10.2 years (range = 5–13 years, SD = 3.9). A sixth chimpanzee received training but was dropped from the study for failing to achieve the training criterion. Chimpanzees participated voluntarily, were not food deprived, and had water available *ad libitum*.

Design

Chimpanzees participated in a variable number of training sessions (12 trials each) until they reached criterion (see below). Before each test session, they were also given four refresher training trials and had to get at least three correct to be tested that day. They received a total of 20 False belief and 8 True belief trials over seven test sessions. Sessions 1–4 (Main test) each included four True belief *or* four False belief

again, this speaks to the importance of developing a variety of paradigms for testing false belief understanding.

trials (never both in the same session), with order counterbalanced across participants. Sessions 5–6 (Duping test) and Session 7 (In-room test) each included four False belief trials.

Materials

Rewards were grapes or banana slices. I did not use the box apparatus used with children in Krachun et al. (in preparation) because the chimpanzees appeared to be incapable of learning how the chutes worked, even after hundreds of training trials (Jensen, unpublished data). Instead, the chimpanzees' apparatus (Figure 4.6) consisted of a sliding table mounted below a Plexiglas window looking onto the chimpanzees' enclosure, a small yellow box with lid, and two larger, round, opaque plastic containers (one blue and one white). The blue and white containers were equivalent to the two chute openings in the children's apparatus in Krachun et al. (in preparation) and the yellow box was equivalent to the egg. Chimpanzees could choose a container by reaching through holes at the bottom of the window. The location of the reward was randomized with the constraint that it could not be in the same location for more than two consecutive trials.

Procedure

Chimpanzees were tested in a familiar enclosure by two female experimenters, E1 (a student assistant) and E2 (myself). Prior to testing, chimpanzees had to learn that grapes always went into the blue container on their left and bananas always went into the white container on their right. All sessions began with a two-trial warm-up in which E1

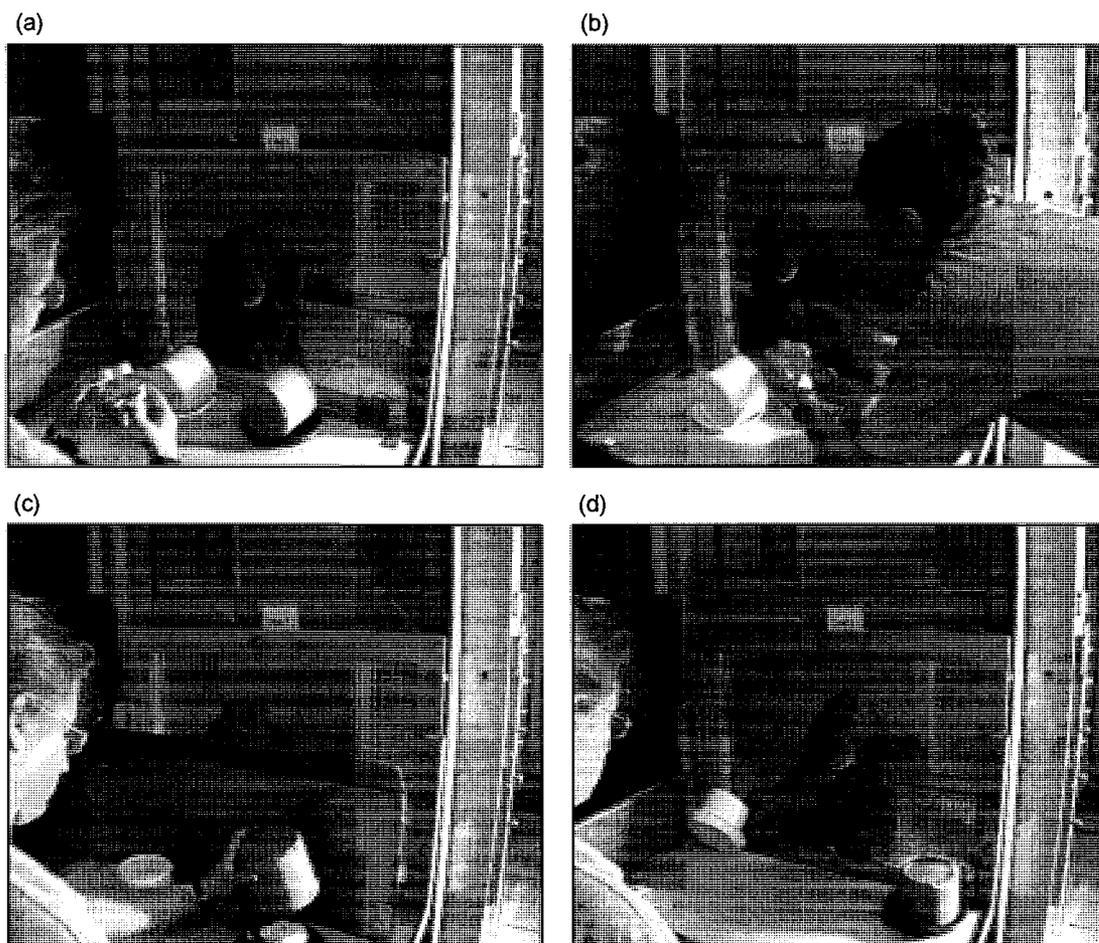


Figure 4.6. False belief procedure for Study 2: (a) E1 puts a grape into the box and then leaves; (b) E2 enters and sneakily switches the grape for a banana slice; (c) E1 returns and, because she thinks it still has a grape inside, hides the box in the 'grapes' container; (d) the chimpanzee incorrectly chooses the 'banana' container.

placed, in random order, a grape (in one trial) and a banana slice (in the other trial) into the appropriate container in plain view and then let the chimpanzee choose. Note, however, that in all subsequent training trials and test trials (described below), E1 *always* blocked chimpanzees' view with an opaque occluder before hiding the reward under the blue or white container. Thus, chimpanzees could only participate in this test if they were

able to learn, during training, the rule stated above: grapes go in the blue container, bananas in the white.

Training procedure. Training proceeded in three stages, with chimpanzees advancing from one stage to the next when they achieved at least 10/12 trials correct in two consecutive sessions. In stage 1, E1 held up the food (grape or banana) and then positioned an opaque occluder to block the chimpanzee's view. She hid the food in the appropriate container, removed the occluder, and let the chimpanzee choose. In stage 2, E1 introduced the yellow box. She held up the food, placed it into the yellow box, and closed the lid. She then positioned the occluder, hid the box in the appropriate container, removed the occluder, and let the chimpanzee choose. In stage 3, E1 placed the food into the yellow box, then put the box down onto the table and left the room, closing the door behind her. She stayed outside the room for 10 seconds (approximately the same amount of time she would be gone in the True belief and False belief trials), then returned and completed the trial as before. E2 also sat in the room off to one side during this last stage but did not participate in the procedure. Chimpanzees proceeded to the test trials when they passed all three training stages.

Main test trials. In False belief trials, E1 baited the yellow box and closed it while chimpanzees watched, then placed the box onto the table and left the room. E2 then approached the window, got the chimpanzee's attention, and switched the food inside the box for the other type of food. She did this sneakily and glanced repeatedly at the door to check if E1 was returning. When E2 was finished she sat down again and secretly signaled to E1 that she was finished by pressing a button on her own walkie-talkie, which

made E1's walkie-talkie beep. E1 then returned, positioned the occluder to block the chimpanzee's view, hid the yellow box in a container, and removed the occluder. If chimpanzees understood that E1 had a false belief about the contents of the box, they should choose the container corresponding to its *original* contents. The True belief procedure was identical to the False belief procedure except that E2 waited until E1 returned before switching the contents (and the switch was not sneaky), and E1 hid the box according to its *current* contents.

Refresher session. For some participants, several weeks elapsed between the Main test and the follow-up tests. I therefore administered another stage-3 training session in between, to ensure that chimpanzees remembered the correct location for each type of reward. Individuals had to get at least 10/12 trials correct to proceed to the follow-up tests. All chimpanzees achieved this criterion.

Duping test (first follow-up). The two-session Duping test investigated whether personally experiencing an unexpected change of contents (just as children personally experience unexpected contents in the Smarties test) would improve chimpanzees' performance in False belief trials. Giving chimpanzees first-hand experience with being mistaken about the contents of the box might help them to grasp the possibility that others could also be wrong about its contents. Additionally, it might prompt them to attend closely to the procedure instead of relying on the simple strategy of always choosing the blue container for grapes and the white container for bananas, as they learned during training.

Three randomly chosen chimpanzees received the duping experience and the remaining two acted as non-duped controls. In the Duping trials, E1 baited the yellow box as the chimpanzee observed, then put it down in the middle of the table and left the room. The blue and white containers were not used in these trials. E2 then pulled a curtain across the window to block the chimpanzee's view. She switched the food inside the box for the other type, put the box back onto the table, opened the curtain, returned to her seat, and secretly signaled to E1 that she was finished. E1 then returned, opened the box, and gave the chimpanzee the food (which was, of course, different from that the chimpanzee had seen placed into the box). The Non-duping control trials were identical to the Duping trials except that E2, behind the curtain, opened and closed the yellow box without switching its contents. The duped group received two Duping trials and two Non-duping trials in each session (alternating, always beginning with a Non-duping trial). The control group received four Non-duping trials in each session. For each group, these were followed immediately by four False belief trials that were exactly the same as in the Main test.

In-room test (second follow-up). In this final session, all chimpanzees received four False belief trials the same as in the Main test except that E1 remained in the room during the switch. After baiting the yellow box and placing it on the table, E1 stood up, turned around, and walked across the room, muttering to herself and pretending to be absorbed in a control panel on the opposite wall. E2 then quietly approached the window and, while the chimpanzee observed, sneakily switched the contents of the box, periodically checking to make sure E1 was not looking. After E2 was finished she sat

down again and secretly signaled E1, who returned and carried out the remainder of the trial as before.

Coding and Analyses

Chimpanzees' choice of container, defined as the first one they touched, was unambiguous and therefore coded live. I also coded from videotapes whether or not chimpanzees looked back and forth in apparent indecision between the containers before choosing. They were considered to have done so if they looked at the ultimately unchosen container *two or more* times before settling on the other container. I used this criterion (looking at least twice, rather than at least once as in Study 1) because chimpanzees almost always looked at both containers at least once before choosing in Study 2, making this behavior uninformative here.

I coded looking for all eight True belief and eight False belief trials of the Main test. Looking was not coded for the follow-up tests because there were no corresponding True belief trials with which to compare them. The mean number of trials in which looking back and forth occurred was then compared across conditions. I also investigated looking in just a particular subset of these data, as in Study 1. Recall that in that study, True belief and False belief trials were examined on an equal footing by using only the subset of trials in which apes chose the same container as the competitor, as they had done throughout the pretest/training period. In keeping with this logic, I analyzed looking for just the trials in which chimpanzees chose the container corresponding to the current contents of the yellow box, as they had learned during training. Because each chimpanzee had a different number of data points for this analysis, I used proportion scores. Statistical

analyses for all measures were nonparametric and all p values reported below are exact and two-tailed.

Reliability

To determine inter-rater reliability, an independent coder blind to condition and naïve to the hypotheses of the study coded 25% of test trials for each measure. Reliability was excellent, with perfect agreement for choice and a Cohen's kappa of 0.80 ($p = 0.001$) for looking.

Results and Discussion

The five chimpanzee participants took an average of 25 sessions to learn the unique locations for the two types of reward (range = 9–45 sessions, mean = 24.4, SE = 6.29). A sixth chimpanzee was still choosing randomly after 50 sessions and was therefore not given the False belief test. There were no effects of order of condition on True belief or False belief performance (Mann-Whitney U tests: $U \geq 0.50$, $N = 5$, $p \geq 0.30$ in both cases) and so I collapsed across order for all analyses. Because the Main test and two follow-up tests included varying numbers of False belief trials, I used proportion scores for easier comparison across tests.

Main Test

Choice. Chimpanzees succeeded in far more True belief trials than False belief trials (Figure 4.7), and the difference in the group means approached significance (Wilcoxon test: $T^+ = 15.00$, $N = 5$, $p = 0.063$). Chimpanzees' performance in False belief trials was not significantly different from the chance proportion of 0.50 ($T^+ = 6.00$, $N = 3$

[2 ties], $p = 0.25$), although True belief performance did approach significance ($T^+ = 15.00$, $N = 5$, $p = 0.063$). Increasing the number of participants to gain power would thus likely yield significant results for True belief but not False belief. I also examined False belief performance as compared to the True belief complement value, as in Study 1 (see Study 1, Experiment 1, Coding and Analyses). There was no significant difference between the proportion of successful False belief trials and the proportion of incorrect True belief trials ($T^+ = 4.50$, $N = 3$ [2 ties], $p = 0.75$), suggesting that chimpanzees did not switch strategies across conditions.

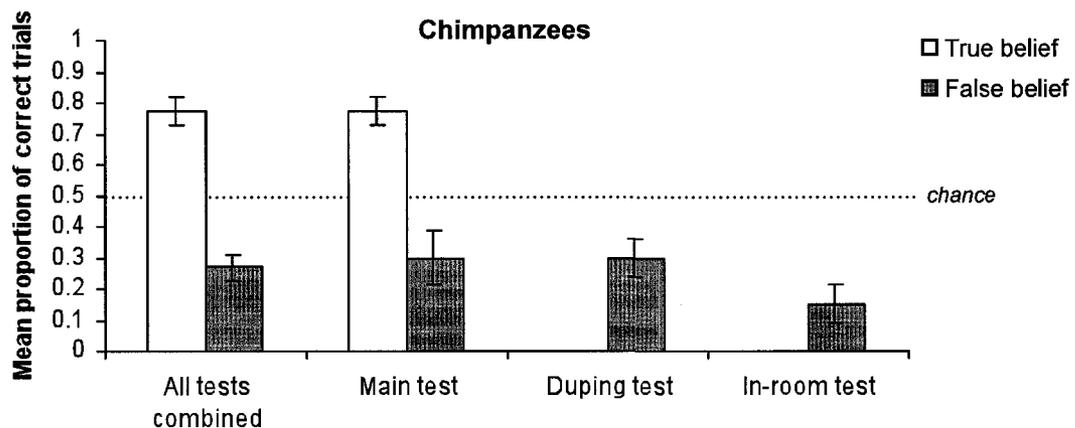


Figure 4.7. Mean proportion of correct trials (\pm SE) for chimpanzees in Study 2.

Performance at the individual was consistent with the group results. All chimpanzees' True belief scores were higher than their False belief scores; and all True belief scores were above the chance value whereas all False belief scores were at or below chance (see Table 4.1). With only five chimpanzees, formal statistical analyses on

the first True belief and False belief only for each participant would not be very informative. However, these trials followed the same overall pattern: four out of five chimpanzees were successful in their first True belief trial whereas only two out of five were successful in their first False belief trial.

Table 4.1

Proportion of correct True belief (TB) and False belief (FB) trials for individual chimpanzees in the Main test and follow-up tests of Study 2.

Chimpanzee	Main test		Duping test	In-room test	All tests
	(8 TB trials)	(8 FB trials)	(8 FB trials)	(4 FB trials)	(20 FB trials)
Alex	0.75	0.25	0.25 ^a	0.00	0.20
Jahaga	0.88	0.50	0.38 ^a	0.00	0.35
Fifi	0.75	0.13	0.13 ^a	0.25	0.15
Trudy	0.63	0.50	0.25 ^b	0.25	0.35
Annett	0.88	0.13	0.50 ^b	0.25	0.30

^a = duped chimpanzee

^b = non-duped chimpanzee

Looking. The positive looking results found in Study 1 (Version A of the procedure) were not replicated here: chimpanzees looked back and forth between containers in a similar proportion of True belief (mean = 0.36, SE = 0.12) and False belief (mean = 0.33, SE = 0.09) trials ($T^+ = 4.00$, $N = 3$ [2 ties], $p = 1.00$). In the first True belief and False belief trials only, the proportion of chimpanzees showing the looking response was the same in both conditions (0.40). Finally, when I examined only the subset of True belief and False belief trials in which chimpanzees chose according to the current contents of the yellow box, the mean proportion of trials with looking was still

similar across the True belief (mean = 0.38, SE = 0.14) and False belief (mean = 0.33, SE = 0.06) conditions ($T^+ = 10.00$, $N = 5$, $p = 0.63$).

Follow-up Tests

In the Duping test, personally experiencing an unexpected change of contents made no difference to chimpanzees' False belief performance. The three duped chimpanzees (see Table 4.1) did no better in the Duping sessions (mean proportion correct = 0.25, SE = 0.07) than in the Main test sessions (mean = 0.29, SE = 0.11) ($T^+ = 1.00$, $N = 1$ [2 ties], $p = 1.00$). Their performance was also not significantly different from the two non-duped chimpanzees, whose average proportion of trials correct in these follow-up sessions was 0.38 (SE = 0.13) ($U = 1.50$, $N = 5$, $p = 0.50$). In the In-room test, chimpanzees' mean proportion of trials correct (0.15, SE = 0.06) was actually lower than in the Main test, although this difference was again not significant ($T^+ = 12.00$, $N = 5$, $p = 0.25$).

Main and Follow-up Tests Combined

Because each chimpanzee received 20 False belief trials in total, including the Main test and two follow-up tests, it was possible to check for statistically significant effects in individual performance using the Binomial test. Two chimpanzees (Alex and Fifi) did significantly worse than chance ($p < 0.01$ in both cases), while all other chimpanzees were at chance performance. Furthermore, the group mean for all 20 False belief trials (0.27, SE = 0.04) approached being significantly worse than chance ($T^+ =$

15.00, $N = 5$, $p = 0.063$). In short, there was no evidence of any kind to suggest that chimpanzees recognized E1's false beliefs.

In creating nonverbal false belief tests, researchers have relied almost exclusively on change-of-location procedures inspired by the verbal Sally-Anne test. In this study, I expanded the repertoire by introducing a new test that focused on contents rather than location. Additionally, this test was neither explicitly cooperative-communicative nor competitive. It thus avoided the potential negative influences of these elements on apes' performance. Nevertheless, the results were consistent with previous research: the chimpanzees displayed no evidence of false belief understanding, neither in their active choice responses nor in their spontaneous looking behavior. The Study 1 finding of differential looking responses in apes across False belief and True belief trials was thus not replicated here. It is possible, however, that in the current study apes' uncertainty was simply too high in *both* the True and False belief conditions. The fact that they almost always looked at both containers at least once before responding in both conditions supports this interpretation.

GENERAL DISCUSSION OF STUDIES 1 AND 2

In Studies 1 and 2, I tested apes with two new nonverbal false belief tests, one of them including two variations on the same procedure. The tests I used were more species-relevant than previous ones because they did not require participants to understand cooperative or communicative intentions. The first test was openly competitive and the other was cast in a neutral social context. The apes failed both tests. There now exist a

number of separate studies in which apes have been unsuccessful in nonverbal false belief tests that human children have passed. In addition, since my own research was completed, a new study has emerged in which chimpanzees competed with conspecifics in a task requiring them to recognize another chimpanzee's false belief about the location of a food item (Kaminski, Call, & Tomasello, 2008). Consistent with my findings, chimpanzees also failed that test, although they did well in a similar test requiring them to understand what others knew or did not know based on past visual access. This provides further evidence that chimpanzees have not done poorly in the false belief tests to date because they were interacting with humans rather than conspecifics.

It appears that in trying to determine what apes might understand about the mind, testing for an understanding of others' false beliefs may be aiming too high. While apes have demonstrated an understanding of what others see and know in some situations, perhaps they simply do not possess the cognitive sophistication necessary to recognize that when others do *not* see some crucial event, what they 'know' may be wrong. False beliefs may exceed the limits of chimpanzees' mentalizing capacities.

In the remainder of my thesis, I focused on an area that has received even less attention in research on ape mentalizing: their abilities with respect to their own mental states. As very little work has been done in this area, I began simply in Study 3 by testing chimpanzees' basic understanding of their own visual perceptual experiences. It was important to establish that chimpanzees have good visual perspective taking abilities with regard to themselves before moving on in Study 4 to investigating their ability to recognize when their own perceptions are mistaken. While chimpanzees failed the False

belief tests in Studies 1 and 2, the possibility remained that they could be capable of recognizing when they themselves were experiencing false perceptual states. In other words, chimpanzees' understanding of the more complex aspects of the relationship between seeing and knowing could be greater when dealing with their own mental experiences than when dealing with others'.

CHAPTER 5 – VISUAL PERSPECTIVE TAKING FOR SELF

STUDY 3 – PERSPECTIVE-SHIFTING STUDY

Up to this point, I have been focusing on what chimpanzees understand about others' mental states. Studies 1 and 2 of this thesis, together with previous research, suggest that chimpanzees are capable of recognizing what others see and know but not what they believe—at least not when their beliefs are false. In this chapter, I move on to examining how much chimpanzees understand about their *own* mental states. Very little research has been carried out in this area. Thus, although I began my study of chimpanzees' understanding of other minds by testing a relatively advanced capacity (false belief understanding), it was necessary to back up somewhat in examining what they understand about their own mental states. I therefore began by testing chimpanzees' basic understanding of their own visual perceptual experiences.

As we saw in Chapter 3, Call and Carpenter (2001) found that chimpanzees discriminated between their own states of seeing versus not seeing and knowing versus not knowing, just as they discriminate between these states in others. This was evident in the chimpanzees' looking into tubes for hidden food more often when they had not seen it hidden than when they had. However, Call and Carpenter's (2001) study was limited in some important respects. In order to see the food, apes merely needed to crouch downward slightly to peer in through the front of the tubes. It is difficult to know whether such effortless responses were purposeful, thoughtful attempts to gather knowledge or simple automatic foraging behaviors. More convincing evidence would be if participants

had to expend more concerted, deliberate effort to locate the food and had to choose the appropriate action from among a number of possible responses. Additionally, Call and Carpenter's (2001) study did not test apes' understanding of the importance of *particular* visual perspectives, as there was only one possible vantage point for viewing the food. And in other studies in which apes moved around barriers to see the target of an experimenter's gaze (Bräuer et al., 2005; Tomasello et al., 1999), they did not have to move immediately to the correct position to be considered correct. They could have moved to various locations before finding the one that allowed them visual access to the target, making it impossible to tell if they anticipated the correct angle before moving. Finally, in these and other visual perspective taking studies to date (e.g., Hare et al., 2000; Hare et al., 2001), the apes could have used the gaze direction of the social partner to help them solve the task. It is thus unknown whether apes could recognize the visual perspective required to see an object without this social cue. In this study, I sought to address these limitations and thereby extend what is known about chimpanzees' abilities regarding their own visual perceptual experiences. Having a sense of what apes are capable of at more fundamental levels with regard to their own mental states will provide a better foundation for investigations of their more advanced abilities, which is the subject of Chapter 6.

My first objective in Study 3 was to replicate Call and Carpenter's (2001) finding that chimpanzees responded differentially when they had not seen a hiding event than when they had, indicating that they discriminated between their own states of knowing and not knowing. Information seeking behavior—namely, shifting position to gain visual

access to the food when they had not witnessed it being hidden more often than when they had—was used as an operational measure of this discrimination. My second and main objective was to investigate chimpanzees' understanding of what they would or would not be able to see from different visual perspectives.¹² Specifically, I asked if chimpanzees could adopt the correct of several different visual perspectives within an enclosure to gain visual access to occluded food.

My task was a variation on Call and Carpenter's (2001) tubes paradigm. It involved presenting chimpanzees with a collection of open containers, one of which contained a hidden food item. Chimpanzees were allowed to witness the hiding of the food in some trials but not others. Before making a choice, chimpanzees were given the opportunity to look into the openings in the containers to locate the food. I used a variety of different container types, each of which provided visual access to its interior from a unique viewing angle. Chimpanzees had to recognize which of a number of possible visual perspectives within their enclosure would be the appropriate one in any given trial, depending on the type of container used. Thus, unlike Call and Carpenter's (2001) task, in which there was only one possible response participants needed to make to see the food, my task required chimpanzees to choose from among a number of possible responses, only one of which would be successful in locating the food. The movements they needed to make to see into the containers also required them to expend more

¹² When I write about chimpanzees' understanding of what they would be able to see from a different perspective, I mean more precisely their understanding of *whether or not* they would be able to see something (i.e., Level 1 perspective taking). This was also Hare et al.'s (2000) meaning.

deliberate effort than simply crouching to look into a tube. Thus, if chimpanzees responded correctly, I could be more confident that it was because they knew not just that they needed to move in order to see the food, but also precisely where they needed to go.

My project also differed from earlier perspective taking studies employing visual occluders in two important respects. First, I required chimpanzees to demonstrate their understanding of visual perspective in the absence of any gaze cues. Thus, this task demanded that chimpanzees have a more general appreciation of visual perspective than previous studies could demonstrate. Second, unlike Bräuer et al. (2005) and Tomasello et al. (1999), I considered chimpanzees to have made the correct response if and only if the very first response they made in an attempt to see the hidden food was the appropriate one for the particular type of container used. Under natural conditions, chimpanzees clearly move around to gain visual access to objects blocked from their view; however, whether or not they know which particular visual perspectives they should take in advance of taking them is not clear. By requiring participants to adopt the correct visual perspective as their initial response, I could be more confident that they were anticipating which perspective would gain them visual access to the occluded food.

Based on the results of past research (Call, 2005; Call & Carpenter, 2001), I expected chimpanzees to look into the containers before choosing more often when they had not witnessed the experimenter hide the food than when they had. And given how well they have performed in studies in which they had to judge what others could or not see in the presence of occluding barriers, I predicted that they would have little trouble anticipating where they had to put themselves in order to see around occluders of various

sorts. Experiment 1, the Perspective-shifting test, was the main test of chimpanzees' capacities; Experiments 2 and 3 were designed to rule out alternative, low-level explanations.

Experiment 1 – Perspective-shifting Test

I first gave chimpanzees a *Tubes* task similar to that used by Call and Carpenter (2001), both to introduce them to the paradigm of searching for food in open containers and also to see if I could replicate the basic finding that chimpanzees looked more often in Unseen than in Seen trials. The Perspective-shifting test that followed constituted the main experimental component. In those trials, I tested whether chimpanzees anticipated the specific perspectives they needed to adopt to see into various types of containers. Before and after the Perspective-shifting test, I also measured baseline levels of responding when food was hidden in completely closed, opaque containers (the *Pretest baseline* and *Post-test baseline*). This allowed me to compare chimpanzees' frequency of looking in the Perspective-shifting test with their frequency of looking when the food was not potentially visible from any viewing angle. If they were making deliberate attempts to see the food when they did not know its location in the Perspective-shifting test, we would expect to find differential looking across Seen and Unseen trials in that test but not in the baseline sessions.

Methods

Participants

Ten chimpanzees housed socially at the Wolfgang Köhler Primate Research

Center (WKPRC) in Leipzig, Germany participated: two males and eight females, all captive-born, five hand-reared and five mother-reared, ranging in age from 5 to 30 years old at the start of the study (mean = 13 years, SD = 8.1). All chimpanzees had taken part in a variety of social and physical cognition experiments, with seven participating in a study 6 years earlier that involved searching in tubes for hidden food (Call, 2005).

Chimpanzees were not food deprived and had access to water *ad libitum*.

Design

Chimpanzees received, in succession, 12 Pretest baseline trials, 18 Tubes trials, 36 Perspective-shifting trials, and 12 Post-test baseline trials. In half the trials chimpanzees witnessed the baiting of the containers (Seen trials) and in half their view was blocked (Unseen trials).¹³ In the Perspective-shifting test, each of the three container types was used the same number of times in each condition. Baiting condition, food location, and container type were randomized with the following constraints: there could not be more than four consecutive Seen or Unseen trials, the food could not be in the same location for more than three consecutive trials, and in the Perspective-shifting test the same container type could not be used for more than two consecutive trials. Chimpanzees received a variable number of trials per session depending on their availability and willingness to participate. The modal number of trials per session was 12 for the baseline

¹³ The first two chimpanzees tested received only six Unseen (and no Seen) pretest baseline trials because I decided only after testing them to include Seen trials in the baseline.

sessions (range = 1–12), 9 for the Tubes trials (range = 9–18), and 18 for the Perspective-shifting test (range = 6–27).

Experimental Set-up

Testing took place in two different testing rooms, each having a three-sided windowed booth (105 cm wide, 90 cm deep) set in between two adjoining enclosures (Figure 5.1). The windows extended to 202 cm above floor level in one testing room and 272 cm in the other. The bottom section of the windows on the left and right sides of the booth comprised a 2.5-cm-thick Plexiglas panel (68 x 49 cm) inside a metal frame. A table (80 x 52 cm) with a sliding section on top (80 x 35 cm) was mounted below one of these panels, allowing the containers to be slid within or out of chimpanzees' reach. Chimpanzees chose a container by poking a finger through one of three holes (diameter = 3.5 cm) spaced evenly across the bottom of the panel. This location was labeled the *home window*.

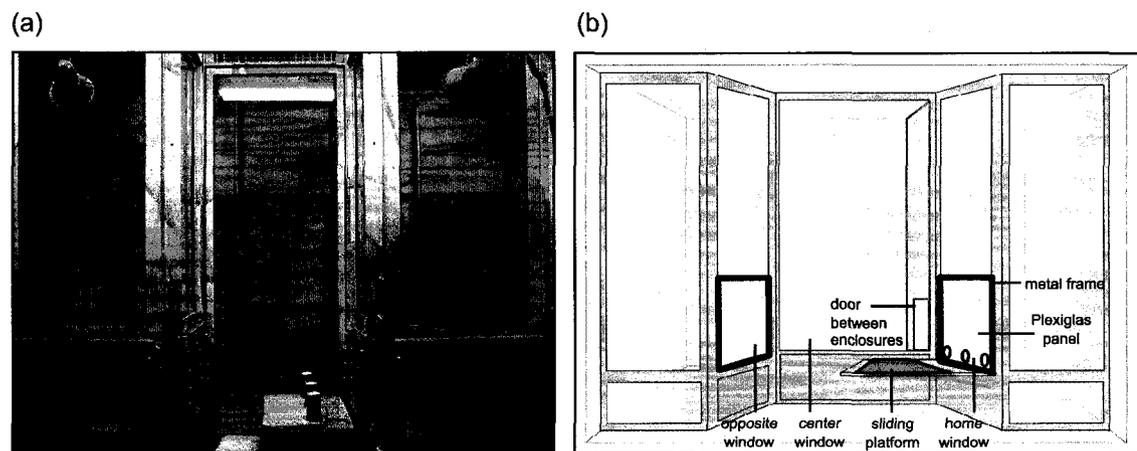


Figure 5.1. (a) Photograph of testing room and (b) schematic diagram with the sides of the booth fanned outward to make the windows visible.

Materials

Baseline containers. Three identical opaque plastic boxes (16 x 11.5 x 4.5 cm) were used for the Pre- and Post-test baseline trials. The boxes had lids and were completely closed, so they did not allow visual access to the food from any viewing angle. The chimpanzees had become familiar with such boxes in previous object-choice studies.

Tubes. For the Tubes task, I used three identical rectangular tubes constructed of 1-cm-thick grey plastic (30 cm long with a 5 x 5 cm opening), similar to those used by Call and Carpenter (2001). The tubes were baited through an opening at the back, and duct tape occluded the top half of the front opening of each tube so that chimpanzees had to clearly lower their head to see inside.

Perspective-shifting test containers. There were three container types (shown in Figure 5.2) for the Perspective-shifting test: (a) cylindrical containers open only at the top and painted white inside for better visibility, (b) triangular containers open only at the sides, and (c) trapezoidal containers open only at the back. The cylinders were cut from grey PVC piping and the triangles and trapezoids were constructed from 1-cm-thick grey plastic. There were three of each type of container, and in any given trial I always used three identical containers. The same containers were used for Seen and Unseen trials so that I could directly compare across conditions. Cylinders could only be seen into from above, triangles only from the sides, and trapezoids only from the back.

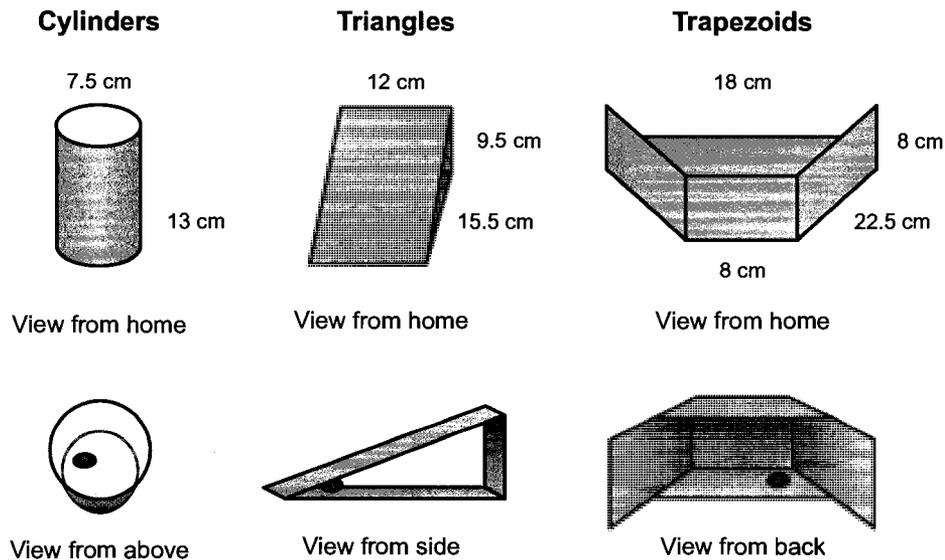


Figure 5.2. Perspective-shifting test containers for Study 3 as viewed from the home window (top row) and from the unique perspective allowing visual access to the food for each type of container (bottom row).

Procedure

Familiarization. Chimpanzees were first familiarized with the properties of the tubes and the Perspective-shifting test containers. The containers were placed in a pile on the floor of an enclosure, and the chimpanzee was then let into the enclosure alone for 10–20 minutes. After completing the Tubes task chimpanzees were again given exclusive access to the Perspective-shifting test containers for 10–20 minutes to refresh their memories. All chimpanzees showed interest in the containers by inspecting, biting, licking, sniffing, carrying, hitting, or throwing them. Interest waned quickly, however, and when the test began chimpanzees no longer appeared to consider the containers to be interesting objects in their own right. The familiarization phase ensured that chimpanzees

knew where the openings in the different types of containers were located. Once they had this knowledge, I could then present them with the containers in position on the platform and see if they recognized specifically where they had to position themselves to see into the various openings.

Baseline, Tubes, and Perspective-shifting test trials. Chimpanzees were tested in familiar enclosures by a female experimenter (myself). The same basic procedure was used for all trials. The experimenter (E) sat facing the home window and placed three identical containers onto the table, spaced evenly apart. She showed chimpanzees that the containers were empty and then placed a food item into one of them. In Seen trials chimpanzees were allowed to watch as E did so, and they thus saw which container she baited. However, in Unseen trials, E first blocked chimpanzees' view by positioning opaque screens at the home and center windows of the booth (the screen at the home window had a lip on top that also obstructed chimpanzees' view from above). If chimpanzees moved to the opposite window during baiting or tried to see around the screens, E blocked the view of the containers with her body and waited until the chimpanzee moved away before proceeding. Thus, in Unseen trials, chimpanzees could only know where the food was located by waiting until E was finished the baiting process and then moving to the appropriate location to look inside the containers used in that trial.

When baiting was complete, E exited the booth and removed the screens in Unseen trials (the screen at the center window first and then the screen at the home window, in quick succession). So as not to inhibit chimpanzees' movements, she sat or stood with her back to the booth for 20 seconds (60 seconds in baseline trials). She then

re-entered the booth and slid the containers forward so chimpanzees could choose. When chimpanzees chose correctly E gave them the food inside the container. Otherwise, E showed them that the container was empty and then removed the food from the correct container and returned it to a nearby bucket.

Coding and Analyses

There were three dependent measures: *choice of container*, *first look*, and *first move*. Choice was coded live and looking and moving were coded from videotapes.

First look. Of most interest in the Perspective-shifting test was the first look measure, defined as the very first attempt chimpanzees made to look into the containers. Three possible responses were coded: UP, SIDE, and OPPOSITE, as described below.

UP: Chimpanzees moved upward and looked down at the containers from above. Doing so provided visual access into the cylinders but was ineffective for the other containers. Chimpanzees' head had to reach or exceed the height of the metal frame around the Plexiglas panel (see Figure 5.1). An UP look could also occur at the center or opposite window, but only if it was not preceded by a look through the lower portion of that window (i.e., below the level of the metal frame).

SIDE: Chimpanzees clearly examined the containers from the side by moving close to the center window and lowering their head. Alternatively, at the home window, they could lean far and low to one side (i.e., into the bottom left or right corner of the window) and examine the sides of the containers from that position. In this case, their head had to reach the edge of the metal frame and their chin had to reach the bottom of

the frame for a SIDE look to be coded. This perspective allowed chimpanzees to see into the triangles but not into the other types of containers.

OPPOSITE: Chimpanzees clearly examined the backs of the containers, either by moving to the opposite window or to the extreme far end of the center window. In the latter case their head had to reach the far edge of the center window frame. Chimpanzees could see into the trapezoids from this perspective but not into the other types of containers. Note that to get from the home to the opposite window chimpanzees obviously had to move past the center window. This was only counted as a SIDE response if chimpanzees clearly hesitated or lowered their head to examine the containers before continuing on to the opposite window.

For all three responses, chimpanzees had to come close to the window (within about half a meter) for a look to be scored. Coding began at the moment the food was potentially visible. In the Seen trials this was when E had completed the baiting and removed her hand from the opening in the container. In Unseen trials, my use of screens to block chimpanzees' view caused the food to become potentially visible at slightly different times for each type of container. For trapezoids, it was as soon as E moved out of the booth so that she no longer blocked chimpanzees' view through the opposite window; for triangles, it was when E next removed the screen from the center window; and for cylinders, it was when E finally removed the screen from the home window (recall that this screen had a lip blocking the view from above). Note that I could not begin coding while all screens were still in place because they obscured my view of the chimpanzees' faces, making it often impossible to tell if they were attempting to look into

the containers. On the other hand, if I waited until all screens were removed I might have missed looks that occurred in the interim (e.g., when chimpanzees ran over to the opposite window to look into the trapezoids after E had exited the booth but had not yet removed the screens from the other windows). Beginning coding as soon as the food was potentially visible was thus a compromise between these two options.

For Seen tubes trials, I also coded looks into the tubes during the couple of seconds it took E to perform the baiting. This was because chimpanzees could look into the tubes while E deposited the grape into the opposite end. Looking during baiting was not an issue with the other containers because chimpanzees could not see the food until E had removed her hand.

In some cases (19% of trials) chimpanzees had already adopted the UP, SIDE or OPPOSITE perspective before coding began, for example because they had moved there as E was leaving the booth or removing the occluding screens. In these cases, if chimpanzees then clearly examined the containers from that perspective once coding began, this was coded as the first look. In addition to the first look, I also noted whether chimpanzees looked from any other perspective throughout the remainder of the trial. If chimpanzees did not attempt to look into the containers from any of the three possible perspectives at any point in the trial (for example, they remained seated at the home window, lay on the floor, or moved about the enclosure without attending to the containers), I coded this as no response.

First move. The looking measure was somewhat conservative because chimpanzees had to not just move to the correct location but also clearly examine the

containers. Instances in which chimpanzees quickly glanced at the containers from the incorrect perspective could have therefore been missed. To address this, I also scored where chimpanzees first moved irrespective of whether or not they looked. The same criteria were used as for first looks except that chimpanzees did not have to approach the window nor visually examine the containers. However, for moves that occurred within the confines of the home window (i.e., UP moves that did not extend beyond the top of the metal frame and SIDE moves into the lower left or right corner of the window) the chimpanzee had to at least be oriented forward. As with the first look measure, I began coding from the moment the food was potentially visible to the chimpanzee. Because it was possible to see where chimpanzees were moving even with all screens in place, I also did a second coding beginning from the moment E finished the baiting and before she had removed any of the screens.

Because looks were irrelevant for this measure, if chimpanzees had already adopted the UP, SIDE or OPPOSITE perspective before coding began then the next location they moved to was coded as the first move. (Note that this meant that in some trials a look could be coded when a move had not been coded.) A few locations within the enclosures were considered to be neutral, in the sense that chimpanzees could not see the food inside the containers from those locations. For example, chimpanzees occasionally waited out the trial on the threshold of the doorway between the two enclosures or on a raised bench at the back wall (they had no reason to remain seated at the home window throughout the entire trial, and some of them simply preferred to sit in other locations). These locations were thus counted as equivalent to the home position

and were not coded as moves. Finally, any trials in which chimpanzees either did not move at all, or moved only to one or more of the neutral positions, were coded as no response.

All statistical tests were non-parametric and reported p values are two-tailed unless otherwise noted. One-tailed tests were used only when results from previous research allowed strong directional predictions.

Reliability

An independent coder naïve to the hypotheses of the study coded 20% of test and baseline trials for choice of container, looking versus not looking, number of different kinds of look (none, one, or more than one kind), first look, and first move. Cohen's kappas were 1.00, 0.80, 0.76, 0.84 and 0.77, respectively ($p \leq 0.001$ in all cases). A second naïve coder judged a third of Tubes trials for chimpanzees' choice and whether or not they looked. Kappas were 1.00 and 0.65,¹⁴ respectively ($p \leq 0.003$ in both cases). All trials were chosen randomly with the constraint that half be Seen and half Unseen.

Results and Discussion

Tubes Task

The Tubes task confirmed that the chimpanzees understood the task of searching in containers for hidden food and were motivated to do so. Looking was in fact at ceiling levels, with chimpanzees looking into the tubes in 100% of Unseen trials and an average

¹⁴ The main and reliability coders agreed on 58/60 of trials, but the kappa was only moderate because responses were heavily skewed in one direction (i.e., there was a positive looking response in almost every trial).

of 89% (SE = 5%) of Seen trials. This difference, although not large, was statistically significant in a one-tailed exact Wilcoxon test ($T^+ = 15.00$, $N = 5$ [5 ties], $p = 0.031$). Frequency of looking was much higher than in Call and Carpenter (2001). This can be explained by changes I made to their procedure to make it more comparable to my Perspective-shifting test. First, in my test the tubes were 50 cm above floor level, making them easier to look into than Call and Carpenter's tubes, which were at a height of 35 cm. Second, I had a 20-second delay between baiting and letting chimpanzees choose, whereas Call and Carpenter had a 5-second delay or no delay at all. My chimpanzees therefore had far more time to look into the tubes than theirs, and they may have done so because they became bored or forgot where the food was, or simply to double-check since no other response was possible at that time. If we exclude any looks that happened beyond the first 5 seconds, chimpanzees looked, on average, in 60% (SE = 10%) of Seen trials and 72% (SE = 11%) of Unseen trials ($T^+ = 29.50$, $N = 9$ [1 tie], $p = 0.22$, one-tailed), which is more similar to Call and Carpenter's 5-second delay condition (2001, Experiment 1) in which chimpanzees looked in approximately 52% of Seen trials and 87% of Unseen trials.

Perspective-shifting Test

As Figure 5.3 shows, in the baseline sessions looking was comparable across Seen and Unseen trials (Pretest: $T^+ = 11.00$, $N = 6$ [2 ties], $p = 0.88$; Post-test: $T^+ = 29.00$, $N = 9$ [1 tie], $p = 0.49$).¹⁵ In the Perspective-shifting test trials, however, chimpanzees looked

¹⁵ To make the baseline trials equivalent to the Perspective-shifting test trials, I

significantly more often in Unseen than in Seen trials ($T^+ = 55.00$, $N = 10$ [0 ties], $p = 0.002$). Looking was effective in locating the food: in Unseen trials, chimpanzees chose correctly significantly more often than the chance proportion (0.33) when they looked

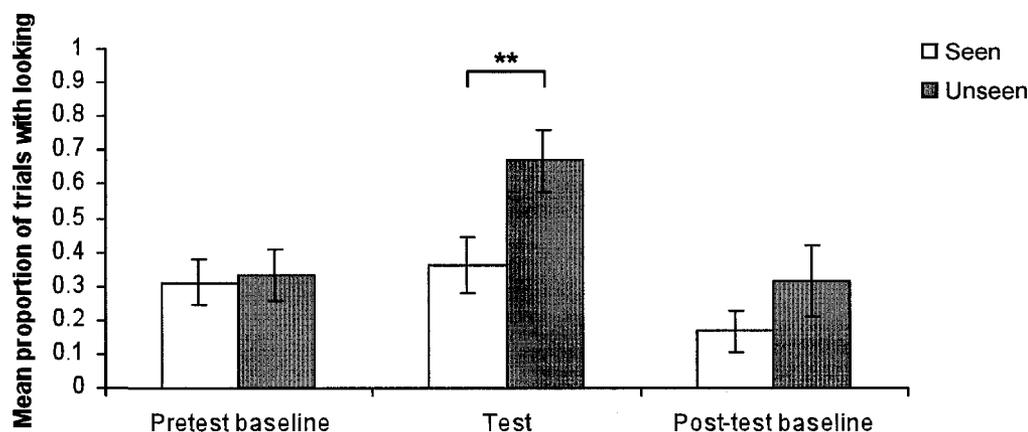


Figure 5.3. Mean proportion of trials (\pm SE) in Study 3, Experiment 1, in which chimpanzees attempted to look into the containers in Seen versus Unseen trials, across the Pretest baseline ($N = 8$), Perspective-shifting test ($N = 10$), and Post-test baseline ($N = 10$) trials. ** $p < 0.01$

into the containers ($T^+ = 55.00$, $N = 10$ [0 ties], $p = 0.002$), whereas their performance was at chance when they did not look ($T^+ = 26.00$, $N = 8$ [0 ties], $p = 0.31$). In the Seen trials, chimpanzees performed above chance whether they looked or not (for both analyses, $T^+ = 55.00$, $N = 10$ [0 ties], $p = 0.002$). Overall, chimpanzees looked far less

included only the first 20 seconds in the analysis. However, with the entire 60 seconds included there were still no significant differences in looking between Seen and Unseen trials (Pretest: $T^+ = 9.00$, $N = 5$ [3 ties], $p = 0.88$; Post-test: $T^+ = 25.00$, $N = 8$ [2 ties], $p = 0.38$). Also note that only eight chimpanzees were included in the pretest comparison of Seen and Unseen trials because two chimpanzees did not receive the Seen trials, as mentioned earlier.

often than they had in the Tubes trials, likely because the effort required to look into the Perspective-shifting test containers was greater.

First look. Chimpanzees' performance in the Perspective-shifting test trials demonstrated that they anticipated the correct visual perspective for seeing into the different containers. As seen in Figure 5.4, chimpanzees' first look was almost always the correct one for the type of container used. For Unseen trials only, exact Friedman tests for each container type were all statistically significant ($F_r \geq 10.41$, $df = 2$, $N = 10$, $p \leq 0.003$ in all cases). Pairwise tests revealed that the correct response for each container type occurred first in a significantly greater proportion of trials than either of the two incorrect responses (Wilcoxon exact tests: $T^+ \geq 43.00$, $N \geq 9$ [ties ≤ 1], $p \leq 0.016$ in all cases), whereas there were no significant differences in frequency of the two incorrect responses ($T^+ \leq 3.00$, $N \leq 2$ [ties ≥ 8], $p \geq 0.50$ in all cases). Additionally, when chimpanzees did look, they rarely looked from more than one different perspective. Of the 184 trials in which chimpanzees looked, there were only 11 in which they performed more than one type of look.

For Seen trials only, the pattern of responding was generally the same, except that the Friedman test was not significant for the cylinders (cylinders: $F_r = 4.57$, $p = 0.14$; triangles: $F_r = 16.20$, $p < 0.001$; trapezoids: $F_r = 12.00$, $p = 0.004$; $df = 2$, $N = 10$ in all cases). For the triangles and trapezoids, the correct perspective was adopted in a significantly greater proportion of trials than either of the two incorrect perspectives ($T^+ \geq 21.00$, $N \geq 6$ [ties ≤ 4], $p \leq 0.031$ in all cases), and the two incorrect responses were not

significantly different ($T^+ \leq 1.50$, $N \leq 2$ [ties ≥ 8], $p = 1.00$ in both cases). In the baseline trials, there were no significant differences in the incidence of different types of response.

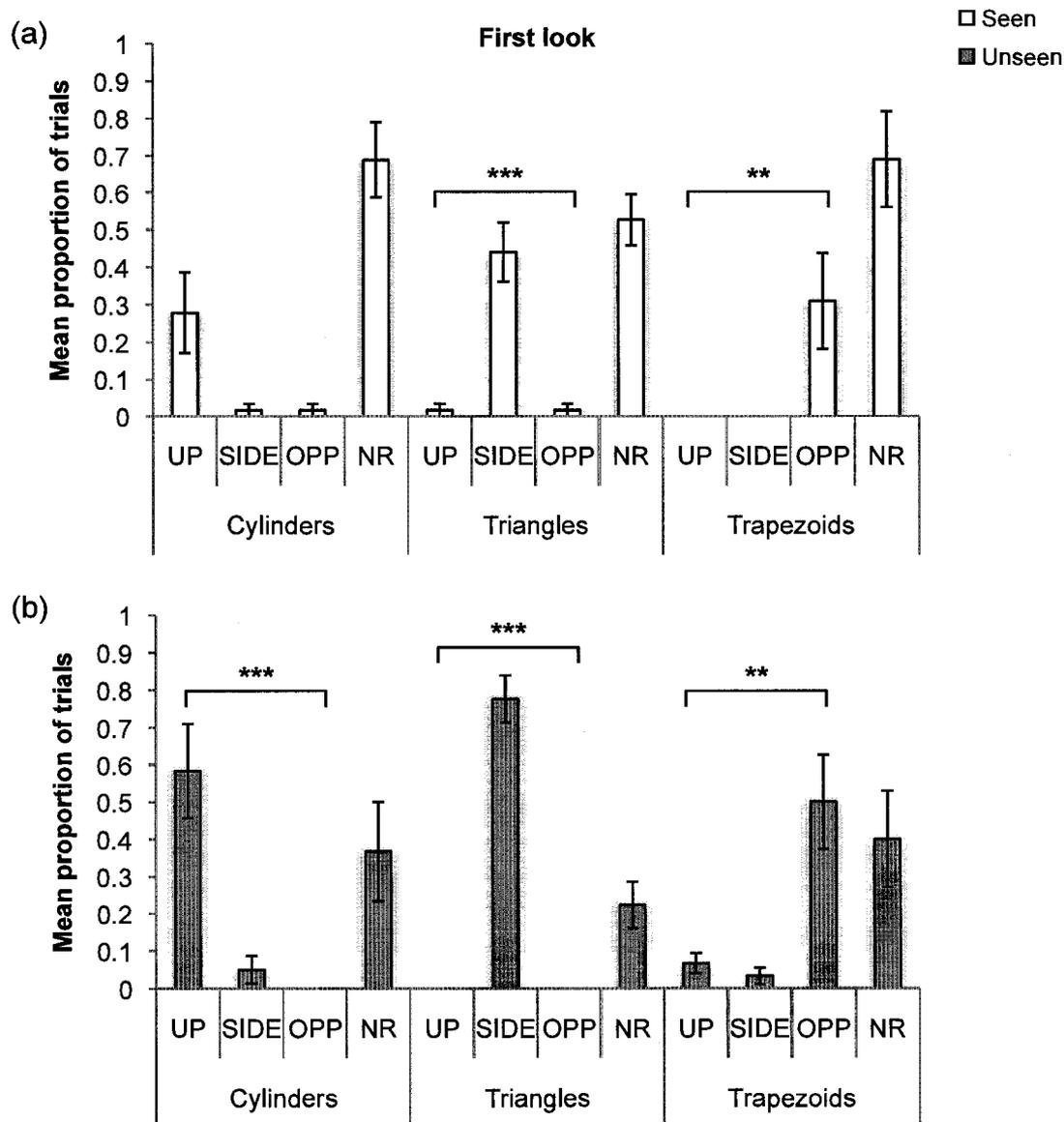


Figure 5.4. Mean proportion of trials (\pm SE) in Study 3, Experiment 1, in which chimpanzees looked first from the UP, SIDE, or OPPOSITE (OPP) perspective or made no response (NR), across all three container types: (a) Seen baiting; (b) Unseen baiting. ** $p < 0.01$, *** $p < 0.001$

Collapsing across the Pre- and Post-test baselines and across the Seen and Unseen conditions (all of which were not significantly different from one another: $p \geq 0.125$ in every case), chimpanzees looked from the UP, SIDE and OPPOSITE perspectives in the same average proportion of trials ($F_r = 4.34$, $df = 2$, $N = 10$, $p = 0.12$).

To investigate whether chimpanzees learned over time to perform the correct response for each type of container, I compared their performance in the first versus the second half of trials in both the Seen and Unseen conditions. There was no detectable sign of learning, as chimpanzees looked just as often in early trials as in later ones (Seen: $T^+ = 31.50$; $N = 10$ [0 ties], $p = 0.71$; Unseen: $T^+ = 23.00$, $N = 8$ [2 ties], $p = 0.52$). And when chimpanzees did look, it was from the correct perspective just as often in the first half as in the second half of trials (Seen: $T^+ = 6.00$; $N = 3$ [5 ties], $p = 0.25$; Unseen: $T^+ = 18.00$, $N = 6$ [4 ties], $p = 0.16$). Further, chimpanzees did not succeed by simply learning the correct responses during Seen trials and then transferring them over to the Unseen trials. Of the 30 trials in which chimpanzees experienced a given container for the very first time, 13 were Unseen trials. Examining just those trials, I found that chimpanzees visually searched for the food in 8 of them, and in every case their first look was from the correct perspective.

First move. I examined the first move chimpanzees made after the food became potentially visible (Figure 5.5). The results were consistent with the first look measure. In Unseen trials, chimpanzees' first move was most often the correct one ($F_r \geq 9.92$, $df = 2$, $N = 10$, $p \leq 0.005$ in all cases). And again, the correct response occurred more often than either of the two incorrect responses ($T^+ \geq 26.50$, $N \geq 7$ [ties ≤ 3], $p \leq 0.047$ in all cases),

while the two incorrect responses occurred with equal frequency ($T^+ \leq 12.50$, $N \leq 5$ [ties ≥ 5], $p \geq 0.25$ in all cases).

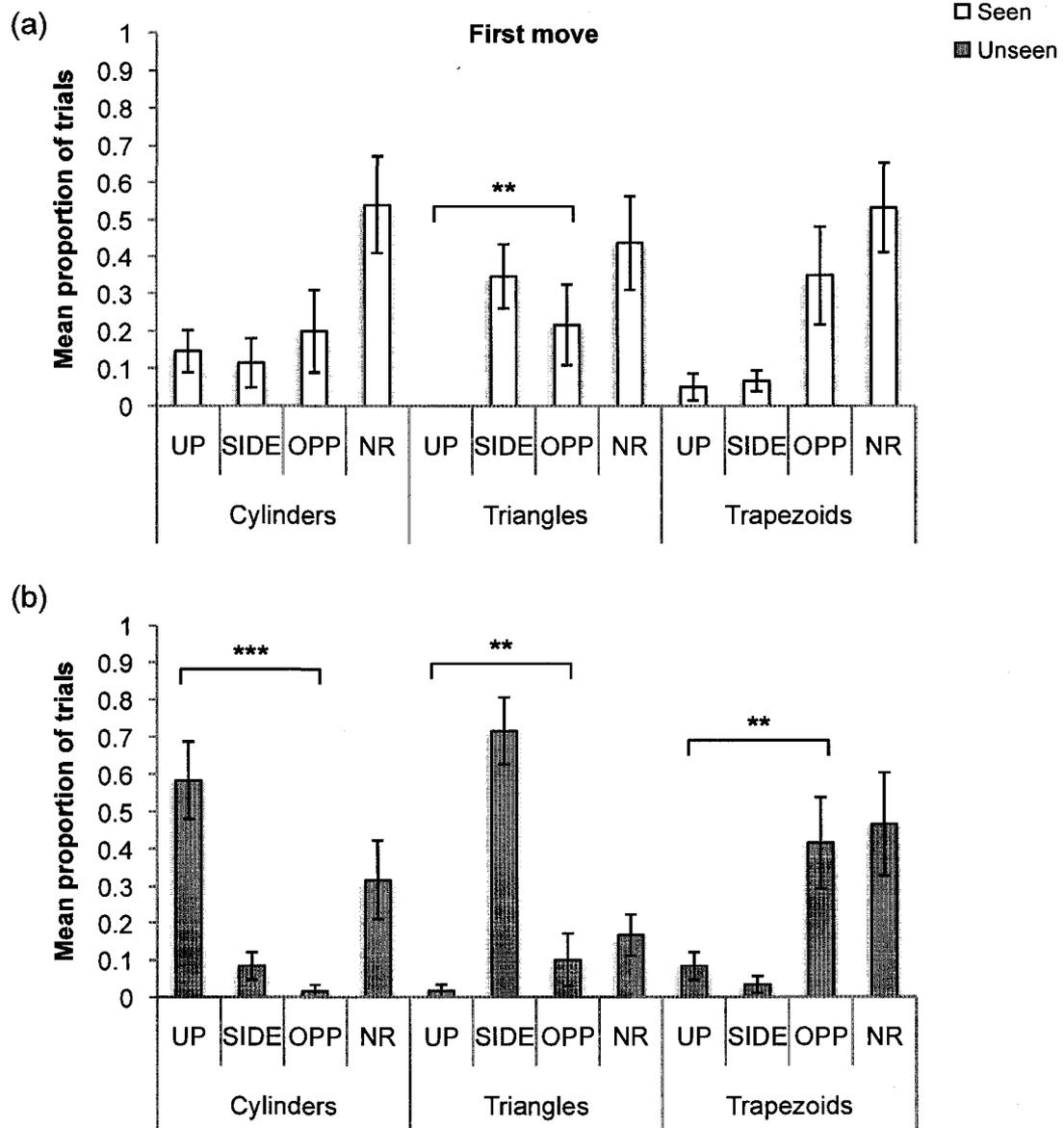


Figure 5.5. Mean proportion of trials (\pm SE) in Study 3, Experiment 1, in which chimpanzees first moved UP, SIDE, or OPPOSITE (OPP), or made no response (NR), across all three container types, irrespective of whether or not they looked: (a) Seen baiting; (b) Unseen baiting. ** $p < 0.01$, *** $p < 0.001$

Moving in Seen trials was less systematic than in Unseen trials, as only the Friedman test for the triangles reached significance (cylinders: $F_r = 1.13$, $p = 0.60$; triangles: $F_r = 11.21$, $p = 0.002$; trapezoids: $F_r = 4.79$, $p = 0.11$; $df = 2$, $N = 10$ in all cases). For the triangles, the SIDE versus UP comparison was significantly different ($T^+ = 36.00$, $N = 8$ [2 ties], $p = 0.008$) but the other two paired comparisons were not ($T^+ \leq 21.00$, $N \leq 7$ [ties ≥ 3], $p \geq 0.125$ in both cases). These results make sense if chimpanzees were moving in Seen trials *not* in an attempt to find the food but for other reasons such as boredom or frustration at the long wait. Finally, I considered the possibility that coding from the moment the food became visible in Unseen trials could have biased the moving results, because chimpanzees might have simply moved to a window at which a screen was currently being removed. To address this possibility, I reran the analyses using my second coding that began from the moment E finished the baiting, before any screens were removed. The pattern of results did not change ($F_r \geq 9.75$, $df = 2$, $N = 10$, $p \leq 0.007$ in all cases).

To summarize, in Experiment 1, chimpanzees searched more often when they had not seen the baiting, but only when there was a real possibility of seeing the food (i.e., when the containers had openings in them). Searching was systematic: chimpanzees' most common response by far was to adopt the correct viewing perspective for the particular containers used, look into the containers, and then return to the home window to choose a container. The results were robust across different measures (first look or first move) and across different starting points for coding (from the moment the food was potentially visible, or from the moment E finished baiting).

These findings strongly suggest that the chimpanzees anticipated which visual perspectives would be effective in gaining visual access to the hidden food and which would not, depending on the type of container. This was the high-level explanation for chimpanzees' success in the Perspective-shifting test. I considered, however, an alternative low-level explanation. When chimpanzees were seated at the home window, the openings in the cylinders and triangles were visible to them, and the shape of the trapezoids was suggestive of an opening at the back. Chimpanzees could have therefore solved the task by aligning themselves with visible openings when they knew food to be hidden somewhere among an array of containers. In other words, they could have simply responded to a general tendency to look into visible openings, calling into question whether they were truly anticipating which angle of view they needed to adopt. I therefore tested this alternative explanation in Experiment 2.

Experiment 2 – UP *versus* OPPOSITE

In this experiment, I examined whether chimpanzees preferred to search in a visible opening for hidden food when there was also a non-visible opening available. If they did so, this would weaken the findings from Experiment 1 by leaving open the possibility that chimpanzees succeeded in the Perspective-shifting test by responding to a general tendency to align themselves with visible openings during foraging. In contrast, if they did not favor the visible opening, it would strengthen the high-level interpretation that they were not simply exercising such a tendency, but were instead anticipating the correct angle of view required to see the food. The task for Experiment 2 was as follows: chimpanzees chose between two identical cylinders open on one end. One cylinder was

oriented vertically and the other was oriented horizontally with its open end turned away from chimpanzees. If chimpanzees had a general tendency to search in visible openings they should look more often into the vertical cylinder. Otherwise, they should look into both cylinders with equal frequency. Note that if chimpanzees looked into one cylinder and did not find the food, they could infer by exclusion that it must be in the other cylinder, and so looking into either cylinder was an effective response. Call (2004) and Call and Carpenter (2001) found that some chimpanzees were capable of making such inferences by exclusion.

Methods

Participants

The same 10 chimpanzees as in Experiment 1 participated. Five months elapsed between experiments.

Design

All chimpanzees received 48 trials. The reward in the first 32 trials was a grape or a small paper cup of yoghurt. Because of waning motivation, in the last 16 trials I used a more desirable reward: a cup of yoghurt perched atop a banana slice and surrounded by several grapes (the *Cornucopia* trials). Within each block of 16 trials, there were equal numbers of Seen and Unseen trials, the visible opening was on the right and left an equal number of times, and the food was on the right and left an equal number of times. The various combinations of these variables were presented in random order with the

constraint that none of the factors could repeat for more than three consecutive trials.

Chimpanzees received a variable number of trials per session (mode = 8, range = 2–20).

Experimental Set-up

The set-up was as for Experiment 1 except that the opposite window was occluded during baiting and the center window remained occluded throughout the entire trial.

Materials

Containers were two identical 7.5-cm diameter cylinders similar to those used in Experiment 1 except that they were 16.5 cm in height and had a removable cap on one end. I used slightly taller cylinders than for Experiment 1 because I wanted the effort of moving up to look into them from above to be comparable to the effort of moving over to the opposite window to look into them from the other end. A plastic rectangle attached to the side of each cylinder kept it from rolling when horizontal.

Procedure

The procedure was as in Experiment 1 with two changes. First, in Seen trials, to minimize the possibility that chimpanzees would simply look into openings where they saw food disappear a moment before, the cylinders were baited through the end opposite the opening. E removed the cap from the bottom of the vertical cylinder or the front of the horizontal cylinder, placed the food inside, and then replaced the cap. Second, I shortened the trials to 10 seconds because 20 seconds had proven to be unnecessarily long in Experiment 1 (chimpanzees typically responded within the first few seconds).

Coding and Analyses

Dependent measures were as for Experiment 1, except that first moves were no longer coded. This measure was now unnecessary because chimpanzees had to climb above the level of the metal frame (because the cylinders were taller) or move all the way over to the opposite window (because the center window was occluded) to see into the containers. These UP and OPPOSITE responses, respectively, were almost invariably accompanied by extremely clear attempts to look into the containers, so that first looks and first moves amounted to essentially the same measure. All statistical tests were non-parametric and reported p values are two-tailed.

Reliability

A naïve coder blind to hypotheses and experimental condition coded 20% of randomly chosen Seen and Unseen trials for choice, looking versus not looking, number of different looks, and first look. Kappas were 1.00, 0.90, 0.89, and 0.90, respectively ($p \leq 0.001$ in all cases).

Results and Discussion

As in Experiment 1, chimpanzees looked significantly more often in Unseen than in Seen trials ($T^+ = 45.00$, $N = 9$ [1 tie], $p = 0.004$), and they chose correctly significantly more often than the chance probability (0.50) in Unseen trials only when they looked (looked: $T^+ = 45.00$, $N = 9$ [0 ties], $p = 0.004$; did not look: $T^+ = 35.50$, $N = 10$ [0 ties], $p = 0.45$). Overall, looking frequency was lower than in Experiment 1, possibly because looking required more effort. Additionally, there were only two containers in this

experiment, so chimpanzees had a 50% chance of guessing correctly without looking. The greater effort of looking, combined with the lower cost of not looking, may have caused reduced motivation to search. To see if the reward's desirability made a difference, I compared looking in the first 32 trials (grape/yoghurt reward) with the last 16 trials (Cornucopia reward). Chimpanzees looked in a significantly greater proportion of Cornucopia trials than grape/yoghurt trials in the Unseen condition ($T^+ = 34.50$, $N = 8$ [2 ties], $p = 0.023$) and results approached significance in the Seen condition ($T^+ = 15.00$, $N = 5$ [5 ties], $p = 0.063$). Increasing the desirability of the reward thus boosted chimpanzees' motivation to search, especially when they did not know the reward's location.

Most important was whether chimpanzees showed evidence of preferring the visible to the non-visible opening. In the Seen and Unseen conditions, chimpanzees moved OPPOSITE just as often as UP (Seen: $T^+ = 6.00$, $N = 4$ [6 ties], $p = 1.00$; Unseen: $T^+ = 34.00$, $N = 9$ [1 tie], $p = 0.20$). Thus, there was no clear evidence that chimpanzees preferred to look into a visible opening when looking into a non-visible one was just as effective (Figure 5.6). Chimpanzees' success in Experiment 1 therefore cannot be explained by a general, over-riding tendency to search in visible openings for hidden food. However, although the effect was not significant, there was a slight trend in Unseen trials for chimpanzees to look more often into the visible opening than into the non-visible one. A high-level interpretation of chimpanzees' success in the Perspective-shifting test (that they anticipated the correct angle of view) needs to be strengthened further. This was my goal in Experiment 3. If chimpanzees could be shown to

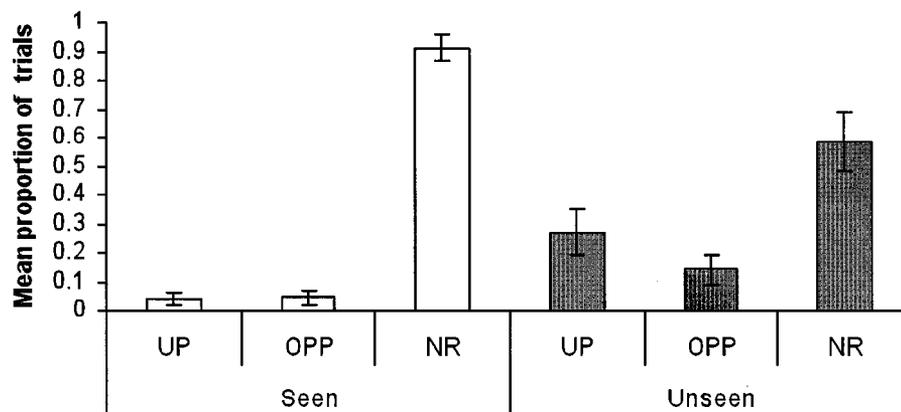


Figure 5.6. Mean proportion of Seen and Unseen trials (\pm SE) in Study 3, Experiment 2, in which chimpanzees first looked from the UP or OPPOSITE (OPP) perspective or made no response (NR).

demonstrate a clear preference for *non*-visible over visible openings in a food-finding task, this would indicate that they were not just responding inflexibly to a general tendency to search in visible openings when foraging.

Experiment 3 – UP *versus* DOWN

In Experiment 3, I tested whether chimpanzees would prefer non-visible openings to visible ones when looking into the non-visible ones was the less effortful of the two responses. If so, this would provide stronger evidence *against* the hypothesis that chimpanzees simply look into the first visible opening they see when searching for food. I presented chimpanzees with a situation in which both visible and non-visible openings were present at the same time, as in Experiment 2. In this case, however, the non-visible openings required far less effort to look into than the visible ones. To create visible and non-visible openings, I stood cylinders open on both ends on top of a raised transparent

platform. Chimpanzees could climb upward to look into visible openings at the top of the cylinders, or they could crouch downward to look into the non-visible openings at the bottom. Because apparent lessened motivation to search in Experiment 2 may have been partly caused by having only two containers, for this experiment I increased the number of containers used in each trial to four.

Methods

Participants

Nine of the same 10 chimpanzees participated (one adult female failed the pretest training criterion). Three months elapsed between Experiments 2 and 3.

Design

Chimpanzees received a variable number of training sessions as required and then 24 test trials across two sessions (all Unseen). At the beginning of each test session chimpanzees were given two refresher trials exactly like the training trials.

Experimental Set-up

The set-up was similar to the earlier one, with the following changes: (1) the door between the enclosures was closed; (2) because there were four containers, I replaced the three-holed Plexiglas panel with a steel mesh panel through which chimpanzees could poke their fingers to choose a container; (3) a third testing room with a different window height (192 cm) was also used; and (4) instead of placing containers directly onto the table, I placed them onto a transparent, 1/2-cm-thick Plexiglas platform (75 x 30 cm) that

stood on top of the table and was 8–10 cm high, depending on the height of the chimpanzee (see Figure 5.7).

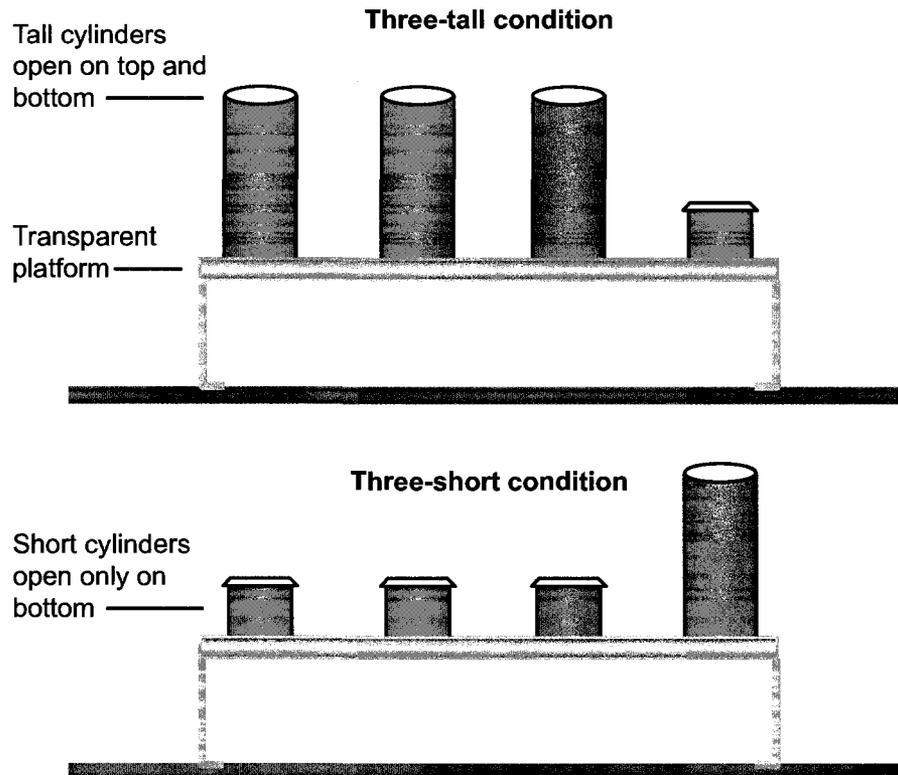


Figure 5.7. Experimental set-up for the two conditions in Study 3, Experiment 3: the Three-tall condition (top) and the Three-short condition (bottom).

Materials

Training cups. For training I used four white opaque plastic drinking cups, 13 cm tall with an 8 cm diameter opening.

Test containers. I again used 7.5-cm diameter cylinders similar to those in Experiments 1 and 2, but with different heights. There were three identical tall cylinders

(16.5–23.5 cm high, depending on the height of the window) and three identical short cylinders (11 cm high). The tall cylinders were open on both ends but the short cylinders were only open on the bottom.

Procedure

In Experiments 1 and 2, the chimpanzees had gained experience climbing upward to look into containers from above (the UP response), but they had never encountered a raised transparent platform that provided visual access from underneath. I therefore needed to be sure that they could also crouch downward to look into containers from below (the DOWN response). Once I was confident they had both responses in their repertoire I could present them with the test, in which there were both visible and non-visible openings present at the same time.

Pretraining. Just before training, E showed chimpanzees that the platform was transparent by holding it vertically in front of her face and also holding up a grape and an empty cup behind it (and she repeated this after the first two training trials). E then gave chimpanzees two warm-up trials in which she placed four cups onto the platform, all upright in one trial and all upside-down in the other, with order determined randomly. She put a grape into one of the cups as the chimpanzee watched and then slid the table forward so the chimpanzee could choose.

Training. E showed chimpanzees that all four cups were empty and positioned a screen to block their view. She positioned the cups all upright or all upside-down in a row, hid a grape in one of them, and removed the screen. E waited until the chimpanzee made the appropriate response (UP for upright cups and DOWN for upside-down cups)

or for 2 minutes, whichever came first. She then slid the table forward so the chimpanzee could choose.

Training was administered in blocks of four trials. Within each block, the cups were upright in two trials and upside-down in two trials and the food was at each location once, with order determined randomly. The chimpanzee had to perform the appropriate response (not necessarily first) in all four trials to proceed to testing; otherwise another block was run. After each block, if chimpanzees had not reached criterion then E provided training tailored to individual needs. For chimpanzees who were not moving downward, she tilted the platform upward until the grape was visible from below. For chimpanzees who were not moving upward, she tilted the cups forward until the grape was visible from above. If chimpanzees continued having trouble E demonstrated the responses. One or two blocks a day were administered until chimpanzees reached criterion. Three chimpanzees did so within two blocks, three within four blocks, and three within five blocks. One chimpanzee did not reach criterion within six blocks (she would not move upward) and was therefore dropped from the experiment. In summary, the goal for training was to ensure that chimpanzees could make both types of response, UP and DOWN, so that they could choose between these responses in the test trials.

Test trials. For the test, there were again four cylinders on the transparent platform in each trial. I ran two conditions: (1) the *Three-tall* condition included three tall cylinders plus one short one, and (2) the *Three-short* condition included three short cylinders plus one tall one. The fourth, odd cylinder was always positioned on the far left or right side of the others (see Figure 5.7). It was never baited and chimpanzees knew

this, for reasons that will become clear below. The height of the three identical cylinders (tall or short), the position of the fourth cylinder (left or right), and the location of the food (left, right or center) were all counterbalanced and randomly ordered, with the stipulation that none of these variables could repeat for more than three consecutive trials.

I included the fourth, odd cylinder in the array so that I could run the strongest test I could think of for whether chimpanzees prefer to look into visible openings. Namely, if there were three non-visible openings and one visible opening present at the same time and the visible opening clearly had not been baited (i.e., the Three-short condition), would chimpanzees nevertheless prefer the visible opening to the non-visible ones? If they did so, this would indicate an undeniable preference for visible openings. Given the results of Experiment 2, I expected that in the Three-short condition, chimpanzees would have little trouble ignoring the one visible opening in the tall cylinder, especially given that that cylinder was clearly not baited. I predicted that they would instead choose the only effective response in that condition, which was to move DOWN and look into the cylinders from underneath. The Three-tall condition, in which either the UP or DOWN response would be effective in locating the food, was therefore of more interest. In the Three-tall condition, chimpanzees could choose to either climb UP and look into the cylinders from above or crouch DOWN and look into them from below. While both responses were effective, the UP response required significantly more effort than the DOWN response.

The procedure for all trials was as follows. E positioned the three identical cylinders in a vertical line in the center of the raised platform, one in front of the other,

and let chimpanzees watch as she baited one of them. She also placed the fourth cylinder onto the platform, apart from the others so it was clear she did not bait it. Next, E blocked the chimpanzee's view of the three identical cylinders and positioned them in a horizontal row across the platform, with the fourth cylinder at the far left or far right of the row and still visible to the chimpanzee. (When the fourth cylinder was a tall one, E also placed a lid on it so that chimpanzees could not look down into it during this time.) Note that although the chimpanzees saw the baiting, they did not see the positioning of the containers and so did not know where the food was located. All trials were thus Unseen. When E had finished positioning the containers she removed the screen blocking the chimpanzee's view (and the lid from the fourth, tall cylinder, when relevant), sat for 10 seconds looking at her stopwatch, and then pushed the containers within chimpanzees' reach. E did not leave the window during the delay because she was not obstructing chimpanzees' view by remaining there. Further, the metal grating made it difficult to see the chimpanzees' faces on video and so it was easier to code the responses live.

Coding and Analyses

E coded live whether chimpanzees chose the correct container and also whether they performed either of two responses: moving UP beyond the level of the metal frame and looking into the containers from above, or crouching DOWN and looking into them from below. All statistical tests were non-parametric and reported p values are two-tailed.

Reliability

A naïve coder judged 20% of randomly selected test trials for container chosen,

first look, and whether participants looked from more than one perspective. Kappas were 1.00, 0.94, and 0.94, respectively ($p \leq 0.001$ in all cases).

Results and Discussion

As Figure 5.8 shows, looking was at ceiling levels, with every chimpanzee looking into the containers in every trial. It was also very effective, as there was only a single trial in which a chimpanzee chose incorrectly. Most interestingly, chimpanzees almost always crouched downward to look into the containers from below rather than climbing upward to look into them from above, and the difference was highly significant ($T^+ = 45.00$, $N = 9$ [0 ties], $p = 0.004$). There were only 11 test trials in 216 in which a chimpanzee performed more than one type of response, either moving UP to look after previously moving DOWN (2 trials) or moving DOWN to look after previously moving UP (9 trials).

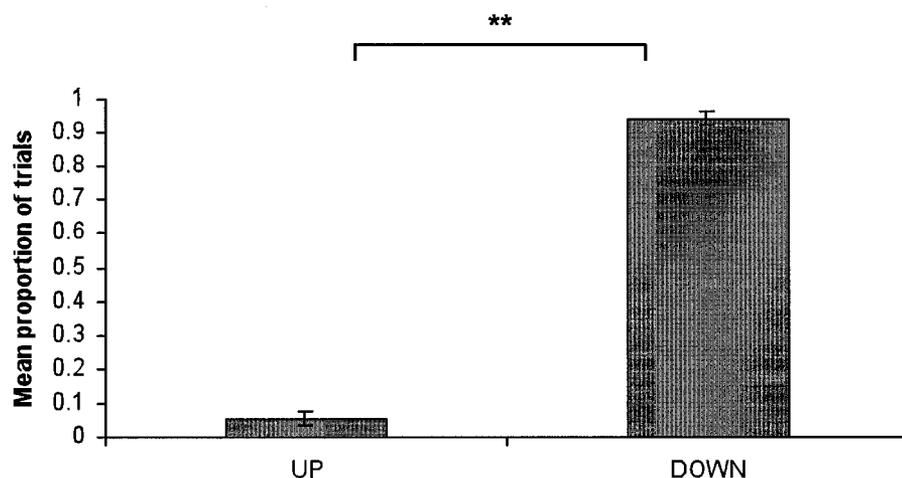


Figure 5.8. Mean proportion of trials (\pm SE) in Study 3, Experiment 3, in which subjects first looked from the UP or DOWN perspective. ** $p < 0.01$

It is very likely that chimpanzees preferred the DOWN response because it was the far less effortful of the two responses. Supporting this view is the fact that when chimpanzees did climb upward as their first response, it was almost always (except in two cases) within the first six test trials. Chimpanzees thus quickly learned that the DOWN response was less effortful than the UP response and they came to rely on this response almost exclusively as time went on. Furthermore, there were only 12 trials in which a chimpanzee's first or only response was to climb upward. Of these 12 trials, only two were in the Three-short condition, in which looking into the one visible opening (at the top of the fourth, tall cylinder) could not have possibly been effective in locating the food because that container was never baited. Thus, as predicted, I found that chimpanzees did not show a preference for visible openings. Instead, they clearly preferred to look into non-visible openings when this was the easier of the two options. Together with the results of Experiment 2, this finding rules out the possibility that chimpanzees succeeded in Experiment 1 by acting on a general tendency to look into visible openings.

General Discussion

Chimpanzees visually searched significantly less often when they had witnessed the hiding of a food reward than when they had not, replicating findings from previous studies (Call, 2005; Call & Carpenter, 2001). In doing so, they demonstrated that they could discriminate between situations in which they knew the location of a food reward and those in which they did not. More notably, chimpanzees immediately recognized which one of various viewing perspectives they needed to adopt to spy a reward hidden

inside a container. In Experiment 1, chimpanzees were presented with a situation in which they could move around an enclosure to visually search for food hidden among three identical open containers. The correct perspective for seeing into the containers depended on the location of the openings, which varied across trials. In an overwhelming proportion of cases, chimpanzees' first response was to go immediately to the appropriate location for gaining visual access to the hidden food. Experiments 2 and 3 ruled out the possibility that chimpanzees succeeded in Experiment 1 because of a general tendency to align themselves with visible openings during foraging. In Experiment 2 they looked into visible and non-visible openings equally as often when both responses required similar levels of effort, and in Experiment 3 they even preferred non-visible openings to visible ones when looking into the non-visible openings was the easier option. In short, chimpanzees demonstrated that they could anticipate the effect that specific shifts in their own perspective would have on their ability to see the food. Whether their responses involved explicit reasoning of any kind is impossible to know with certainty. However, the fact that they often readily left the home window (the only place where food could be physically accessed) with the clear intention of visually locating the food before making a choice suggests a deliberateness to their responses that goes far beyond any reflex foraging response. Further, it is important to note that the chimpanzees did not move in an attempt to acquire the food. They were quite aware of the fact that the food could only be acquired at the home window, as evidenced by their returning to the home window to choose a container after they had spotted the food.

One common criticism of visual perspective taking research with animals is that they could solve the tasks presented to them by simply responding to the observable behavior of others. For example, instead of understanding that a social partner with its back turned cannot see them, chimpanzees might learn from experience that when others are oriented away from them they are typically unresponsive to their gestures (for an extended and ongoing debate on this issue see Heyes, 1998; Penn & Povinelli, 2007; Povinelli & Vonk, 2003, 2004; Tomasello et al., 2003a; Tomasello, Call, & Hare, 2003b). My task involved no social partners, so there were no social cues available for participants to make use of. Furthermore, chimpanzees did not learn—either through previous experience with similar situations or within the context of this study—where they needed to position themselves to see into the different types of containers. From the beginning, they almost invariably adopted the correct perspective for seeing into the particular containers they were confronted with in any given trial, despite the fact that they had never encountered these containers before the start of this study. While it is true the cylinders were somewhat similar to drinking cups, with which some of the chimpanzees may have had previous experience, the triangles and trapezoids were entirely novel to them, as was the transparent platform in Experiment 3. Their immediate mastering of the task thus suggests that they were putting their knowledge of the containers' openings, gained during the familiarization phase, together with their current perceptions of how the containers were positioned in front of them to judge where they needed to move. Although numerous species search for occluded objects and find appropriate viewing (and approach) angles during their search (Chapuis, 1987; Regolin,

Vallortigara, & Zanforlin, 1995; Zucca, Antonelli, & Vallortigara, 2005), chimpanzees were capable of *anticipating* the appropriate viewing perspective. Thus, chimpanzees clearly demonstrated excellent visual perspective taking abilities with regard to themselves, just as they have with regard to others in previous studies.

In human children, as we saw in Chapter 3, an understanding of visual perception appears earlier in development than the ability to recognize false mental states and is likely foundational to it. For example, children have to first understand that seeing leads to knowing before they can understand how a false belief is formed by witnessing one element in a sequence of events but not another. All evidence to date suggests that chimpanzees, although they have demonstrated a good understanding of their own and others' visual perceptual experiences, never acquire the ability to consider others' false beliefs. While they are capable of recognizing that there can simultaneously exist two different perceptions of a scene (one's own and another's, or one's own from two different perspectives), they do not appear to recognize that another individual's beliefs can differ from their own or from reality. The question that follows is, do chimpanzees fare any better with false mental states when they themselves are experiencing those states? This question is addressed in the following chapter as I continue my investigation of how much chimpanzees understand about their own (as opposed to others') mental states.

CHAPTER 6 – APPEARANCE *VERSUS* REALITY

STUDY 4 – LENS STUDY

So far, we have seen chimpanzees perform very poorly in false belief tests, despite all efforts to make these tests as transparent as possible for them (Studies 1 and 2). In contrast, they did very well in a test requiring them to make judgments about their own visual perceptions (Study 3), just as they have done well in past studies requiring them to judge what others can or cannot see and also what others do or do not know based on past visual experiences. Inferring states of seeing and knowing may define the limit of chimpanzees' mental-state attribution capacities. Or perhaps false mental states are only problematic for chimpanzees when those states belong to others. It is possible that chimpanzees' abilities regarding the more complex aspects of the relationship between seeing, knowing and believing may be greater with regard to their own mental states. Hence, in Study 4, I developed a nonverbal appearance-reality task for chimpanzees to see if they were capable of recognizing when they themselves were experiencing an erroneous mental state: a mistaken visual perception.

As noted in Chapter 1, researchers have often observed similarities in the performance of children on false belief and appearance-reality tasks. However, while chimpanzees have so far performed poorly in false belief tests, it does not necessarily follow that that they should perform poorly in appearance-reality tests, for several reasons. First, there are potentially important differences between the two types of test that may affect chimpanzees' performance, including the mental states at issue. Gopnik

and Astington (1988) speculate that it might be easier to “construct metarepresentations of representations that are more perceptually based than those that are more abstract” (p. 34). If this is so, then representing a mistaken perception may be easier for chimpanzees than representing a false belief (this possibility is discussed further in Chapter 7). Perhaps more crucially, false belief tests focus on others’ mental contents whereas appearance-reality tests focus on one’s own. Chimpanzees could possess a narrower capacity for dual representation than humans—one that is limited to their own sphere of mental experience. This is not an unreasonable prospect, given that one’s own mental states can be directly experienced while others’ must be indirectly inferred. Intuitively, it would seem to be easier to recognize when one has a perception in conflict with one’s own knowledge of reality than when someone else has a belief in conflict with reality. Research with apes (Call, 2005; Call & Carpenter, 2001) and monkeys (e.g., Hampton, 2001; Kornell, Son, & Terrace, 2007; Washburn, Smith, & Shields, 2006) does suggest that they may possess some ability to reflect upon their own mental states, although Carruthers (2008), among others, has challenged the metacognitive interpretation of these findings. Carruthers does grant, however, that more convincing evidence of metacognitive processes might be found by pitting “behavior guided by a current belief against behavior that is guided by a belief about current perception” (p. 86), which is the approach I take in the current study.

False belief tests differ from appearance-reality tests in another way that may give children an unfair advantage over apes. False belief tests typically entail a fairly complex sequence of events played out over time by two or more actors, giving the tests a

narrative structure that is not a feature of most appearance-reality tests. With their bedtime stories, cartoons, and so on, even toddlers have amassed far more experience processing such narrative structures than apes. Although chimpanzees do poorly in false belief tests, they might therefore perform better in more simply structured appearance-reality tests.

A final reason to suspect that chimpanzees might succeed in appearance-reality tests is that they have succeeded in other paradigms that draw on related cognitive skills. For example, in the children's literature, visual perspective taking skills have been found to correlate positively with performance on appearance-reality tasks (Flavell, 1986; Flavell et al., 1986). And some investigators (e.g., Braine & Shanks, 1965a; Braine & Shanks, 1965b; Murray, 1968) have theorized that grasping the appearance-reality distinction is foundational to succeeding in liquid conservation tasks, with which chimpanzees have had some success (Call & Rochat, 1996, 1997; Muncer, 1983; Suda & Call, 2004; Woodruff et al., 1978). With regard to the latter, however, as we saw in Chapter 3, there was some evidence that chimpanzees were influenced by the appearance of the containers in those studies. The extent to which they used strategies that did not rely upon appearance-reality understanding, such as tracking the larger amount of liquid through the transfer, was also unclear. Investigations aimed more directly at testing apes' appearance-reality understanding are therefore needed.

In this study, I asked whether chimpanzees could gain a preferred reward by ignoring a visual illusion and acting instead on what they knew to truly be the case. Specifically, I investigated whether chimpanzees, when faced with an illusion that made a

food item look larger than it really was, could ignore its apparent size and respond on the basis of what they knew to be the actual size of the item. My appearance-reality *Lens* test capitalized on chimpanzees' unmitigated passion for grapes—the bigger, the better. I used magnifying and minimizing lenses to make a small grape appear bigger and a big grape appear smaller. Chimpanzees who recognized that the apparently smaller grape was truly the bigger one should choose that grape, thereby demonstrating that they distinguished appearance from reality.

My test was comprised of several components or subtests. I began by demonstrating for chimpanzees the properties of the magnifying and minimizing lenses (the *Demo* trials). I then administered a *Basic* test inspired by the standard procedures commonly used to test for appearance-reality understanding in children. In this test, chimpanzees watched as an experimenter placed the small and big grapes behind the magnifying and minimizing lenses, respectively. Chimpanzees who could distinguish appearance from reality would be expected to choose the grape that looked smaller but was actually bigger. Subsequent tests ruled out simpler explanations for the positive results I found in the *Basic* test. Most notably, I included a *Tracking* test to address the possibility that individuals who passed the *Basic* test might have done so by visually tracking the bigger grape as it was placed behind the minimizing lens. My *Tracking* test involved a procedural modification that made it impossible for chimpanzees to solve the task by this method. Very briefly, in this test the experimenter again allowed chimpanzees to watch as she placed the small and big grapes behind the lenses, as in the *Basic* test, but then she blocked chimpanzees' view so that they could not see where she

positioned the containers (Unseen trials). For comparison, in half the trials (Seen trials) chimpanzees were also allowed to watch the placement of the containers. I next included a *Transfer* test to test whether chimpanzees who passed the Basic and Tracking tests could transfer their skills over to a different type of food: apple cubes. If so, this would suggest that they understood that the distorting effects of the lenses were not specific to grapes. Following the Transfer test, the *Mixed* test was used to examine whether chimpanzees had learned to avoid the magnifying lens, and also to test how flexibly they could apply the knowledge they displayed in earlier sessions. Finally, I tested whether chimpanzees could have solved my task by learning a simple reverse contingency rule: *choose the smaller grape to obtain the larger one*, without truly understanding that the appearance of the grapes was misleading. To investigate this possibility, I administered two slightly different reverse contingency control tests (*RC-1* and *RC-2*) in which chimpanzees were required to choose the smaller of two grapes to obtain the larger one, and vice versa. If chimpanzees could choose the apparently smaller (truly bigger) grape in my Lens test but could not choose a smaller grape to obtain a bigger one when there were no lenses involved, this would indicate that they were *not* using a simple reverse contingency rule to solve the Lens test. Two reverse contingency tests were administered because after completing the RC-1 test I realized that it differed from the Lens test in one potentially important respect (explained in the Procedure section). This lessened my confidence in it as a control test, so I addressed this problem in the RC-2 test.

For comparison and validation purposes, I also tested 4- and 4.5-year-old children with a child-adapted version of the Lens test and with several other commonly used

appearance-reality tests. I focused on this age range because the transition from failing to passing appearance-reality tests usually occurs during this time (Flavell et al., 1983; Flavell et al., 1986; Gopnik & Astington, 1988). The children's test is described first below, followed by the test for chimpanzees.

Experiment 1 – Children

The children's Lens test included essentially the same components as the chimpanzees'. However, so that all trials could be administered in one session, children were not given the Basic test, Mixed test, or reverse contingency tests (RC-1 and RC-2), and they received a much smaller number of trials overall than the chimpanzees. The Mixed test and the reverse contingency tests were not necessary for children because the small number of trials they received made it unlikely that they would learn to avoid the magnifying lens or learn a reverse contingency rule. With regard to the Basic test, this was deemed unnecessary for children because the Seen tracking trials also allowed visual tracking of the food, so they served the same function as the Basic trials. For the chimpanzees, however, I included the Basic trials because I wanted to eliminate early on those individuals who could not pass even a very simple version of my test. Further, I was concerned that moving directly into the Tracking test might be too confusing for the chimpanzees, given that the experimenter could not even minimally explain the task for them as she could for the children.

Children could not be given food because of the potential for choking or allergic reactions. I therefore had them choose between big and small pieces of artificial food to give to a small plush duck, which I creatively named "Duck".

Methods

Participants

Thirty-two children from kindergartens in Leipzig, Germany were recruited for participation by letters sent to parents. There were 16 children in each of two age groups: 4-year-olds that ranged from 48 to 52 months old (mean = 50 months, SD = 1.5) and 4.5-year-olds that ranged from 54 to 59 months old (mean = 56 months, SD = 1.8). The 4-year-olds included eight males and eight females; the 4.5-year-olds included nine females and seven males. Eight further children were dropped from the study, five because of uncooperativeness and three because of procedural error.

Design

All children received an introduction to the test followed by a demonstration (Demo) of the properties of the magnifying and minimizing lenses. The various components of the Lens test were then administered in a fixed order: (1) a four-trial Tracking test consisting of alternating Seen and Unseen trials, always beginning with a Seen trial; (2) a Transfer test consisting of two Unseen trials; and (3) a Post-test consisting of two Unseen trials similar to the initial Demo trials (see Procedure section for details). Within each test component, the truly bigger reward was on the left or right an equal number of times and the each container was baited first an equal number of times, with side and order determined randomly.

Upon completion of the Lens test, children were administered four additional appearance-reality tests (one trial of each) based on procedures developed by other

researchers. Immediately following the Lens test, they were first given a *Standard-format lens* test, in which they were directly asked which piece of food looked bigger and which was really bigger, as in standard verbal appearance-reality tests. They were then given a Rock-sponge test and two tests involving color filters: the *Nonverbal color filter* test and the *Verbal color filter* test (adapted from Flavell et al., 1983; Flavell, Flavell, & Green, 1987; Flavell et al., 1986). The two color filter tests were always grouped together with the nonverbal version administered first. Half the children received the Rock-sponge test before the two color filter tests and half received it afterwards.

Materials

Materials included two identical plastic containers (16 x 16 x 8 cm), one with a magnifying lens on front and one with a minimizing lens on front (Figure 6.1). The lenses were mounted into a hole (6.5-cm diameter) in a removable panel on the front of each container. Because the panels were angled 40 degrees away from participants and the food was placed just behind the lenses, the food could not be seen from above.¹⁶ Each container had a narrow shelf on top on which the food could be displayed.

The food for Duck was made out of modeling clay and consisted of two types: (1) small (1.75 cm diameter) and big (2.5 cm diameter) green balls that were flattened on the bottom to keep them from rolling; and (2) small (1.25 cm³) and big (1.75 cm³) cubes of white clay. These green balls and white cubes were perceptually similar to the grapes and apple cubes used for the chimpanzees. For the Rock-sponge test I used a sponge painted

¹⁶ This was verified independently by two different people for the children's experimental set-up and by three different people for the chimpanzees' set-up.

to look like a rock. For both color filter tests I used a transparent yellow filter (20 x 12 cm) mounted into a cardboard A-frame so that it could be made to stand upright at an angle of approximately 45-degrees; a hole puncher that made flower-shaped holes; and yellow and white pieces of paper.

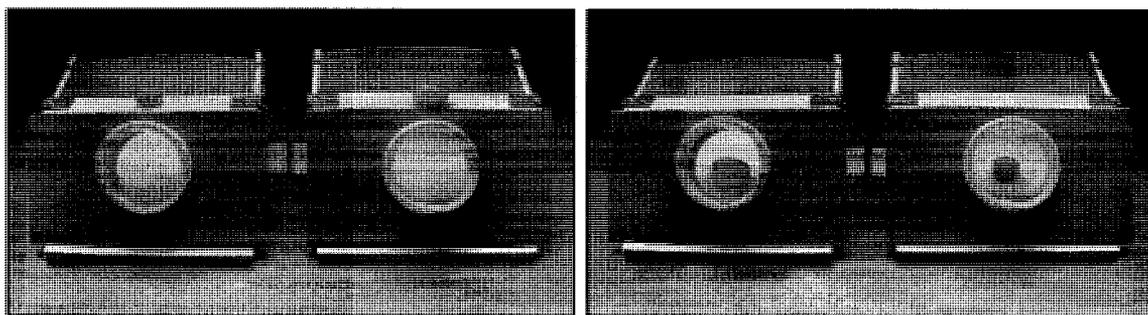


Figure 6.1. Containers used in the Lens test in Study 4, showing the (a) real and (b) apparent sizes of the grapes (magnifying lens on left, minimizing lens on right in both photographs).

Procedure

Children were tested individually in a quiet room in their kindergarten by a female experimenter (a native German-speaking research assistant) who sat opposite the child at a small table.

Introduction. The experimenter (E) first introduced children to Duck, a small toy duck that made quacking sounds when squeezed. E explained that Duck had a fat belly because he liked to eat a lot. E told children that when Duck got a lot to eat he became happy and quacked, but that when he did not get enough to eat he became sad and remained quiet. E then demonstrated to children that Duck preferred bigger pieces of

food to smaller pieces. E put three small and three big green balls onto the table, explaining that they were Duck's food. E then put one big piece of food in front of the child, saying, "Let's see what happens when we give Duck this." Duck approached the food, 'ate' it, and then quacked and danced about happily. E explained that Duck was happy because he had gotten lots to eat. E then put a small piece of food in front of the child. Duck approached and ate it, then hung his head sadly as E said, "Duck is sad because he didn't get enough to eat." Finally, E put both a big and a small piece of food in front of the child and said "Okay, now Duck can choose between two pieces himself, and we'll see which piece he takes." Duck looked back and forth between the two pieces, then approached the bigger piece, ate it, and began quacking and dancing about. E explained again that Duck was so happy because he had gotten lots to eat.

Demo. The purpose of the demo was twofold: (1) to familiarize participants with the properties of the lenses before testing, and (2) to verify that participants were in fact subject to the visual illusions created by the lenses. If so, they should choose the magnified food in the very first trial, before they had any reason to suspect that the appearance of the food was misleading.

E baited the containers out of view of children, putting a small piece of food behind the magnifying lens and a big one behind the minimizing lens. When viewed through the lenses, the apparent relative size of the food items would thus be the reverse of reality. E then presented the containers and asked children to indicate which piece of food they should give Duck so that he would be happy and quack. When the child pointed to one of the containers, E said, "Okay, but before we give Duck the food, let me

show you something.” E then slowly removed the food from the container and replaced it again until the child had seen its visual transformation at least twice. She then repeated this with the other container. Finally, E removed both pieces of food from the containers and gave Duck the one the child had indicated. Duck reacted appropriately, quacking and dancing when he got the truly bigger piece and hanging his head quietly when he got the smaller piece. E did two Demo trials, one with the magnifying container on the left and one with it on the right, with order counterbalanced across participants.

Lens test. The Lens test, which was administered immediately after the Demo trials, investigated participants’ ability to distinguish appearance from reality. It included various components administered in fixed order: a Tracking test, a Transfer test, and a Post-test. Each of these is discussed in turn below.

Tracking test. The Tracking test included two conditions, Seen and Unseen, and always began with a Seen trial.¹⁷ In the Seen trials, E put one container onto the middle of the table and placed the food behind the lens as the child watched. She then stacked the second container directly on top of the first and lowered the second food item into that container. Again, the small food item always went behind the magnifying lens and the big one behind the minimizing lens. E then positioned the containers on the left and right sides of the table and asked children which food they should give Duck so that he

¹⁷ As mentioned earlier, while children did not receive the Basic test that chimpanzees received, the Seen trials of the Tracking test performed the same function as the Basic trials. Thus, for the children’s Tracking test, I always began with a Seen trial to preserve the same fundamental sequence of testing as for chimpanzees.

would be happy and quack. For the Unseen trials, which alternated with Seen trials, E again baited the containers and stacked them one atop the other as children watched. In this case, however, just before positioning the containers, she blocked children's view with an opaque screen so that they could not visually track the movement of the food items. After positioning the containers, E removed the screen and asked children which food they should give to Duck to make him quack.

Transfer test. The Transfer test was exactly the same as the Tracking test except that white cubes were used instead of green balls. Because the cubes were very similar in color to the interior of the containers, E placed a black piece of paper on the floor of each container for this test. To help keep the testing time for children within reasonable limits, only Unseen trials were administered (chimpanzees also received Seen trials).

Post-test 'Demo' trials. Now that children had more experience with the lenses, I was interested in knowing how they would perform when they could no longer witness the baiting of the containers, as was the case in the initial Demo trials. The procedure for these trials was exactly the same as for the initial Demo trials except that E did not perform the demonstration after the child made a choice. Instead, she immediately removed the food indicated by the child and gave it to Duck.

Additional appearance-reality tests. Children were given three additional appearance-reality tests, described in turn below.

Standard-format lens test. I was curious to know how children's performance in my Lens test might differ if I asked them directly about appearance and reality, as in standard verbal tests. Thus, after children indicated a container in the second Post-test

trial, E said “Okay, but before I give Duck the food, I would like to ask you something.” Pointing simultaneously to the front of each container, E asked children two questions, in counterbalanced order: (1) When you look at the food through here, which piece looks bigger? and (2) Which piece is really bigger? Children had to answer both questions correctly to pass.

Rock-sponge test. In this standard test, E first placed the rock-sponge on the table and asked, “Can you tell me what this is?” When children identified the object as a rock/stone/etc. (which they did in every case), E allowed them to handle it and discover that it was a sponge. E then put the rock-sponge back onto the table and asked the child what the object looked like and what it really was, counterbalanced for order. Children had to answer both questions correctly to pass. Occasionally a child was hesitant to answer, in which case E repeated the question(s) while stating the two possible options (rock or sponge), in varying order. Thus, E began with open-ended questions and only provided the two response options if necessary. Sapp et al. (2000) used a similar open-ended question format in their verbal response paradigm. I wanted to avoid providing the two response options when possible because of evidence that doing so may lead children to err by giving the same answer to both questions (Deák, 2006).

Nonverbal color filter test. In this test, E stood the frame containing the yellow filter onto the table and made sure the child watched as she moved a piece of white paper behind the filter and out again at least twice. E then set the paper down behind the filter and used the hole puncher to make a flower-shaped hole in it. E then removed the paper flower in her closed hand and put two paper flowers, one white and one yellow, in front

of the child. Pointing to the hole in the paper behind the filter, E asked “Can you show me which of these flowers I just took out of here?” Children were correct if they indicated the white flower.

Verbal color filter test. This test followed immediately upon the Nonverbal color filter test. With the paper still behind the filter, E asked children what color the paper looked like behind the filter and what color it really was, in counterbalanced order. As with the Rock-sponge test, if children were hesitant to respond E restated the question(s) including both possible answers, yellow or white. Children had to answer both questions correctly to pass.

Coding and Analyses

Responses were coded live by the experimenter. For all trials of the Lens test, regardless of condition, the correct response was to choose the truly bigger piece of food. Children did so by touching or pointing at the container in which the bigger piece of food was located. Unless otherwise noted, all statistical analyses were nonparametric and all reported p values are exact and two-tailed.

Reliability

An independent coder naïve to hypotheses and blind to condition coded the choice responses of 25% of randomly chosen participants in each age group. There was perfect agreement with the main coder.

Results and Discussion

Demo

Children clearly demonstrated that they perceived the magnified food as bigger: 27 out of 32 children wrongly chose the magnified food in their first Demo trial; and four of the five children who chose correctly went on to choose wrongly in the second Demo trial.

Lens Test

I chose to report group medians rather than means for the children (Figure 6.2) to be consistent with the chimpanzees' analysis, in which means were not appropriate

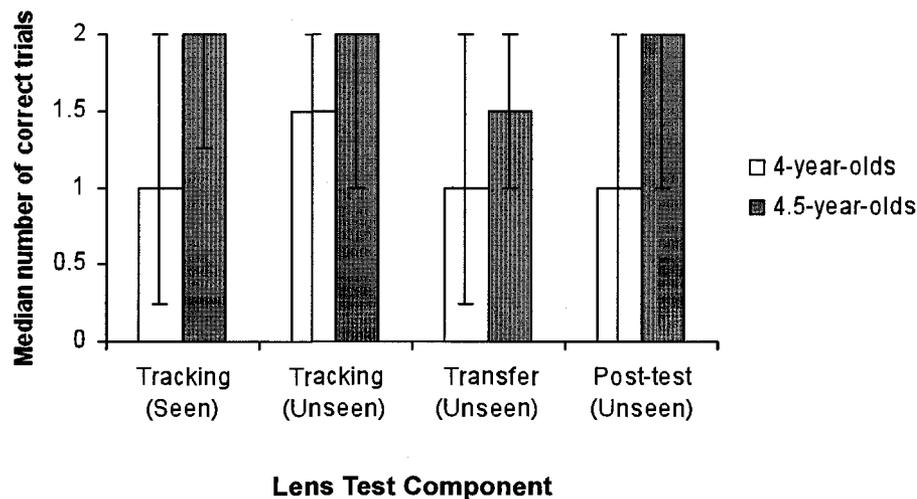


Figure 6.2. Children's median number of correct trials (\pm interquartile ranges) in the various components of the Lens test, Study 4 ($N = 16$ for each bar).

because the data were bimodally distributed (see Experiment 2). There were no significant differences across the various components of the Lens test for either age group

(Friedman tests: 4-year-olds: $X^2 = 0.69$, $df = 3$, $N = 16$, $p = 0.88$; 4.5-year-olds: $X^2 = 2.46$, $df = 3$, $N = 16$, $p = 0.53$), so I analyzed all eight test trials combined. As a group, 4.5-year-olds succeeded in a significantly greater number of trials (median = 7, interquartile range = 5–8) than 4-year-olds (median = 5, interquartile range = 2.25–6.75) (Mann-Whitney U test: $U = 75.00$, $N_{\text{each group}} = 16$, $p = 0.042$). Additionally, 4.5-year-olds performed significantly better than the chance proportion of 0.50 (Wilcoxon test: $T^+ = 109.50$, $N = 15$ [1 tie], $p = 0.003$) but 4-year-olds did not ($T^+ = 66.50$, $N = 14$ [2 ties], $p = 0.39$). This is in keeping with previous studies in which dramatic improvements in appearance-reality test performance have been observed between 4 and 5 years of age.

Additional Appearance-reality Tests

Figure 6.3 shows, for each of the additional appearance-reality tests I administered, the number of children in each age group who passed. Order of

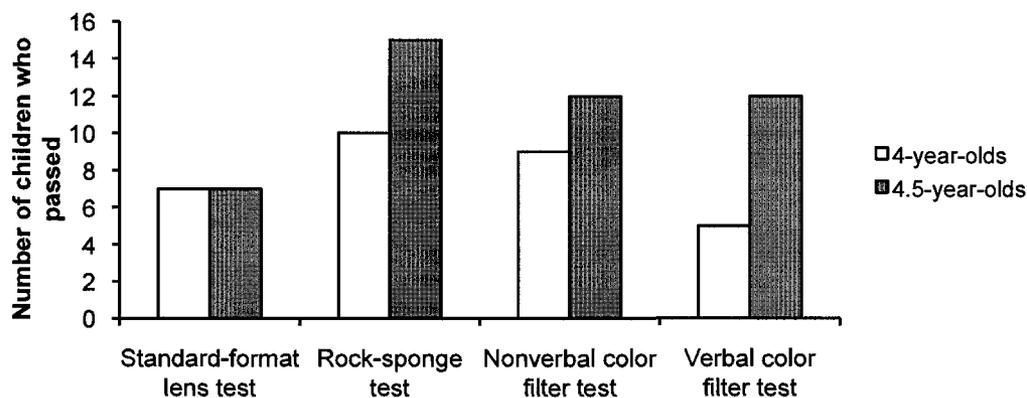


Figure 6.3. Number of children passing the various additional appearance-reality tests administered in Study 4 ($N = 16$ for each bar).

administration of the Rock-sponge test and the two color filter tests did not affect performance ($X^2 \leq 5.33$, $df = 1$, $N = 16$, $p \geq 0.08$ in all cases), and there were no effects of question order within each test ($X^2 \leq 4.27$, $df = 1$, $N = 16$, $p \geq 0.12$ in all cases). I therefore collapsed across these variables.

Within the younger group, there were no significant differences across tests (Cochran's $Q = 3.34$, $df = 3$, $N = 16$, $p = 0.39$), and performance did not differ from chance in any of the tests (Binomial test: $p \geq 0.21$ in all cases). However, for the 4.5-year-olds, there were significant differences across tests ($Q = 11.65$, $df = 3$, $N = 16$, $p = 0.007$). This was due mainly to the older children's poor performance in the Standard-format lens test, which was significantly worse than in the Rock-sponge test (Sign test: $p = 0.008$) and marginally worse than in the Verbal color filter test (Sign test: $p = 0.063$). All other paired comparisons were nonsignificant at $p \geq 0.125$. Older children's performance did not exceed chance in the Standard-format lens test (Binomial test: $p = 0.80$), but it was better than chance in the Rock-sponge test ($p = 0.001$) and approached significance in both color filter tests ($p = 0.077$ in both cases).

It is curious that the 4.5-year-olds did so poorly in the Standard-format lens test, given how well they did in the Lens test. I suspect they were confused over why the experimenter was suddenly questioning them about which piece of food looked bigger and which was really bigger, when in most cases their responses up to that point clearly indicated that they knew this. Children's somewhat weaker performance in both color filter tests relative to the Rock-sponge test may also seem surprising, but it must be kept in mind that the sample size was not particularly large ($N = 16$). Furthermore, many 4.5-

to 5-year-old children may still be in a transitional phase with regard to appearance-reality understanding, and so varying task factors may exert a significant influence in this age range. Both the Rock-sponge test and the color filter tests were administered after the relatively lengthy Lens test, and children's motivation to attend may have been waning by that point. This could have had more of an effect on performance in the color filter tests, which were somewhat longer and more involved than the Rock-sponge test.

Correlations Among Tests

I next examined relations among the various tests, collapsing across age groups to maximize power. To make the Lens test data binary so that it would be consistent in format with the other appearance-reality tests, I converted the data to pass/fail scores using a passing criterion of six or more trials correct out of eight. I used a Phi (Φ) correlation test for dichotomous data, with approximate significance values. As expected, the Lens test was positively correlated with the Standard-format lens test ($\Phi = 0.37, N = 32, p = 0.039$). The latter was also marginally correlated with the Verbal color filter test ($\Phi = 0.32, N = 32, p = 0.067$) which is not surprising given their procedural similarities: in both tests, children witnessed the visual transformation of the object and viewed the object through a transparent barrier. There were no other significant correlations ($p \geq 0.075$ in all cases).

As children received one trial only of each of the additional appearance reality tests I administered, it is possible that their performance in any given test did not truly reflect their capacities. Momentary distraction might have caused them to fail a test they would have otherwise passed, for example. Alternatively, they could have chosen the

correct answer by chance, without understanding the distinction between appearance and reality. Reasons such as these, as well as differing task factors, may have been responsible for the lack of correlation found among the various tests. It has been suggested that it is important to take an aggregate approach when testing for false belief understanding (Hughes et al., 2000; Krachun et al., in preparation), and the same may be true for appearance-reality understanding. Thus, I created an aggregate proportion score for each child from the four additional appearance reality tests (the Standard-format lens test, the Rock-sponge test, and the two color filter tests). I then compared these to the proportion of trials correct in the Lens test, using a Pearson correlation test with approximate significance values. The correlation was highly significant ($r = 0.44$, $N = 32$, $p = 0.01$) and it remained marginally significant when I controlled for age ($r = 0.33$, $df = 29$, $p = 0.070$). This positive correlation was even more apparent when I included only the Unseen trials for the Lens test, which were the more meaningful trials ($r = 0.46$, $N = 32$, $p = 0.009$; $r = 0.42$, $df = 29$, $p = 0.018$).

To summarize, performance in the Lens test was typical for children in the age range I tested. Four-year-olds did not exceed chance performance as a group whereas 4.5-year-olds did. Such a shift in appearance-reality test performance between 4 and 5 years of age is common in research using standard verbal tests (e.g., Flavell et al., 1986). This shift was also apparent in the additional appearance-reality tests I administered, with the exception of the Standard-format lens test. Further, there were positive correlations between the Lens test and the aggregate of the four additional appearance reality tests I administered. All of these factors suggest that the Lens test is a valid measure of

appearance-reality understanding. I next investigated how chimpanzees would perform in an adapted version of the test.

Experiment 2 – Chimpanzees

Methods

Participants

Fourteen chimpanzees (*Pan troglodytes*) housed socially at the Wolfgang Köhler Primate Research Center (WKPRC) in Leipzig, Germany participated: 10 females and four males, 6 to 31.5 years old (mean = 19.3 years, SD = 9.5), all born in captivity, six nursery-reared and eight mother-reared. Three further chimpanzees were dropped from the study, one for being uncooperative, one because of procedural error, and one for failing the pretest criterion (i.e., the *Preference test*; see below). Participants were not food deprived and had water available *ad libitum*. Eight chimpanzees had previously taken part in studies employing reverse contingency procedures (see Table 6.1).

Design

To take part in the Lens test, chimpanzees first had to demonstrate in a Preference test that they did indeed prefer larger grapes to smaller ones. Those who did so were then given a demonstration (Demo) of the properties of the magnifying and minimizing lenses. This was followed by the various components of the Lens test in a fixed order: (1) a two-session Basic test, (2) a two-session Tracking test, (3) a two-session Transfer test, and (4) a four-session Mixed test. Detailed procedures for these tests are given below in the

Procedure section. Each session consisted of 12 trials. For the Basic test all trials were Seen; for all other tests half the trials were Seen and half Unseen.

Because of the large number of trials involved for chimpanzees, doing well in earlier stages of testing served as a criterion for participation in later stages. Only those individuals who exceeded chance performance in the Basic test proceeded to the Tracking test, and only those who exceeded chance in the Tracking test proceeded to the Transfer and Mixed tests. Exceeding chance performance was defined as choosing the truly bigger food significantly more often than chance in either the second session or across both sessions combined. In addition to the test trials, all chimpanzees received the first reverse contingency control test (RC-1), consisting of four 12-trial sessions. Half the chimpanzees did this test before the Lens test and half afterwards. Chimpanzees who exceeded chance in the Basic test were also given the second reverse contingency control test (RC-2), consisting of two 12-trial sessions, after all other testing was complete.

For all sessions, the bigger and smaller food items were on the left and right an equal number of times and each container was baited first an equal number of times. The position of the food items and which container was baited first were randomly determined with the constraint that neither factor could repeat for more than three consecutive trials.

Materials

Lens test. The magnifying and minimizing containers were the same as the children's (see Figure 6.1). Small grapes (2.0 x 1.5 cm) and big grapes (2.75 x 1.75 cm) were used for the Basic test and Tracking test (sizes refer to approximate diameters along

the longer and shorter dimensions). A thin slice was cut from the bottom of each grape to keep it from rolling. For the Transfer test and Mixed test, small (1.25 cm³) and big (1.75 cm³) apple cubes were used. To maximize the visibility of the food while minimizing glare, a light was shone down into the containers from above. This was not necessary for the children because the daycare centers where they were tested were very brightly lit. Note that, unlike in the previous liquid conservation studies carried out with apes (Call & Rochat, 1996, 1997; Muncer, 1983; Suda & Call, 2004; Woodruff et al., 1978), the containers I used to hold the rewards were identical in appearance—all that varied was the appearance of the food reward itself. This avoided the problem of chimpanzees being influenced by differences in the appearance of the containers

Reverse contingency (RC) control tests. Containers for both reverse contingency tests also required chimpanzees to view the food through a transparent barrier (equivalent to the lenses in the Lens test), but the appearance of the food was not modified (Figure 6.4). For all trials, I used small and big grapes that approximated the apparent sizes of the grapes as viewed through the lenses in the Lens test.

RC-1 test. Two identical opaque rectangular tins (16 x 11 x 7 cm) open at the back end were used. A piece of rectangular clear plastic (10 x 8 cm) with a narrow shelf on top was mounted to the front of each container. Chimpanzees could see food that was placed onto the surface of the container just behind the plastic, but they could not see food that was placed inside the container.

RC-2 test. Containers were two semicircular pieces of clear plastic (10 cm wide, 14 cm high), each mounted onto a 10 x 10 cm base. The top half of each container was

covered in black plastic and contained a hidden shelf onto which a grape could be placed. Chimpanzees could thus see food in the bottom half of the container but not the top half.

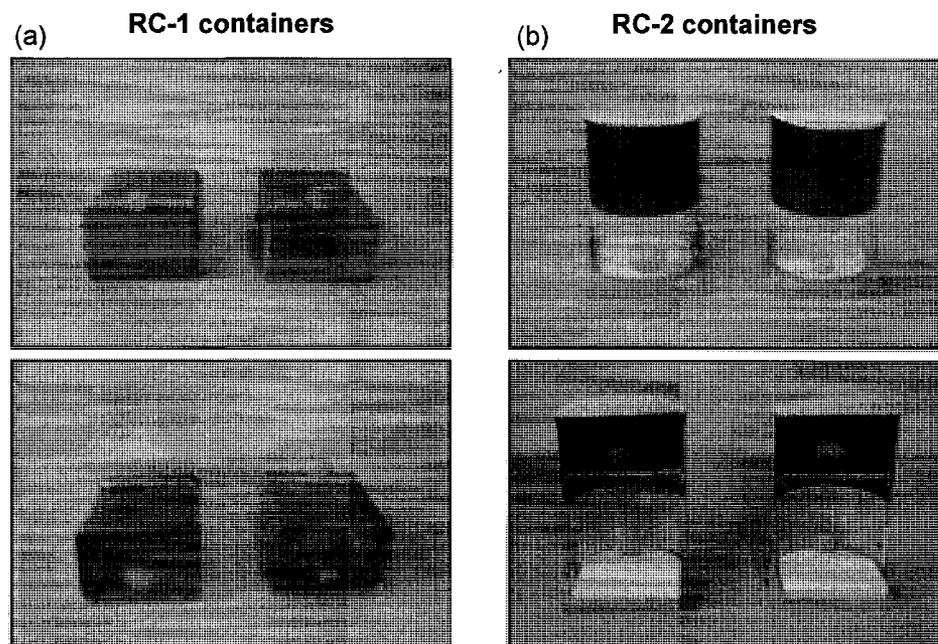


Figure 6.4. Containers used in the (a) RC-1 and (b) RC-2 reverse contingency control tests of Study 4, as viewed from the ape's perspective (top row) and the experimenter's perspective (bottom row).

Procedure

Chimpanzees were tested individually in a familiar enclosure by a female experimenter (myself). The experimenter (E) sat facing the chimpanzee through a Plexiglas window (69 x 49 cm) with a sliding table mounted below it, approximately 50 cm above floor level. When E slid the table forward, chimpanzees could choose a container by poking a finger through one of two holes spaced 58 cm apart on the bottom left and right sides of the window.

Preference test. To take part in the study, chimpanzees had to first demonstrate that they preferred bigger to smaller grapes by choosing the bigger one significantly more often than the chance proportion (0.50). E put the containers onto the table, one on the left and one on the right, and simultaneously placed a big grape on top of one and a small grape on top of the other. She then slid the containers within reach of chimpanzees so they could choose. The grapes were not lowered behind the lenses in these trials. Chimpanzees only received a second session if they did not exceed chance in the first session. For the remainder of the study, two Preference trials were repeated at the beginning of each test session. If chimpanzees did not choose the bigger grape in both trials they were not tested that day (this happened only twice).

Demo. The Demo trials were particularly important for chimpanzees because they had absolutely no prior experience with magnifying or minimizing lenses. Additionally, it is as yet unknown whether chimpanzees are subject to visual illusions in the same way that humans are. Thus, the Demo trials allowed me to both familiarize chimpanzees with the properties of the lenses and to verify that chimpanzees actually perceived the magnified grape as larger. If they did perceive the illusion, they would be expected to choose the magnified grape in the very first Demo trial, because they had no reason to suspect at that point that they were being visually misled.

The Demo procedure was as follows: E put an occluder in front of the window to block the chimpanzee's view. She placed a small grape behind the magnifying lens in one container and a big grape behind the minimizing lens in the other container, so that the grapes' apparent relative sizes would be the reverse of reality. Because of their spheroid

shape, the grapes scaled upward and downward in appearance with minimal distortion. E then removed the occluder blocking the chimpanzee's view of the containers. When she was confident the chimpanzee had looked at the front of both containers she slid them forward so the chimpanzee could choose. However, before giving chimpanzees their chosen grape, E slowly removed the grape from behind the lens and replaced it again, repeating this action until the chimpanzee had witnessed the visual transformation of the grape at least twice. E then performed the demonstration with the other grape behind the other lens. Two such Demo trials were repeated at the beginning of every test session, immediately before the test trials, in order to refresh chimpanzees' memory of the properties of the lenses. Additionally, these trials allowed me to see how chimpanzees, after they had gained more experience with the lenses, would perform in trials in which they could not witness the baiting of the containers (recall that this was the same reason for giving the children two Post-test trials of this type in Experiment 1).

Additional pretest experience. I was curious to know if more varied experience with the lenses, beyond that chimpanzees received in the Demo trials, would make a difference to their performance in the test trials. I therefore gave six chimpanzees two additional pretest sessions (12 trials each) in which they experienced the effect of the lenses on a variety of foods including pretzels, peanuts, tea biscuits, and pieces of banana, kiwi, and orange.¹⁸ Grapes were still used for the two Preference and Demo trials at the beginning of these sessions. Each of the chimpanzees experienced the effects

¹⁸ Two chimpanzees who were ultimately dropped from the study also received these sessions.

of the lenses on at least four of these foods depending on individual preferences, and the foods were presented in randomized order. The procedure for these sessions was the same as for the Basic test below. The eight chimpanzees who did not participate in these sessions proceeded directly from the Demo to the Lens test.

Lens test. The various components of the appearance-reality Lens test were as follows: a Basic test, a Tracking test, a Transfer test, and a Mixed test. Each of these is described in turn below.

Basic test. This first stage was my simplest test of chimpanzees' ability to distinguish appearance from reality. As in the Preference trials, E placed a small grape atop the magnifying container and a big grape atop the minimizing container. As chimpanzees watched, she slowly lowered each grape in turn into the container and behind the lens. When E was confident the chimpanzee had looked at both containers (either during baiting or afterwards), she allowed the chimpanzee to choose. E then removed both grapes and rewarded the chimpanzee with the chosen grape. To ensure that chimpanzees would not just choose the first or last container E touched, she simultaneously touched both containers before sliding them forward. Also, so that chimpanzees could not learn to avoid the magnifying lens by noting minute differences in the appearance of the containers, halfway through each session E surreptitiously switched the front panels of the containers, along with the lenses mounted in them. As a further precaution, before each session E also removed each lens and remounted it in the other panel.

Chimpanzees who exceeded chance in the Basic test proceeded directly to the Tracking test. Those who failed the Basic test, however, may have done so because they were unable to inhibit reaching for the grape that appeared bigger, even if they knew it was truly smaller. These individuals were therefore given a *Reduced inhibition test*, consisting of two sessions the same as the Basic sessions except that E covered the lenses just before letting the chimpanzee choose. Thus, if being confronted by the image of the grapes at the moment of choice was causing inhibition problems for these chimpanzees in the Basic test, they would be expected to perform better in the Reduced inhibition test. If they did indeed perform better, they would then be allowed to proceed to the Tracking test. The experimenter would also continue, throughout the remainder of testing, to cover the lenses for these chimpanzees just before allowing them to choose. This never occurred, however, because all six chimpanzees who failed the Basic test also failed the Reduced inhibition test.

Tracking test. I recognized that chimpanzees who did well in the Basic test could have done so by visually tracking the bigger grape as E placed it behind the minimizing lens. Of course, to do so would necessarily involve ignoring the misleading appearance of the grapes behind the lenses, suggesting some understanding that the grapes' appearance conflicted with reality. However, stronger evidence that chimpanzees recognized that the apparently bigger grape was in reality the truly smaller one would be if they could succeed without the opportunity to visually track the grapes.

As with the children, there were two conditions in the Tracking test: Seen and Unseen. Seen trials were similar to the Basic test in that chimpanzees could still visually

track the movement of the bigger grape, although they now had to pay slightly closer attention in order to successfully do so. In Unseen trials, however, E made it impossible for chimpanzees to visually track the grapes. In both conditions, E began by placing one container in the middle of the table and lowering a grape into it as the chimpanzee observed. E then stacked the second container on top of the first and lowered the second grape into it. Again, the small grape always went behind the magnifying lens and the big grape behind the minimizing lens. In Seen trials, E positioned the containers on the left and right sides of the table in full view of chimpanzees and then let them choose. In Unseen trials, however, before positioning the containers E blocked chimpanzees' view with an occluder so that they could not visually track the movement of the grapes. Chimpanzees received six Seen and six Unseen trials in each session, randomly ordered with not more than three consecutive trials of the same kind. Those who exceeded chance proceeded to the Transfer test.

Transfer test. It was important to see if chimpanzees who performed well in previous sessions would transfer their abilities over to a different type of food. If they could not do so, it would suggest that they believed the distorting effects of the lenses to be specific to grapes. In contrast, succeeding in the Transfer test would suggest that chimpanzees had come to understand something about the lenses more generally—namely, that they caused whatever was placed behind them to have a misleading appearance. Two Transfer sessions were administered identical to the Tracking test described above, except that small and big apple cubes were used instead of grapes (for

the Preference and Demo trials as well as the test trials).¹⁹ Regardless of their performance, all chimpanzees who participated in the Tracking test were allowed to participate in the subsequent Mixed test, to avoid reducing the number of participants too greatly for this final test.

Mixed test. This last stage of the Lens test was the most challenging for chimpanzees. It allowed me to see how flexibly chimpanzees could apply the knowledge they demonstrated in earlier trials. It also allowed me to test whether chimpanzees had succeeded in previous sessions by simply learning to avoid the magnifying lens. For half the trials, the procedure was exactly the same as in the Transfer test, with E placing the big apple cube behind the minimizing lens (the *Big-to-small* condition). Chimpanzees were thus expected to perform the same in these trials as in the previous Transfer trials. For the other half of trials, however, E now placed the big food behind the magnifying lens (the *Big-to-bigger* condition). If chimpanzees had simply learned to avoid the magnifying lens, they would be expected to avoid the apparently bigger (and now also truly bigger) food in these trials and to choose the apparently (and truly) smaller piece.²⁰

¹⁹ For the first three chimpanzees tested, I lined the interior of the containers with black plastic to improve the visibility of the food, as was done for the children. However, I decided that this was not necessary for the chimpanzees because the apple cubes were off-white and lit with supplementary lighting from above, so they were clearly visible through the lenses. I was also concerned that changing the appearance of the containers might make chimpanzees think that they were not the same containers as in previous trials. For these reasons, I left the interior of the containers white throughout all trials for the remaining chimpanzees.

²⁰ Before this Mixed test, the first three participants (Fifi, Jahaga, and Trudy) also received one session consisting entirely of Big-to-bigger trials. Its purpose was to determine, *before* moving on to the more complex Mixed test, whether the chimpanzees

If they were *not* simply avoiding the magnifying lens, however, they would continue to respond correctly in these trials by choosing the truly bigger piece of food. Note that to prepare chimpanzees for this test, I also modified the two Demo trials given at the start of each session. First, there was a *Pre-demo* in which E repeatedly lowered a big and a small apple cube, in turn, behind each lens while chimpanzees observed. The two Demo trials were then administered as before, except that in one of these trials (counterbalanced for order) E put the big apple cube behind the magnifying lens and the small apple cube behind the minimizing lens.

To test how flexibly chimpanzees could apply their appearance-reality knowledge, in both conditions (Big-to-small and Big-to-bigger) I made half the trials Seen and half Unseen, with the Unseen trials constituting the greatest test of chimpanzees' abilities. To succeed in these trials, chimpanzees had to note the transformation in the appearance of the food as E lowered it into the container: Did the truly bigger piece now appear relatively bigger or smaller than the other piece? E then blocked chimpanzees' view before positioning the containers on the table. When E then unblocked chimpanzees' view and allowed them to choose, they had to remember which food they should search for, the apparently smaller one or the apparently bigger one.

had learned to avoid the magnifying lens. All three chimpanzees chose the apparently (and truly) bigger piece of food in every trial. Because the results were quite clear, and because the Mixed test included trials of the same type, I decided this session was unnecessary. I was also concerned that experiencing an entire session of Big-to-bigger trials was leading chimpanzees to adopt the simple strategy of always choosing the apparently bigger food, which would negatively affect their overall performance in the Mixed test. I therefore did not administer this session to the remaining chimpanzees.

Chimpanzees were given 24 Big-to-small trials and 24 Big-to-bigger trials, three Seen and three Unseen of each per session for four sessions. Order of the various conditions was randomized, with the constraint that no factor could repeat for more than three consecutive trials.

First reverse contingency test (RC-1). I anticipated the argument that chimpanzees could pass the Lens test by learning a simple reverse contingency rule. I therefore administered a reverse contingency test, which I made as similar as possible to the Lens test without including misleading appearances. If chimpanzees could pass the Lens test but not the reverse contingency test, this would indicate that they were not solving the Lens test by learning a reverse contingency rule.

To minimize any possible transfer effects (positive or negative) that could occur if chimpanzees connected this test with the Lens test, I had a different experimenter (a female student assistant) administer the reverse contingency trials, using the containers shown in Figure 6.4a. The procedure was as follows: E blocked the chimpanzee's view with an occluder and placed a big grape onto the shelf of one container and a small grape onto the other. E also placed a grape of the opposite size inside of each container. Thus, the container with the big grape displayed on its shelf had a small grape hidden inside of it, and vice versa. E then removed the occluder and slowly lowered each grape in turn down onto the surface of the container, just behind the transparent barrier. This action was analogous to lowering the grapes behind the lenses in the Lens test, but the appearance of the grape was not transformed. When chimpanzees made their choice, E took the chosen grape from behind the transparent barrier and dropped it into a nearby

bucket. She then took the hidden grape out of the container and gave it to the chimpanzee. Removing the chosen grape from view just before revealing the other grape was analogous to lifting the chosen grape out from behind the lens in the Lens test.

After running RC-1, I realized that the test differed from the Lens test in one potentially important respect. In the RC-1 test, at the beginning of each trial, E blocked chimpanzees' view before placing one grape on top of each container and one (of opposite size) inside each container. Thus, chimpanzees did not have a chance to see the actual big grape that they would receive if they picked the small one, and vice versa. In the Lens test, however, chimpanzees watched as E lowered the grapes into the containers, so they were able to see what their actual reward would be if they were to choose a given container. This difference could have made the RC-1 test harder for the chimpanzees than the Lens test. To address this potential problem, I created a modified reverse contingency test (the RC-2 test), described below.

Second reverse contingency test (RC-2). A third experimenter (the study supervisor) administered the RC-2 test, again to avoid possible transfer effects. Also, to minimize the chances that chimpanzees would associate this test with the RC-1 test, different containers were used (Figure 6.4b). So that the RC-2 test would be more similar in procedure to the Lens test than RC-1, E did not block chimpanzees' view of the containers while placing the hidden grapes inside. Instead, E held up both a small and big grape together in one hand for the chimpanzee to see and then moved the grapes downward, depositing one into the opaque top part of the container and the other into the transparent lower part. E then repeated this with the other container, reversing the

locations of the small and big grapes. Thus, analogous to the Lens test, chimpanzees saw a small grape disappear into one container while a big grape became visible and a big grape disappear into the other container while a small grape became visible.

Coding and Analyses

Chimpanzees' choice of container was coded live. All statistical analyses were nonparametric and all p values reported below are exact and two-tailed.

Reliability

An independent coder naïve to the hypotheses of the study and blind to condition coded chimpanzees' choice in 10% of trials. A smaller percentage of trials was coded for chimpanzees than for children because chimpanzees received a much larger number of trials overall. Sessions were chosen randomly, with the constraint that each chimpanzee and each type of test, including reverse contingency tests, be represented at least once. Reliability was excellent (Cohen's kappa = 0.98, $p < 0.001$).

Results and Discussion

Preference Test and Demo

Chimpanzees unequivocally demonstrated that they preferred big grapes to small ones. Thirteen out of 14 chimpanzees chose the bigger grape in at least 10/12 trials of the Preference test (Binomial test: $p < 0.02$). The remaining chimpanzee scored 9/12 and was thus given a second session, in which she scored 12/12. Chimpanzees clearly also perceived the visual illusion created by the lenses. In their very first Demo trial, before they had any experience with the lenses, 12/13 chimpanzees wrongly chose the magnified grape (one

chimpanzee, Fifi, was inadvertently not given the opportunity to choose in her first Demo trial). Eleven of these 12 chimpanzees again chose the magnified grape in their second Demo trial.

Lens Test

The data were bi-modally distributed, with most chimpanzees doing either very well or very poorly in the test. I therefore examined individual rather than group performance, using Binomial tests to determine whether chimpanzees passed or failed the various test components (see Table 6.1). In the Basic test, chimpanzees passed if they

Table 6.1

Number of correct trials for individual chimpanzees in various components of the Lens test, Study 4, broken down by session. Bold text indicates a passing score (significantly better than chance at $p < 0.05$). Chimpanzees who did not pass a given test by the second session are excluded here from the subsequent test.

Ape	Age (years)	Rearing	RC ^b history	Basic test		Unseen ^a tracking test		Unseen transfer test		Unseen mixed test (Big-to-small) trials	
				Sess 1 (12 trials)	Sess 2 (12 trials)	Sess 1 (6 trials)	Sess 2 (6 trials)	Sess 1 (6 trials)	Sess 2 (6 trials)	Sess 1&2 (6 trials)	Sess 3&4 (6 trials)
Corrie	30.5	nursery	none	2	4
Dorien	26.5	nursery	1, 2	1	0
Fraukje	31	nursery	2	2	0
Natascha	27	nursery	none	6	6
Ulla	30	nursery	none	3	6
Patrick	10	mother	2	1	0
Robert	31.5	nursery	none	7	10	1	1
Trudy	14	mother	2	7	12	5	3
Frodo	13.5	mother	1	12	12	5	4
Lome	6	mother	none	10	12	2	6	4	5	.	.
Jahaga	14	mother	1	11	12	6	6	1	6	3	1
Sandra	14	mother	1, 2	10	12	6	6	4	6	5	3
Pia	8	mother	none	12	12	5	6	5	6	4	3
Fifi	14	mother	2	2	10	6	6	1	6	2	0

^a Unseen trials only are shown for the Tracking, Transfer and Mixed tests because these trials were the true test of apes' abilities.

^b RC history = previous reverse contingency experience: 1 = Vlamings, Uher and Call (2006); 2 = Uher and Call (2008)

chose the truly bigger food significantly more often than chance ($p < 0.05$) by the second session. Eight chimpanzees passed the Basic test, and five of them even exceeded chance within the first session. Furthermore, looking at just the first six trials, five chimpanzees got at least 5/6 trials correct. Thus, in the Basic test, more than half the chimpanzees were capable of ignoring the misleading appearance of the magnified grape to choose the truly bigger one, and about a third were capable of doing so from the earliest trials.

Of course, the fact that chimpanzees watched the baiting introduces the possibility that they succeeded by visually tracking the bigger grape, without attending to the appearance of the grapes behind the lenses. The eight chimpanzees who passed the Basic test were thus also given a Tracking test. The Unseen tracking trials were the true test of chimpanzees' ability to succeed without visually tracking the food. Thus, for this analysis, I used a stricter 'pass' criterion than the criterion I used to determine whether chimpanzees would move on the next stage of testing. I considered chimpanzees to have passed the Tracking test only if they exceeded chance performance in the Unseen trials separately (in addition to exceeding chance in all trials, Seen and Unseen). Five of the eight chimpanzees who received the Tracking test passed according to this criterion, and three of them got all six Unseen trials correct in the first session, thereby demonstrating that they could choose the truly bigger grape (despite its misleading appearance) without tracking. Four of these chimpanzees also passed the subsequent Transfer test with apples (using the same pass criterion as above for the Tracking test), indicating that they did not consider the effect of the lenses to be specific to grapes. The fifth chimpanzee was marginally significant with a total of 9/12 Unseen trials correct. While no chimpanzee got

all six Unseen trials correct in the first Transfer session, one chimpanzee got 5/6 correct and two chimpanzees got 4/6 correct.

For the Mixed test, I first looked at chimpanzees' responses in the Big-to-bigger condition, in which the truly bigger food was placed behind the magnifying lens. Chimpanzees clearly demonstrated in these trials that they had not simply learned in previous sessions to avoid the magnifying lens. All seven chimpanzees who were given the Mixed test correctly chose the truly bigger food in 100% of the Seen trials, and six chimpanzees chose it in 100% of the Unseen trials. I next looked at chimpanzees' performance in the Big-to-small condition, in which the truly big food was placed behind the minimizing lens, as in the Basic, Tracking, and Transfer tests. Four chimpanzees passed the Seen big-to-small trials, choosing the truly bigger food in 10 or more trials out of 12, but the other three chimpanzees failed. Further, not one chimpanzee passed the Unseen big-to-small trials (see Table 6.1) despite the fact that these trials were exactly the same as the previous Unseen transfer trials. I thus assume that the Mixed test, in which the truly bigger food was sometimes magnified and sometimes minimized, was too confusing for the chimpanzees, leading them to respond randomly or to fall back on the simple strategy of always choosing the visually larger food.

Comparison with the Reverse Contingency Control

In previous reverse contingency tests carried out by other researchers, chimpanzees have show a strong bias toward choosing larger quantities or sizes over smaller ones (Boysen & Berntson, 1995; Boysen et al., 1996; Boysen et al., 2001). Individuals who successfully overcome this bias do so only after hundreds of trials (Uher

& Call, 2008; Vlamings, Uher, & Call, 2006). In keeping with these findings, and in sharp contrast to the Lens test, not one chimpanzee passed either of the reverse contingency control tests (RC-1 or RC-2), regardless of whether I counted all sessions or just the last session (Binomial test: $p > 0.05$ in all cases). For each reverse contingency test, I also compared group performance in each session with each corresponding session

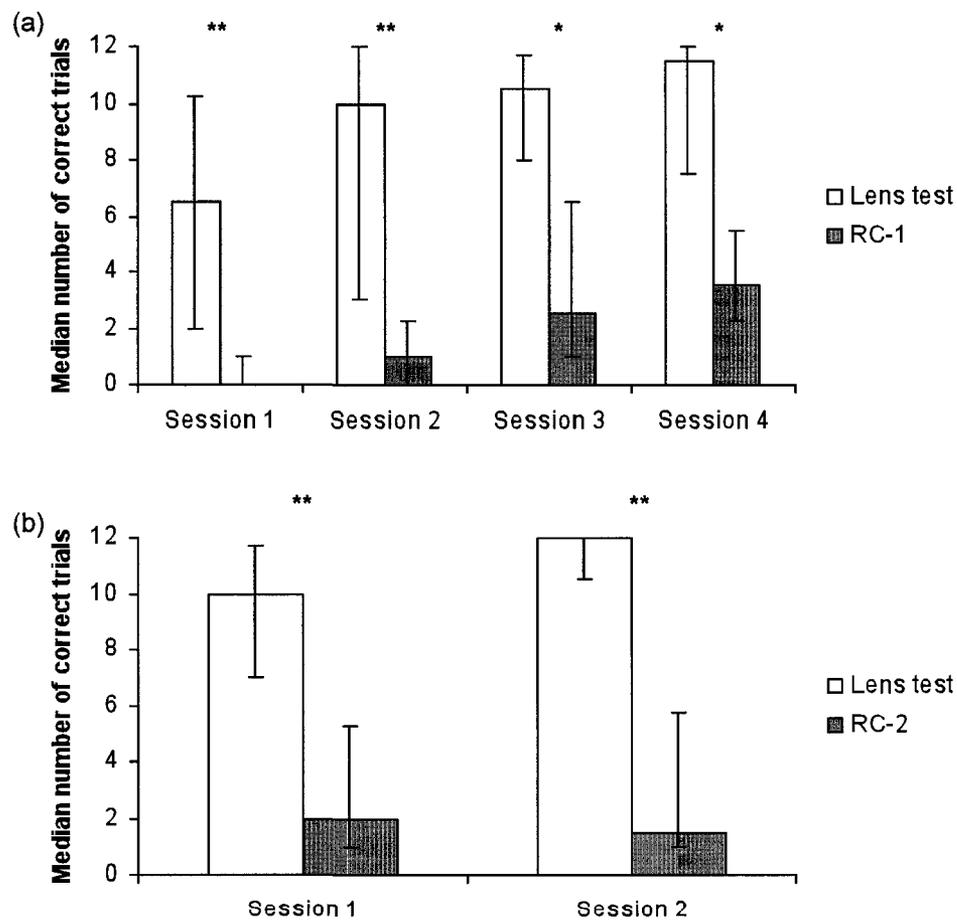


Figure 6.5. Chimpanzees' median number of correct trials (\pm interquartile ranges) in Study 4 in the Lens test as compared to the two reverse contingency control tests: (a) RC-1 ($N = 14$ in sessions 1 and 2; $N = 8$ in sessions 3 and 4); and (b) RC-2 ($N = 8$ in both sessions). * $p < 0.05$, ** $p < 0.01$

of the Lens test, including all Seen and Unseen trials (see Figure 6.5). The median number of trials correct was significantly higher in the Lens test than in RC-1 (Wilcoxon tests: $T^+ \geq 77.00$, $N = 12$ [2 ties], $p \leq 0.005$ in both Sessions 1 and 3; $T^+ \geq 35.00$, $N = 8$ [0 ties], $p \leq 0.016$ in both Sessions 3 and 4). This was also the case when comparing the Lens test with RC-2 ($T^+ = 36.00$, $N = 8$ [0 ties], $p = 0.008$ in both sessions). The fact that chimpanzees did so poorly in the reverse contingency tests relative to the Lens test suggests that they were *not* using a simple reverse contingency rule to solve the Lens test.

Other Variables

There were no significant effects on performance of (1) extra pretest experience with different foods, (2) previous reverse contingency experience, or (3) order of administration of the Lens and reverse contingency tests (Mann-Whitney U tests: $p \geq 0.10$ in all cases). However, the eight mother-reared chimpanzees did significantly better as a group than the six nursery-reared chimpanzees in the Basic test ($U = 7.00$, $p = 0.027$). This was surprising given that, generally, when differences are detected it is usually the human-reared apes outperforming mother-reared apes (see review by Call & Tomasello, 1996). However, the average age of the mother-reared chimpanzees (11.7 years, $SD = 3.2$) was also significantly younger than that of the nursery-reared group (29.4 years, $SD = 2.13$) ($U = 0.00$, $p < 0.001$). Some researchers have found similar age effects for chimpanzees in other tasks that may involve self-knowledge, including mirror self-recognition (Povinelli, Rulf, Landau, & Bierschwale, 1993) and episodic memory (Martin-Ordas, Call, Haun, & Colmenares, under review). In those studies, as in this one,

adolescents and young adults performed better than very young or older individuals. In the Basic test, only two individuals did not fit this pattern of results: Patrick was younger (10 years old) but failed, and Robert was older (31.5 years old) but passed. It is therefore possible that the difference in performance between mother-reared and nursery-reared chimpanzees was in fact due to age.

Comparison of Chimpanzees and Children

Although the chimpanzees and children were not given exactly the same collection of tests (e.g., the children did not receive the Basic test and chimpanzees did not receive the additional appearance reality tests), it is still possible to make some direct comparisons across species. Both chimpanzees and children participated in the Seen and Unseen tracking trials of the Lens test, for example. There were no significant differences in performance across species in those trials (Mann-Whitney U tests: chimpanzees vs. 4-year-olds: $U \leq 55.00$, $N_{\text{chimpanzees}} = 8$, $N_{\text{children}} = 16$, $p \geq 0.39$; chimpanzees vs. 4.5-year-olds: $U \leq 59.00$, $N_{\text{chimpanzees}} = 8$, $N_{\text{children}} = 16$, $p \geq 0.26$). There were also no significant differences across species in the Unseen transfer trials, which both species also received (chimpanzees vs. 4-year-olds: $U = 53.00$, $N_{\text{chimpanzees}} = 7$, $N_{\text{children}} = 16$, $p = 0.84$; chimpanzees vs. 4.5-year-olds: $U = 42.00$, $N_{\text{chimpanzees}} = 7$, $N_{\text{children}} = 16$, $p \geq 0.36$). The children's two Post-test trials might also be considered equivalent to the chimpanzees' two Demo trials in the second session of their Transfer test, because in both cases participants had experienced fundamentally the same sequence of tests by that point (namely, the Basic test and/or Seen tracking test, followed by the Unseen tracking test, followed by the Transfer test). Once again, there were no differences in performance

across species in those trials (chimpanzees vs. 4-year-olds: $U = 45.00$, $N_{\text{chimpanzees}} = 7$, $N_{\text{children}} = 16$, $p = 0.56$; chimpanzees vs. 4.5-year-olds: $U = 41.00$, $N_{\text{chimpanzees}} = 7$, $N_{\text{children}} = 16$, $p \geq 0.31$).

In summary, I found no significant differences in performance between children and chimpanzees on any of the tests they had in common. However, chimpanzees received a far greater number of trials than children, and so one could argue that the very first test trial for all participants (i.e., the first Basic trial for chimpanzees and the first Seen tracking trial for children) may be a more meaningful comparison. When I examined just that trial there were still no significant differences between chimpanzees and 4-year-old children ($X^2 = 1.27$, $df = 1$, $N_{\text{chimpanzees}} = 14$, $N_{\text{children}} = 16$, $p = 0.30$). However, the proportion of 4.5-year-old children responding correctly (0.75) approached being significantly greater than the proportion of chimpanzees responding correctly (0.36) ($X^2 = 4.69$, $df = 1$, $N_{\text{chimpanzees}} = 14$, $N_{\text{children}} = 16$, $p = 0.063$). This suggests that chimpanzees were more similar in performance to the 4-year-olds than to the 4.5-year-olds. Further supporting this is that the proportion of 4-year-old children who passed the Lens test (0.25) was very similar to the proportion of chimpanzees who passed all of the Basic, Tracking and Transfer components of the Lens test (0.29).

General Discussion

Like the young 4-year-olds I tested, chimpanzees did not do well as a group in the Lens test. Also like the children, however, some individual chimpanzees did succeed. Thus, while the capacity to distinguish appearance from reality may not be widespread among chimpanzees, it may nevertheless be present in some proportion of the

population, perhaps especially among adolescents and young adults. Similar age effects have been found for chimpanzees regarding other abilities, such as mirror self-recognition (Povinelli et al., 1993) and episodic memory (Martin-Ordas et al., under review). Of course, it is also possible that appearance-reality understanding is common among chimpanzees but that some of them found my task difficult for other reasons. Similarly, some researchers have suggested that younger children's problems with appearance-reality tasks are more a function of problematic task factors rather than any inability to distinguish appearance from reality (e.g., Deák, 2006; Hansen & Markman, 2005; Sapp et al., 2000). One possibility is that the chimpanzees who failed even my Basic test may have been unable to inhibit choosing the grape that appeared larger, even if they did recognize that it was not truly larger (Boysen et al., 2001). Although I attempted to remove this 'pull of the apparent' by covering the lenses just before letting these individuals choose, this may not have effectively removed the impact of what they had seen a brief moment before. Completely eliminating response-inhibition issues in tests with young children or nonhuman animals is a continuing challenge (see, for example, Bialystok & Senman, 2004).

The other side of the coin is that the chimpanzees who passed the test may have done so not because they were capable of distinguishing appearance from reality but for some other reason. Based on the results of the control tests and other manipulations, we can rule out a number of alternative explanations. We know that successful chimpanzees did not simply learn to choose the small grape to obtain the big one because they could not learn to do this in the reverse contingency control tests. We also know that they did

not just learn to avoid the visually larger food behind the magnifying lens because they did not do so when the truly bigger piece of food was magnified in the Big-to-bigger trials of the Mixed test. And some individuals clearly did not solve the task by visually tracking the grapes because they still succeeded when I made this impossible in the Unseen tracking trials. The data, at the very least, show that some chimpanzees were able to overcome the misleading perceptual appearance of the food with regard to its size. Additionally, data on object individuation (Mendes, Rakoczy, & Call, 2008) has demonstrated that apes who watch an object disappear into a box and then later retrieve it from that box know that it is the very same object. Thus, I conclude that several of the chimpanzees did indeed appreciate that what they saw was at odds with what they knew to truly be the case. They recognized that the very same piece of food that looked smaller was really the bigger one, and they chose accordingly.

These findings have important implications for current debates on metacognition in nonhuman animals. For example, can the successful chimpanzees in this study be described as showing evidence of metacognition, in the sense of reflecting upon their own mental states? Flavell et al. (1983) suggests that appearance-reality understanding is indeed metacognitive: “although always susceptible to being deceived by appearances, we have acquired the metacognitive knowledge that appearance-reality differences are always among life’s possibilities” (p. 96). Carruthers (2008, p. 86) also implies that if chimpanzees were to pass an appearance-reality test, this would be a more convincing indicator of metacognition than previous studies have offered. Put very simply, Carruthers’ issue with the evidence to date is that the behaviors claimed to be indicative

of metacognitive processes in animals can be explained without invoking any kind of reflection on one's own mental states. An animal need only have different mental states of varying strengths competing with one another, with the stronger exerting more influence over behavior than the weaker. When no potential response is more strongly motivated than any other, for example because the mental states involved are of similar strength, animals tend to react by seeking more information or choosing randomly.

Succeeding in my appearance-reality test is difficult to explain in Carruthers' terms, because the beliefs and perceptions pitted against one another would seem to have equal strengths (i.e., a strong belief that the big grape is in container A is matched by a strong perception that the big grape is in container B). Under these conditions, and when further information seeking is not an option (as it was not in this case) chimpanzees would be expected to choose randomly. Conceivably, beliefs could win out over perceptions if they were for some reason naturally stronger; however, if anything, we would expect perceptions to exert the stronger influence. In Chapter 3, I discussed studies demonstrating that when young children are asked about the real and apparent properties of objects, such as size or color, they most often err by answering both questions based on the object's perceptual appearance (Flavell et al., 1983; Taylor & Flavell, 1984). In keeping with this, in my Standard-format lens test children made 15 such phenomenism errors but only three realism errors (i.e., giving reality answers to both questions). Thus, if chimpanzees did not recognize that the apparently bigger grape was the truly smaller one, we would expect them to be heavily biased toward choosing that grape. Some chimpanzees did so, but some did not, and those that did not are

evidence that at least some individual chimpanzees are capable of grasping the appearance-reality distinction.

As far as I am aware, the Lens test presented in this chapter is the only appearance-reality test ever developed for apes. The current study thus constitutes the first experimental demonstration that chimpanzees are capable of passing an appearance-reality test. This is exciting news, given their repeated failures in false belief tests. It shows that although representing others' false belief states may be beyond chimpanzees, they are not entirely incapable of recognizing that mental states can be false. And many would argue that this recognition signifies that one understands something very fundamental about the mind—that it does not simply reflect, but rather *represents*, reality (e.g., Wellman et al., 2001). The puzzle at this point, then, is why some chimpanzees are capable of passing appearance-reality tests but not false belief tests, especially given how closely related these abilities appear to be in human children (Flavell, 1993). I come back to this in the following, and final, chapter of this thesis.

CHAPTER 7 – FINAL COMMENTS

Summary and Meta-analysis of Findings

In this thesis, I investigated various aspects of mentalizing in apes that have up until now received little attention. I focused on the mental states of seeing, knowing and believing in particular because unlike intentions, desires, and other internally driven states, they have the crucial feature of being potentially wrong. Apes' mentalizing capacities with regard to both others and themselves were investigated. My main goal was to determine whether apes were capable of recognizing when others, or they themselves, were experiencing a false mental state. In the Competitive reaching study (Study 1), I tested both chimpanzees' and bonobos' ability to recognize that an experimenter had a false belief about the location of food that was moved when she was not looking. In the Change-of-contents study (Study 2), I further tested whether chimpanzees could recognize that an experimenter had a false belief about the contents of a box because she did not see the contents being switched. The apes failed these tests, despite my attempts to make them more species-relevant than earlier false belief tests. This is not to say, however, that apes do not have any capacity to infer mental states in others. As we saw in Chapter 3, earlier research shows that they are capable of recognizing and responding to others' perceptual and knowledge states; it is others' false belief states, in particular, that present problems for them.

In Studies 3 and 4, I tested chimpanzees' ability to make judgments about their own mental states. Very little work has been done in this area, so I first sought to establish chimpanzees' basic understanding of their own visual perceptual experiences in

the Perspective-shifting study (Study 3). Chimpanzees demonstrated good visual perspective taking skills with regard to themselves by immediately adopting the correct perspective for seeing into various types of containers. This study served as a bridge to the Lens study (Study 4), in which I investigated whether chimpanzees could recognize when they were experiencing a *false* visual perception. In Study 4, some individuals passed an appearance-reality test by consistently choosing the visually smaller (but truly bigger) of two food items, indicating that some chimpanzees are capable of recognizing when they themselves are experiencing false mental states, at least when those states are visual perceptions.

Given that the studies I presented here were all designed to test various mentalizing capacities in apes, it would be instructive to know how individuals performed across studies. If the tests I devised do indeed tap into related capacities that all stem from the same fundamental ability to recognize and respond to mental states, then we should find some consistency in performance across tasks. Of the 23 total apes I tested, 15 of them (all chimpanzees) took part in two or more of my studies, making some cross-test comparisons possible. Table 7.1 shows the mean proportion of correct trials in the various tests for these individuals. All individuals who received the false belief tests did very poorly in them, irrespective of how they performed in the other tests. This floor effect makes comparisons of false belief performance with the other tests largely uninformative. However, in the two tests designed to test chimpanzees' mentalizing capacities with regard to their own mental states, most individuals tended to perform

Table 7.1

Proportion of correct trials across tests for the 15 apes participating in at least two of the four studies. Blank cells indicate that the ape did not participate in that test.

Chimpanzee	False belief tests		Perspective-shifting test		Lens test	
	Competitive reaching	Change-of-contents	First look (Unseen)	First move (Unseen)	Basic (Seen)	Tracking (Unseen)
Trudy	0.42	0.35	0.83	0.78	0.79	0.67
Jahaga	0.00	0.35	1.00	0.89	0.96	1.00
Sandra	0.42	.	0.81	0.78	0.92	1.00
Frodo	0.17	.	0.72	0.67	1.00	0.75
Fifi	0.42	0.15	0.44	0.44	0.50	1.00
Dorien	0.25	.	0.28	0.39	0.04	.
Fraukje	0.58	.	0.28	0.28	0.08	.
Alex	0.25	0.20	0.50	0.33	.	.
Alexandra	0.00	.	0.39	0.28	.	.
Annett	0.58	0.30	0.94	0.89	.	.
Pia	0.36	.	.	.	1.00	0.92
Patrick	0.17	.	.	.	0.04	.
Corrie	0.42	.	.	.	0.25	.
Natascha	0.50	.	.	.	0.50	.
Ulla	0.17	.	.	.	0.38	.
<i>Chance proportion:</i>	0.50	0.50	0.33	0.33	0.50	0.50

either well in both or poorly in both. Further, as Figure 7.1 shows, for the seven chimpanzees who participated in both the Perspective-shifting test and the Lens test, performance was very highly correlated (Pearson's $r = 0.93$, $N = 7$, $p = 0.002$), suggesting that these two tests were indeed tapping into the same understanding of one's own mental states. Note that the age effect discussed in Chapter 6, whereby adolescent and young adult chimpanzees do better in tests of self-knowledge than older chimpanzees, is also apparent here.

The overall picture that emerges from the current research, together with past research, is that chimpanzees have a good appreciation of what others can or cannot see and also what others do or do not know based on past visual experiences. They

additionally recognize what they themselves can or cannot see from different visual perspectives, and they distinguish between their own states of knowledge and ignorance as determined by past visual experiences. False belief states, however, present more of a problem for them. All research to date suggests that apes are unable to respond adaptively to others' false belief states, and measures of spontaneous behavior provide only very weak evidence that they may recognize false beliefs on some level not evident in their

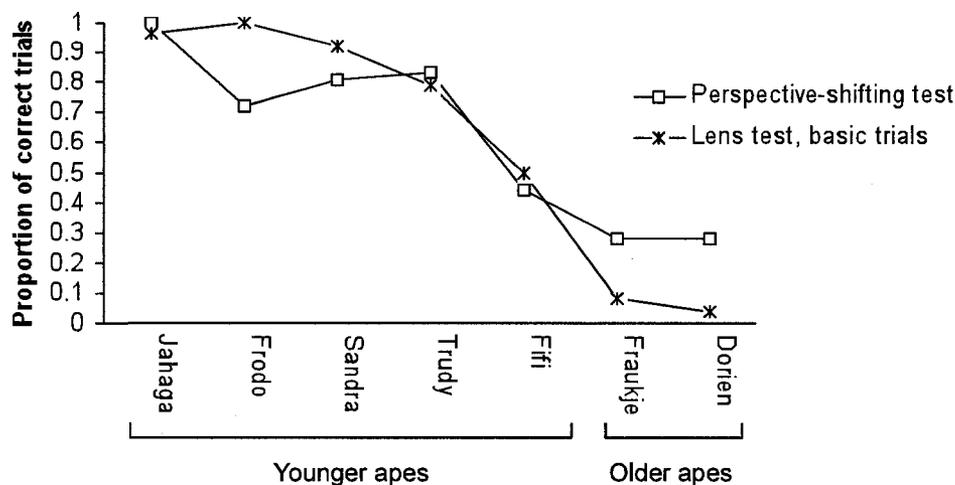


Figure 7.1. Performance of individual chimpanzees in the Perspective-shifting test of Study 3 (Unseen trials, first look measure) and the Lens test of Study 4 (Basic trials). The seven apes shown here were the only ones participating in both tests.

active responses. On the other hand, some chimpanzees do appear to be capable of recognizing when they *themselves* are experiencing a false mental state (namely, a mistaken visual perception). It is important to reiterate here that when I write about chimpanzees 'understanding' or 'recognizing' their own or others' mental states, I am not

claiming that this necessarily involves *conscious* reasoning or reflection on their part. It very well might, but there is currently no way to know this with certainty. We can say, however, that chimpanzees' responses in a variety of experimental tasks are consistent with their understanding at least some mental states on some level, including some false mental states. Further, their responses in control conditions make it improbable that their success in these tasks can be accounted for by low-level mechanisms such as learning to read behavioral or other cues.

Implications and Future Directions

At the start of this thesis I mentioned that research into ape mentalizing was still in its earliest stages. The four studies I have presented here help to address some of the outstanding questions, but there is still much work to do. One obvious question raised by my results is why some chimpanzees would be capable of passing appearance-reality tests but not false belief tests. Why would the same chimpanzee who can hold two conflicting representations of an object in mind simultaneously (e.g., a grape as apparently big but truly small) be incapable of simultaneously representing the real location of an object and someone else's mistaken belief about its location? The answer to this question may lie in the enormous difference in complexity between these two situations. The first involves variation along just one dimension: the real versus apparent size of the grape. In contrast, the second varies along a number of dimensions: my mental state versus someone else's mental state, the true location of the object versus the false representation of its location, the particular events in a sequence that I have witnessed versus the ones that someone else has witnessed, and so on. Thus, the sheer complexity

of false belief tests, at least the ones developed so far, may just be too cognitively taxing for chimpanzees.

It is also possible that some particular features of false belief tests could be more problematic for chimpanzees than other features. For example, as mentioned earlier, the kinds of mental states that chimpanzees are required to infer in false belief tests may be far more difficult for them than the kinds of mental states to be inferred in other tests, such as in visual perspective taking tests. Gopnik and Astington (1988) speculated that this is why children show good visual perspective taking abilities earlier in development than they begin to pass false belief tests. Of course, the particular kinds of false beliefs studied in false belief tests are also perceptually based, in the sense that they are formed by visually perceiving some events but not others. An important difference, however, is that the source of the false belief is not immediately perceptually apparent as participants respond in false belief tests, which could make these tests more difficult for chimpanzees than appearance-reality tests.

Another possibility is that recognizing others' mental states, of any kind, may be harder than recognizing one's own mental states. In the literature on mentalizing in humans, there is much debate about the relation between self-knowledge and knowledge of other minds. Some people think that an understanding of one's own mental states is more fundamental and develops earlier than an understanding of others' mental states; others think that they develop concurrently and draw upon the same underlying cognitive abilities; and still others argue that recognizing one's own mental states is harder and develops later than the ability to recognize others' (e.g., see Bogdan, 2005). A number of

researchers have attempted to resolve the debate by comparing children's relative success on tests requiring them to recognize others' false belief states versus tests requiring them to recognize their own prior false belief states. Findings have been mixed on which tests children find easier (see review in Müller et al., 2007), but a meta-analysis of a large number of studies suggests that there is no consistent difference between children's performance on self and other false belief tests (Wellman et al., 2001). Nevertheless, it could still be the case that chimpanzees find it easier to recognize their own mental states than to recognize others', even if this is not so for human children.

In short, the false belief and appearance-reality tests reported in this thesis differ in two important ways that could make the former more difficult for chimpanzees than the latter. Future research could help to determine the relative importance of each of these factors. Two lines of research in particular would be particularly useful: (1) 'self' versions of false belief tests, to determine if chimpanzees are capable of recognizing their own false belief states; and (2) 'other' versions of appearance-reality tests, to determine whether chimpanzees can recognize when others are experiencing false visual perceptions.

The first of these will be an extremely challenging problem to solve. Consider the difficulty, for example, of having nonverbal participants report on their own prior false beliefs about the contents of a container, as in the standard Smarties test (Perner et al., 1987). Such a test might proceed as follows: chimpanzees could be videotaped as they encountered a Smarties box and then as they looked inside the box to discover its unexpected contents (e.g., crayons). Two still images from the video might then be

presented to chimpanzees in varying order: (1) one showing the chimpanzee with the closed box just before it was opened, and (2) one showing the chimpanzee looking inside the open box. In each case, chimpanzees would be required to choose one of two photos depicting what they thought was in the box at the moment depicted in the video image, Smarties or crayons. Such a test seems easy enough in theory, but it would involve an enormous amount of training, and chimpanzees might never catch on to what was expected of them. Second, even if chimpanzees succeeded in such a task, it could be argued that they had simply learned to choose the Smarties photo when the video image depicted a closed box and to choose the crayons photo when the video image depicted open box, without really understanding why they were doing so. That is, success in this task could not provide clear evidence that chimpanzees were reporting on their own past false belief states. When one considers the ease with which Smarties tests for children are carried out and interpreted, the unique problems involved in studying mentalizing capacities in nonlinguistic creatures is very apparent.

A social version of the Lens test would be somewhat easier to implement than a self version of the false belief test. Such a test could involve, for example, pitting subordinate and dominant chimpanzees against one another in a contest for food, as in Hare et al.'s (2000, 2001) Competition paradigm described in Chapter 3. In this case, however, the food items would not be placed behind opaque barriers versus out in the open, but would instead be displayed behind magnifying versus minimizing lenses. If participants successfully predicted that a chimpanzee who was naïve to the lenses would choose the magnified grape over the minimized one, this would suggest that they

recognized the naïve chimpanzee's false perception, and consequent false belief, about the size of the grape. This in turn would suggest that judging others' false beliefs has been difficult for chimpanzees in previous false belief tasks because the source of the other's false belief has not been immediately perceptually apparent. Another advantage of a social version of the Lens task is that it would be less subject to explanation in terms of behavior reading than many other tests of mental-state attribution. Povinelli and Vonk (2003) suggested that the most convincing paradigms for investigating mentalizing in apes would be those requiring them to extrapolate from unique personal mental experiences to others' mental states. Thus, if some chimpanzees were to pass a social version of the Lens test, this might satisfy those who are skeptical of apes' success in any task in which they must infer others' mental states from observable physical or behavioral cues.

A social version of the Perspective-shifting test described in this thesis could also be instructive in determining the relation between chimpanzees' mentalizing abilities with regard to the self and others. While the Perspective-shifting test for self requires chimpanzees to recognize 'what I would be able to see from that particular perspective', a social version of the test would require them to recognize 'what he or she is able to see from that particular perspective'. Recall that in the self version of the Perspective-shifting test reported here (Study 3), chimpanzees knew from the earliest trials where they needed to position themselves to see into various types of containers. However, whether chimpanzees are also capable of recognizing what others can or cannot see from a variety of different perspectives is still unclear. With a social version of the

Perspective-shifting test we could ask, for example, do chimpanzees understand that a social partner who looks down onto the trapezoids from above cannot see the hidden food, but that one who moves opposite to look into them from the back can see it? If so, this would support the view that they have a general understanding of visual perspective that applies to cases involving both themselves and others. If not, it would suggest that their more advanced visual perspective taking abilities may be better with regard to the self than with regard to others.

More broadly, comparing chimpanzees' abilities on comparable social and nonsocial tasks could help to illuminate the relation between their mentalizing capacities with regard to the self and others. Are they two distinct kinds of abilities, or are they both grounded in a more general understanding of mental states as personal, internal phenomena that are largely determined by ones' perceptual experiences and can thus be incorrect? So far, evidence from this thesis and previous studies suggests that chimpanzees' mentalizing capacities with regard to the self match or exceed their capacities regarding others, but much more research is needed in this area before this can be concluded with any confidence. Especially important would be work aimed at charting the developmental progression of mentalizing capacities in chimpanzees, as it has been charted for children. Such investigations would allow us to determine whether skills for the self precede skills for others in chimpanzees or, alternatively, develop concurrently with them. Developmental studies have the potential to inform debates about the mechanism behind the capacities in question. For example, if chimpanzees demonstrate that they are capable of solving mentalizing tasks for the self earlier in

development than comparable tasks for others, it might support the Simulation theory view that recognizing others' mental states is accomplished by using one's own mental experience as a model.

Also greatly needed are direct investigations of chimpanzees' understanding of how knowledge states form, especially their understanding of how the different sensory modalities produce different types of knowledge (i.e., the modality-specific knowledge discussed in Chapter 3). Knowing how others come to know or believe what they do is fundamental to making correct mental-state attributions. That chimpanzees can use others' past visual exposure to a hiding event to infer what others know suggests that they understand the relation between seeing and knowing, but little is known about their understanding of how other kinds of perceptual experiences lead to knowledge. The knowledge-gathering paradigm used in my Perspective-shifting test (Study 3) could be adapted to test for modality-specific knowledge in chimpanzees, similar to the tests for children developed by O'Neill and colleagues (1992). For example, chimpanzees could be presented with two identical opaque, closed containers, one containing a desirable treat (e.g., food pellets) and the other containing something undesirable (e.g., sawdust). The containers would sound differently when shaken but could be made to smell alike. Would chimpanzees recognize that others who looked at or smelled the containers would not know what was inside, but that those who shook the containers would know? In another potential paradigm to test chimpanzees understanding of knowledge formation in themselves, they could first be familiarized with two food items, both of which looked identical. However, one item would be soft to the touch and taste bitter while the other

would be hard to the touch and taste sweet. If the items were then hidden inside of two identical containers, would chimpanzees recognize that they needed to feel inside the containers, rather than look inside of them, to determine the location of the desirable item?

Further studies to investigate whether chimpanzees recognize what they themselves do and do not know are also needed. Call and Carpenter (2001) made some initial progress on this question by finding that apes could discriminate between situations in which they knew versus did not know where food was hidden, and I replicated their findings with chimpanzees in Study 3. However, it is unclear whether one needs to reflect upon one's own mental states in this procedure or only act upon them (Carruthers, 2008). Further progress could be made using paradigms that require animals to effectively report upon their own states of knowledge or ignorance in an experimental task (Smith et al., 2003). The basic paradigm is as follows: animals are first trained to discriminate between random arrays of dots that are either sparsely packed or densely packed, and to press one button when they are presented with a sparse array and another button for a dense array (other stimuli, such as auditory tones of high or low frequency, may also be used). A correct response wins the animal a reward. In the test situation, they are occasionally presented with arrays that are difficult to judge as either sparse or dense because they fall somewhere in between. If they choose incorrectly a penalty ensues, such as a time-out period before the next trial. However, the animals also have the option of pressing a third key (the 'escape' or 'uncertainty' response) that results in no reward but advances them immediately to the next trial. Both monkeys and dolphins have

demonstrated that they are capable of employing the uncertainty response in this and similar procedures (Hampton, 2001; Kornell et al., 2007; Smith et al., 2003; Washburn et al., 2006), suggesting that they recognize when they do not know how to respond. If chimpanzees were also able to do so, it would provide more convincing evidence than Call and Carpenter's (2001) test that they are capable of recognizing their own states of knowledge and ignorance, just as they are capable of recognizing others' (although see Carruthers, 2008, for further skepticism).

Finally, while evidence continues to mount that false belief understanding is a uniquely human capacity, it may nevertheless still be too early to say this with absolute certainty. A decade ago researchers were ready to conclude that apes had no mentalizing capacities at all (Povinelli & Eddy, 1996b; Tomasello & Call, 1997). Since then, however, evidence to the contrary has been found in a steadily growing number of studies, many of which are reviewed in Chapter 3. Similarly, for children, the age of emergence of false belief understanding has traditionally been set at somewhere between 4 and 5 years old, and it still is by many researchers (see Wellman et al., 2001). However, innovative ways of asking much younger children what they understand about false beliefs has seriously challenged this view. Recall, for example, that 3-year-old children performed better in false belief tests when they were personally involved in carrying out the deceptive act (Chandler et al., 1989; Hala et al., 1991; Sullivan & Winner, 1993). A possible fruitful direction for future research could be to create similar false belief tests for chimpanzees and other apes, given that they have been observed to actively deceive others under natural conditions (Byrne & Whiten, 1988, 1990; Whiten &

Byrne, 1988). Another approach that has resulted in evidence of possible false belief understanding in infants just 1 to 2 years of age has been to use spontaneous looking rather than active choice measures (Onishi & Baillargeon, 2005; Southgate et al., 2007; Surian et al., 2007). The positive looking results obtained for apes in Study 1 (Version A of the procedure) provide the glimmer of a possibility that they too might understand more about false beliefs than their active responses suggest. Granted, those results were not replicated in Version B of the procedure or in Study 2. However, possible reasons for this were pointed out: the apes may have become tired or confused by previous testing (Study 1, Version B) or may have been experiencing high uncertainty across both the True and False belief conditions (Study 2). There may therefore be some value in pursuing this approach further.

Conclusion

This thesis included a number of firsts, among them the first appearance-reality test for apes, the first false belief test requiring apes to compete rather than cooperate with humans, and the first time that bonobos have been tested for false belief understanding. The results, as is most often the case, raise as many questions as they answer. My hope is that they will nevertheless be seen as an important contribution to the larger project of understanding the ape mind, and thereby also understanding more about the human mind. Together with previous research in this area, the current results suggest that the evolutionary story is not as simple as some have suggested. Povinelli and Giambrone (2001), for example, have argued that while both apes and humans are capable of reading and responding to others' behavior, the additional ability to

reinterpret that behavior in terms of its underlying mental states evolved in the human line after the evolutionary split from the chimpanzee line (called the “reinterpretation hypothesis”). However, there is now more than ample evidence to suggest that chimpanzees are capable of understanding at least some mental states. It is thus very likely that some mentalizing abilities were already present in the common ancestor of humans and chimpanzees. This does not mean, of course, that these abilities were not elaborated and built upon to a greater extent in the human line over the subsequent five million years than they were in the chimpanzee line. Clearly, adult humans have far more sophisticated and flexible mentalizing abilities than adult chimpanzees, perhaps aided in large part by our capacity for language. It does mean, however, that humans can no longer be seen as unique amongst all animals in the basic ability to even conceive of a mental realm of experience.

REFERENCES

- Abelev, M., & Markman, E. (2006). Young children's understanding of multiple object identity: Appearance, pretense, and function. *Developmental Science, 9*, 590-596.
- Allen, C., & Bekoff, M. (1997). *Species of mind: The philosophy and biology of cognitive ethology*. Cambridge, MA: The MIT Press.
- Andrews, G., Halford, G. S., Bunch, K. M., Bowden, D., & Jones, T. (2003). Theory of mind and relational complexity. *Child Development, 74*, 1476-1499.
- Astington, J. W., & Baird, J. A. (Eds.). (2005). *Why language matters for theory of mind*. New York: Oxford University Press.
- Avramides, A. (2001). *Other minds*. London: Routledge.
- Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: The MIT Press.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21*, 37-46.
- Baron-Cohen, S., Tager-Flusberg, H., & Cohen, D. J. (Eds.). (2000). *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed.). Oxford: Oxford University Press.
- Bartsch, K., & Wellman, H. M. (1989). Young children's attribution of action to beliefs and desires. *Child Development, 60*, 946-964.
- Bethell, E. J. (2003). Rank effects on chimpanzee social monitoring. *Primate Eye, 80*, 4.
- Bethell, E. J., Vick, S. J., & Bard, K. A. (2007). Measurement of eye gaze in chimpanzees (*Pan troglodytes*). *American Journal of Primatology, 69*, 562-575.

- Bialystok, E., & Senman, L. (2004). Executive processes in appearance-reality tasks: The role of inhibition of attention and symbolic representation. *Child Development, 75*, 562-579.
- Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*, B25-B31.
- Bogdan, R. J. (2005). Why self-ascriptions are difficult and develop late. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 90-206). New York: The Guilford Press.
- Boysen, S. T., & Berntson, G. G. (1995). Responses to quantity: Perceptual versus cognitive mechanisms in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes, 21*, 82-86.
- Boysen, S. T., Berntson, G. G., Hannan, M. B., & Cacioppo, J. T. (1996). Quantity-based interference and symbolic representations in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes, 22*, 76-86.
- Boysen, S. T., Berntson, G. G., & Mukobi, K. L. (2001). Size matters: Impact of item size and quantity on array choice by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 115*, 106-110.
- Braine, M. D. S., & Shanks, B. L. (1965a). The conservation of a shape property and a proposal about the origin of the conservations. *Canadian Journal of Psychology, 19*, 197-207.
- Braine, M. D. S., & Shanks, B. L. (1965b). The development of conservation of size. *Journal of Verbal Learning and Verbal Behavior, 4*, 227-242.

- Bräuer, J., Call, J., & Tomasello, M. (2005). All great ape species follow gaze to distant locations and around barriers. *Journal of Comparative Psychology, 119*, 145-154.
- Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition, 10*, 439-448.
- Brentano, F. (1995 [1874]). *Psychology from an empirical standpoint*. London: Routledge.
- Brook, A., & Stainton, R. J. (2000). *Knowledge and mind*. Cambridge, MA: The MIT Press.
- Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: How infants interpret adult looking behavior. *Developmental Psychology, 38*, 958-966.
- Buttelmann, D., Call, J., & Tomasello, M. (2007, March). *Great apes use emotional expressions to infer desires*. Paper presented at the Biennial Meeting of the Society for Research on Child Development, Boston, MA.
- Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2007). Enculturated chimpanzees imitate rationally. *Developmental Science, 10*, F31-F38.
- Butterworth, G. (1991). The ontogeny and phylogeny of joint visual attention. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading* (pp. 223-232). Oxford: Blackwell.
- Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy. *British Journal of Developmental Psychology, 9*, 55-72.

- Byrne, R. W. (1995). *The thinking ape: Evolutionary origins of intelligence*. New York: Oxford University Press.
- Byrne, R. W., & Whiten, A. (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans*. Oxford: Clarendon Press.
- Byrne, R. W., & Whiten, A. (1990). Tactical deception in primates: The 1990 database. *Primate Report*, 27, 1-101.
- Call, J. (2004). Inferences about the location of food in the great apes (*Pan paniscus*, *Pan troglodytes*, *Gorilla gorilla*, and *Pongo pygmaeus*). *Journal of Comparative Psychology*, 118, 232-241.
- Call, J. (2005). The self and other: A missing link in comparative social cognition. In H. S. Terrace & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (pp. 321-341). Oxford: Oxford University Press.
- Call, J., Agnetta, B., & Tomasello, M. (2000). Cues that chimpanzees do and do not use to find hidden objects. *Animal Cognition*, 3, 23-34.
- Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition*, 3, 207-220.
- Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). 'Unwilling' versus 'unable': Chimpanzees' understanding of human intentional action. *Developmental Science*, 7, 488-498.
- Call, J., Hare, B., & Tomasello, M. (1998). Chimpanzee gaze following in an object-choice task. *Animal Cognition*, 1, 89-99.

- Call, J., & Rochat, P. (1996). Liquid conservation in orangutans (*Pongo pygmaeus*) and humans (*Homo sapiens*): Individual differences and perceptual strategies. *Journal of Comparative Psychology, 110*, 219-232.
- Call, J., & Rochat, P. (1997). Perceptual strategies in the estimation of physical quantities by orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology, 111*, 315-329.
- Call, J., & Tomasello, M. (1996). The effect of humans on the cognitive development of apes. In A. E. Russon & K. A. Bard (Eds.), *Reaching into thought: The minds of the great apes* (pp. 371-403). Cambridge: Cambridge University Press.
- Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development, 70*, 381-395.
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72*, 1032-1053.
- Carlson, S. M., Wong, A., Lemke, M., & Cosser, C. (2005). Gesture as a window on children's beginning understanding of false belief. *Child Development, 76*, 73-86.
- Caron, A. J., Butler, S., & Brooks, R. (2002). Gaze following at 12 and 14 months: Do the eyes matter? *British Journal of Developmental Psychology, 20*, 225-240.
- Carpenter, M., Call, J., & Tomasello, M. (2002). A new false belief test for 36-month-olds. *British Journal of Developmental Psychology, 20*, 393-420.
- Carruthers, P. (1996). Simulation and self-knowledge: A defence of theory-theory. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 22-38). Cambridge: Cambridge University Press.

- Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. *Mind and Language*, 23, 58-89.
- Carruthers, P., & Smith, P. K. (Eds.). (1996). *Theories of theories of mind*. Cambridge: Cambridge University Press.
- Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two, three-, and four-year-olds' early theories of mind. *Child Development*, 60, 1263-1277.
- Chapuis, N. (1987). Detour and shortcut abilities in several species of mammals. In P. Ellen & C. Thinus-Blanc (Eds.), *Cognitive processes and spatial orientation in animal and man* (pp. 97-106). Dordrecht: Nijhoff.
- Charman, T., & Campbell, A. (1997). Reliability of theory of mind task performance by individuals with a learning disability: A research note. *Journal of Child Psychology & Psychiatry*, 38, 725-730.
- Churchland, P. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78, 67-90.
- Churchland, P. (1988). *Matter and consciousness* (2nd ed.). Cambridge, MA: The MIT Press.
- Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, 9, 377-395.
- D'Entremont, B., Hains, S. M. J., & Muir, D. W. (1997). A demonstration of gaze following in 3- to 6-month-olds. *Infant Behavior & Development*, 20, 569-572.

- Davies, M., & Stone, T. (Eds.). (1995a). *Folk psychology: The theory of mind debate*. Oxford: Blackwell.
- Davies, M., & Stone, T. (Eds.). (1995b). *Mental simulation: Evaluations and applications*. Oxford: Blackwell.
- de Waal, F., & Lanting, F. (1997). *Bonobo: The forgotten ape*. University of California Press.
- de Waal, F. B. M., & Pokorny, J. J. (2008). Faces and behinds: Chimpanzee sex perception. *Advanced Science Letters, 1*, 99-103.
- Deák, G. O. (2006). Do children really confuse appearance and reality? *Trends in Cognitive Sciences, 10*, 546-550.
- Dennett, D. (1996). *Kinds of minds*. New York: Basic Books.
- Dunham, P., Dunham, F., & O'Keefe, C. (2000). Two-year-olds' sensitivity to a parent's knowledge state: Mind reading or contextual cues? *British Journal of Developmental Psychology, 18*, 519-532.
- Dunphy-Lelii, S., & Wellman, H. (2004). Infants' understanding of occlusion of others' line of sight: Implications for an emerging theory of mind. *European Journal of Developmental Psychology, 1*, 49-66.
- Elkind, D. (1966). Conservation across illusory transformations in young children. *Acta Psychologica, 25*, 389-400.
- Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews, 24*, 581-604.

- Flavell, J. H. (1963). *The developmental psychology of Jean Piaget*. Princeton, NJ: Van Nostrand.
- Flavell, J. H. (1977). The development of knowledge about visual perception. In H. E. Howe Jr. & C. B. Keasey (Eds.), *The Nebraska Symposium on Motivation* (Vol. 25, pp. 43-76). Lincoln: University of Nebraska Press.
- Flavell, J. H. (1986). The development of children's knowledge about the appearance-reality distinction. *American Psychologist*, *41*, 418-425.
- Flavell, J. H. (1992). Perspectives on perspective taking. In H. Berlin & P. Pufall (Eds.), *Piaget's theory: Prospects and possibilities*.
- Flavell, J. H. (1993). The development of children's knowledge of false belief and the appearance-reality distinction. *International Journal of Psychology*, *28*, 595-604.
- Flavell, J. H. (2000). Development of children's knowledge about the mental world. *International Journal of Behavioral Development*, *24*, 15-23.
- Flavell, J. H. (2004). Theory-of-mind development: Retrospect and prospect. *Merrill-Palmer Quarterly*, *50*, 274-290.
- Flavell, J. H., Flavell, E. R., & Green, F. L. (1983). Development of the appearance-reality distinction. *Cognitive Psychology*, *15*, 95-120.
- Flavell, J. H., Flavell, E. R., & Green, F. L. (1987). Young children's knowledge about the apparent-real and the pretend-real distinctions. *Developmental Psychology*, *23*, 816-822.

- Flavell, J. H., Green, F. L., & Flavell, E. R. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development, 51*, i-69.
- Flavell, J. H., Green, F. L., Wahl, K. E., & Flavell, E. R. (1987). The effects of question clarification and memory aids on young children's performance on appearance-reality tasks. *Cognitive Development, 2*, 127-144.
- Flavell, J. H., Shipstead, S. G., & Croft, K. (1978). Young children's knowledge about visual perception: Hiding objects from others. *Child Development, 49*, 1208-1211.
- Flombaum, J. I., & Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology, 15*, 447-452.
- Foote, A. L., & Crystal, J. D. (2007). Metacognition in the rat. *Current Biology, 17*, 551-555.
- Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483-527.
- Garnham, W., & Ruffman, T. (2001). Doesn't see, doesn't know: Is anticipatory looking really related to understanding of belief? *Developmental Science, 4*, 94-100.
- Gergely, G., Bekkering, H., & Kiraly, I. (2002). Rational imitation in preverbal infants. *Nature, 415*, 755-755.
- Gettier, E. (1963). Is justified true belief knowledge? *Analysis, 23*, 121-123.
- Gómez, J. C. (1996). Non-human primate theories of (non-human 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.
- Gómez, J. C. (2004). *Apes, monkeys, children, and the growth of the mind*. Cambridge, MA: Harvard University Press.
- Gómez, J. C. (2005). Species comparative studies and cognitive development. *Trends in Cognitive Sciences*, 9, 118-125.
- Gómez, J. C., & Teixidor, P. (1992, August). *Theory of mind in an orangutan: A nonverbal test of false belief appreciation?* Paper presented at the XIV Congress of the International Primatological Society, Strasbourg, France.
- Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 59, 26-37.
- Gopnik, A., Slaughter, V., & Meltzoff, A. N. (1994). Changing your views: How understanding visual perception can lead to a new theory of mind. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 157-181). Hove: Lawrence Erlbaum.
- Gordon, R. M. (1996). 'Radical' simulationism. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 11-21). Cambridge: Cambridge University Press.
- Hala, S., Chandler, M., & Fritz, A. S. (1991). Fledgling theories of mind: Deception as a marker of three-year-olds' understanding of false belief. *Child Development*, 62, 83-97.

- Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 5359-5362.
- Hansen, M. B., & Markman, E. M. (2005). Appearance questions can be misleading: A discourse-based account of the appearance-reality problem. *Cognitive Psychology*, *50*, 233-263.
- Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., et al. (1996). 'Theory of mind' in the brain: Evidence from a PET scan study of Asperger syndrome. *Neuroreport*, *8*, 197-201.
- 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.
- Hare, B., Call, J., & Tomasello, M. (2006). Chimpanzees deceive a human competitor by hiding. *Cognition*, *101*, 495-514.
- Hare, B., Melis, A. P., Woods, V., Hastings, S., & Wrangham, R. (2007). Tolerance allows bonobos to outperform chimpanzees in a cooperative task. *Current Biology*, *17*, 619-623.
- Hare, B., & Tomasello, M. (2004). Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour*, *68*, 571-581.

- Held, S., Mendl, M., Devereux, C., & Byrne, R. W. (2001). Studies in social cognition: From primates to pigs. *Animal Welfare, 10*, S209-S217.
- 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.
- Hirata, S., & Matsuzawa, T. (2001). Tactics to obtain a hidden food item in chimpanzee pairs (*Pan troglodytes*). *Animal Cognition, 4*, 285-295.
- Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development, 57*, 567-582.
- Holmes, H. A., Black, C., & Miller, S. A. (1996). A cross-task comparison of false belief understanding in a head start population. *Journal of Experimental Child Psychology, 63*, 263-285.
- Horner, V., & Whiten, A. (2007). Learning from others' mistakes? Limits on understanding a trap-tube task by young chimpanzees (*Pan troglodytes*) and children (*Homo sapiens*). *Journal of Comparative Psychology, 121*, 12-21.
- Hughes, C., Adlam, A., Happé, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good test-retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *Journal of Child Psychology and Psychiatry, 41*, 483-490.

- Humphrey, N. K. (1976). The social function of intellect. In P. P. G. Bateson & R. A. Hinde (Eds.), *Growing points in ethology* (pp. 303-317). Cambridge: Cambridge University Press.
- Itakura, S. (1996). An exploratory study of gaze-monitoring in nonhuman primates. *Japanese Psychological Research*, 38, 174-180.
- Itakura, S., & Tanaka, M. (1998). Use of experimenter-given cues during object-choice tasks by chimpanzees (*Pan troglodytes*), an orangutan (*Pongo pygmaeus*), and human infants (*Homo sapiens*). *Journal of Comparative Psychology*, 112, 119-126.
- Johnson, C. M., & Karin-D'Arcy, M. R. (2006). Social attention in nonhuman primates: A behavioral review. *Aquatic Mammals*, 32, 423-442.
- Jolly, A. (1966). Lemur social behavior and primate intelligence. *Science*, 153, 501-506.
- Kafka, F. (1971[1917]). In *Franz Kafka: The complete stories* (W. & E. Muir, T. & J. Stern, Trans.). New York: Schocken Books.
- 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*, 15, 21-54.

- Keenan, T., & Ellis, B. J. (2003). Children's performance on a false-belief task is impaired by activation of an evolutionarily-canalized response system. *Journal of Experimental Child Psychology, 85*, 236-256.
- Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye. *Journal of Human Evolution, 40*, 419-435.
- Koós, O., Gergely, G., Csibra, G., & Biro, S. (1997, April). *Why eating smarties makes you smart? Understanding false belief at the age of three*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Washington, DC.
- Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint-seeking in monkeys. *Psychological Science, 18*, 64-71.
- Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (in preparation). A new nonverbal change-of-contents false belief test for children and chimpanzees.
- Kuhlmeier, V. A., & Boysen, S. T. (2001). The effect of response contingencies on scale model task performance by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 115*, 300-306.
- Kuhlmeier, V. A., Boysen, S. T., & Mukobi, K. L. (1999). Scale-model comprehension by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 113*, 396-402.
- Langer, J., & Strauss, S. (1972). Appearance, reality, and identity. *Cognition, 1*, 105-128.
- LeDoux, J. (1996). *The emotional brain*. New York: Simon & Schuster.

- Leslie, A. M. (1994). Pretending and believing: Issues in the theory of ToMM. *Cognition*, *50*, 211-238.
- Lewis, C., & Osborne, A. (1990). Three-year-olds' problems with false belief: Conceptual deficit or linguistic artifact? *Child Development*, *61*, 1514-1519.
- Liebal, K., Call, J., & Tomasello, M. (2004). Use of gesture sequences in chimpanzees (*Pan troglodytes*). *American Journal of Primatology*, *64*, 377-396.
- Lillard, A. S. (1993). Young children's conceptualization of pretense: Action or mental representational state? *Child Development*, *64*, 372-386.
- Lohmann, H., Carpenter, M., & Call, J. (2005). Guessing versus choosing—and seeing versus believing—in false belief tasks. *British Journal of Developmental Psychology*, *23*, 1-20.
- Martin-Ordas, G., Call, J., Haun, D., & Colmenares, F. (under review). Great ape memories for past events show a human-like developmental pattern.
- Matsui, T., & Miura, Y. (2008). Pro-social motive promotes early understanding of false belief [Electronic Version]. *Nature Precedings* from <http://precedings.nature.com/documents/1695/version/1>.
- Matsuzawa, T. (1985). Use of numbers by a chimpanzee. *Nature*, *315*, 57-59.
- Mayes, L. C., Klin, A., Tercyak, K. P. J., Cicchetti, D. V., & Cohen, D. J. (1996). Test-retest reliability for false-belief tasks. *Journal of Child Psychology and Psychiatry*, *37*, 313-319.
- McGuigan, M., & Doherty, M. J. (2002). The relation between hiding skill and judgment of eye direction in preschool children. *Developmental Psychology*, *38*, 418-427.

- Melis, A. P., Call, J., & Tomasello, M. (2006). Chimpanzees (*Pan troglodytes*) conceal visual and auditory information from others. *Journal of Comparative Psychology*, *120*, 154-162.
- Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, *31*, 838-850.
- Mendes, N., Rakoczy, H., & Call, J. (2008). Ape metaphysics: Object individuation without language. *Cognition*, *106*, 730-749.
- Menzel, E. W. (1974). A group of young chimpanzees in a one-acre field. In A. Schrier & F. Stollnitz (Eds.), *Behavior of nonhuman primates* (Vol. 5, pp. 83-153). New York: Academic Press.
- Moll, H., & Tomasello, M. (2004). 12- and 18-month-old infants follow gaze to spaces behind barriers. *Developmental Science*, *7*, F1-F9.
- Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, *24*, 603-613.
- Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development*, *61*, 722-730.
- Moses, L. J. (2005). Executive functioning and children's theory of mind. In B. F. Malle & S. D. Hodges (Eds.), *Other minds: How humans bridge the divide between self and others* (pp. 11-25). New York: The Guilford Press.
- Müller, U., Miller, M. R., Michalczyk, K., & Karapinka, A. (2007). False belief understanding: The influence of person, grammatical mood, counterfactual

- reasoning and working memory. *British Journal of Developmental Psychology*, 25, 615-632.
- Muncer, S. (1983). "Conversations" with a chimpanzee. *Developmental Psychobiology*, 16, 1-11.
- Murray, F. B. (1968). Phenomenal-real discrimination and the conservation of illusion-distorted length. *Canadian Journal of Psychology*, 22, 114-121.
- Naito, M. (2003). The relationship between theory of mind and episodic memory: Evidence for the development of auto-noetic consciousness. *Journal of Experimental Child Psychology*, 85, 312-336.
- O'Connell, S., & Dunbar, R. I. M. (2003). A test for comprehension of false belief in chimpanzees. *Evolution and Cognition*, 9, 131-140.
- O'Neill, D. K. (1996). Two-year-old children's sensitivity to a parent's knowledge state when making requests. *Child Development*, 67, 659-677.
- O'Neill, D. K., & Chong, S. C. F. (2001). Preschool children's difficulty understanding the types of information obtained through the five senses. *Child Development*, 72, 803-815.
- O'Neill, D. K., Astington, J. W., & Flavell, J. H. (1992). Young children's understanding of the role that sensory experiences play in knowledge acquisition. *Child Development*, 63, 474-490.
- Okamoto, S., Tomonaga, M., Ishii, K., Kawai, N., Tanaka, M., & Matsuzawa, T. (2002). An infant chimpanzee (*Pan troglodytes*) follows human gaze. *Animal Cognition*, 5, 107-114.

- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255-258.
- Panskepp, J. (1998). *Affective neuroscience*. New York: Oxford University Press.
- 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 of London Series B Biological Sciences*, *362*, 731-744.
- Perner, J., Leekam, S., & Wimmer, H. (1987). Three year olds' difficulty with false belief. *British Journal of Developmental Psychology*, *5*, 125-137.
- Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, *308*, 214-216.
- Pika, S., Liebal, K., Call, J., & Tomasello, M. (2005). The gestural communication of apes. *Gesture*, *5*, 39-54.
- Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology*, *47*, 116-129.
- Povinelli, D. J. (1994a). A theory of mind is in the head, not the heart. *Behavioral and Brain Sciences*, *17*, 573-574.
- Povinelli, D. J. (1994b). What chimpanzees (might) know about the mind. In R. W. Wrangham, W. C. McGrew, F. B. M. de Waal & P. Heltne (Eds.), *Chimpanzee cultures* (pp. 285-300). Cambridge, MA: Harvard University Press.
- Povinelli, D. J. (2003). Folk physics for apes: The chimpanzee's theory of how the world works. *Human Development*, *46*, 161-168.

- Povinelli, D. J., Bering, J. M., & Giambrone, S. (2000). Toward a science of other minds: Escaping the argument by analogy. *Cognitive Science, 24*, 509-541.
- Povinelli, D. J., Bierschwale, D. T., & Cech, C. G. (1999). Comprehension of seeing as a referential act in young children, but not juvenile chimpanzees. *British Journal of Developmental Psychology, 17*, 37-60.
- Povinelli, D. J., & DeBlois, S. (1992). Young children's (*Homo sapiens*) understanding of knowledge formation in themselves and others. *Journal of Comparative Psychology, 106*, 228-238.
- Povinelli, D. J., & Eddy, T. J. (1996a). Chimpanzees: Joint visual attention. *Psychological Science, 7*, 129-135.
- Povinelli, D. J., & Eddy, T. J. (1996b). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development, 61*, 1-152.
- Povinelli, D. J., & Eddy, T. J. (1997). Specificity of gaze-following in young chimpanzees. *British Journal of Developmental Psychology, 15*, 213-222.
- Povinelli, D. J., & Giambrone, S. (1999). Inferring other minds: Failure of the argument by analogy. *Philosophical Topics, 27*, 167-201.
- Povinelli, D. J., & Giambrone, S. (2001). Reasoning about beliefs: A human specialization? *Child Development, 72*, 691-695.
- Povinelli, D. J., Nelson, K. E., & Boysen, S. T. (1990). Inferences about guessing and knowing by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 104*, 203-210.

- Povinelli, D. J., Reaux, J. E., Bierschwale, D. T., Allain, A. D., & Simon, B. B. (1997). Exploitation of pointing as a referential gesture in young children, but not adolescent chimpanzees. *Cognitive Development, 12*, 327-365.
- Povinelli, D. J., Rulf, A., & Bierschwale, D. (1994). Absence of knowledge attribution and self-recognition in young chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 108*, 74-80.
- Povinelli, D. J., Rulf, A., Landau, K., & Bierschwale, D. (1993). Self-recognition in chimpanzees (*Pan troglodytes*): Distribution, ontogeny, and patterns of emergence. *Journal of Comparative Psychology, 107*, 347-372.
- 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.
- Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development, 61*, 973-982.
- Premack, D., & Dasser, V. (1991). Perceptual origins and conceptual evidence for theory of mind in apes and children. In A. Whiten (Ed.), *Natural theories of mind: Evolution, development and simulation of everyday mindreading*. Cambridge, MA: Blackwell.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4*, 515-526.

- Preuss, T. M. (2000). What's human about the human brain? In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed.) (pp. 1219-1234). Cambridge, MA: The MIT Press.
- Preuss, T. M. (2001). The discovery of cerebral diversity: An unwelcome scientific revolution. In D. Falk & K. R. Gibson (Eds.), *Evolutionary anatomy of primate cerebral cortex* (pp. 138-164). Cambridge: Cambridge University Press.
- Ravenscroft, I. (2005). *Philosophy of mind*. Oxford: Oxford University Press.
- Reaux, J. E., Theall, L. A., & Povinelli, D. J. (1999). A longitudinal investigation of chimpanzees' understanding of visual perception. *Child Development, 70*, 275-290.
- Regolin, L., Vallortigara, G., & Zanforlin, M. (1995). Object and spatial representations in detour problems by chicks. *Animal Behaviour, 49*, 195-199.
- Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: Evidence from 14- and 18-month-olds. *Developmental Psychology, 33*, 12-21.
- Repacholi, B. M., & Slaughter, V. (Eds.). (2003). *Individual differences in theory of mind: Implications for typical and atypical development*. New York: Psychology Press.
- Rice, C., Koinis, D., Sullivan, K., Tager-Flusberg, H., & Winner, E. (1997). When 3-year-olds pass the appearance-reality test. *Developmental Psychology, 33*, 54-61.
- Ruffman, T., Garnham, W., Import, A., & Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of Experimental Child Psychology, 80*, 201-224.

- Sapp, F., Lee, A., & Muir, D. (2000). Three-year-olds' difficulty with the appearance-reality distinction: Is it real or is it apparent? *Developmental Psychology*, *36*, 547-560.
- Schulkin, J. (2000). *Roots of social sensibility and neural function*. Cambridge, MA: The MIT Press.
- Shettleworth, S. J., & Sutton, J. E. (2006). Do animals know what they know? In S. Hurley & M. Nudds (Eds.), *Rational animals?* (pp. 235-246). Oxford: Oxford University Press.
- Shillito, D., Shumaker, R., Gallup, G., & Beck, B. (2005). Understanding visual barriers: Evidence for Level 1 perspective taking in an orang-utan, *Pongo pygmaeus*. *Animal Behaviour*, *69*, 679-687.
- Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: Review of literature and implications for future research. *Neuroscience and Biobehavioral Reviews*, *30*, 855-863.
- Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor, R., & Erb, L. (1995). The uncertain response in the bottlenosed dolphin (*Tursiops truncatus*). *Journal of Experimental Psychology: General*, *124*, 391-408.
- Smith, J. D., Shields, W. E., & Washburn, D. A. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and Brain Sciences*, *26*, 317-373.

- Sodian, B., Taylor, C., Harris, P. L., & Perner, J. (1991). Early deception and the child's theory of mind: False trains and genuine markers. *Child Development, 62*, 468-483.
- Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 18*, 587-592.
- Suda, C., & Call, J. (2004). Piagetian liquid conservation in the great apes (*Pan paniscus*, *Pan troglodytes*, and *Pongo pygmaeus*). *Journal of Comparative Psychology, 118*, 265-279.
- Suda, C., & Call, J. (2006). What does an intermediate success rate mean? An analysis of a Piagetian liquid conservation task in the great apes. *Cognition, 99*, 53-71.
- Suddendorf, T., & Whiten, A. (2001). Mental evolution and development: Evidence for secondary representation in children, great apes, and other animals. *Psychological Bulletin, 127*, 629-650.
- Sullivan, K., & Winner, E. (1993). Three-year-olds' understanding of mental states: The influence of trickery. *Journal of Experimental Child Psychology, 56*, 135-148.
- Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science, 18*, 580-586.
- Taylor, M. (1988). Conceptual perspective taking: Children's ability to distinguish what they know from what they see. *Child Development, 59*, 703-718.
- Taylor, M., & Flavell, J. H. (1984). Seeing and believing: Children's understanding of the distinction between appearance and reality. *Child Development, 55*, 1710-1720.

- Taylor, M., & Hort, B. (1990). Can children be trained in making the distinction between appearance and reality? *Cognitive Development*, 5, 89-99.
- Terrace, H. S. (2005). Descartes, Darwin, and beyond. In H. S. Terrace & J. Metcalfe (Eds.), *The missing link: Origins of self-knowing consciousness* (pp. ii-vi). New York: Oxford University Press.
- Tomasello, M., & Call, J. (1997). *Primate Cognition*. New York: Oxford University Press.
- 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., & Gluckman, A. (1997). Comprehension of novel communicative signs by apes and human children. *Child Development*, 68, 1067-1080.
- Tomasello, M., Call, J., & Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. *Animal Behaviour*, 55, 1063-1069.
- Tomasello, M., Call, J., & Hare, B. (2003a). 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. (2003b). Chimpanzees versus humans: It's not that simple. *Trends in Cognitive Sciences*, 7, 239-240.

- Tomasello, M., Call, J., Warren, J., Frost, T., Carpenter, M., & Nagell, K. (1997). The ontogeny of chimpanzee gestural signals: A comparison across groups and generations. *Evolution of Communication*, 1, 223-253.
- Tomasello, M., Hare, B., & Agnetta, B. (1999). Chimpanzees, *Pan troglodytes*, follow gaze direction geometrically. *Animal Behaviour*, 58, 769-777.
- Tomasello, M., Hare, B., & Fogleman, T. (2001). The ontogeny of gaze following in chimpanzees, *Pan troglodytes*, and rhesus macaques, *Macaca mulatta*. *Animal Behaviour*, 61, 335-343.
- Uher, J., & Call, J. (2008). How the great apes (*Pan troglodytes*, *Pongo pygmaeus*, *Pan paniscus*, *Gorilla gorilla*) perform on the reversed reward contingency task II: Transfer to new quantities, long-term retention, and the impact of quantity ratios. *Journal of Comparative Psychology*, 122, 204-212.
- Vick, S. J., & Anderson, J. R. (2003). Use of human visual attention cues by olive baboons (*Papio anubis*) in a competitive task. *Journal of Comparative Psychology*, 117, 209-216.
- Vlamings, P. H. J. M., Uher, J., & Call, J. (2006). How the great apes (*Pan troglodytes*, *Pongo pygmaeus*, *Pan paniscus*, and *Gorilla gorilla*) perform on the reversed contingency task: The effects of food quantity and food visibility. *Journal of Experimental Psychology: Animal Behavior Processes*, 32, 60-70.
- Washburn, D. A., Smith, J. D., & Shields, W. E. (2006). Rhesus monkeys (*Macaca mulatta*) immediately generalize the uncertain response. *Journal of Experimental Psychology: Animal Behavior Processes*, 32, 185-189.

- Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 167-187). Malden, MA: Blackwell.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655-684.
- Whiten, A., & Byrne, R. W. (1988). Tactical deception in primates. *Behavioral and Brain Sciences, 11*, 233-273.
- Whiten, A., & Suddendorf, T. (2007). Great ape cognition and the evolutionary roots of human imagination. *Proceedings of the British Academy, 147*, 31-59.
- Wildman, D. E., Uddin, M., Liu, G., Grossman, L. I., & Goodman, M. (2003). Implications of natural selection in shaping 99.4% nonsynonymous DNA identity between humans and chimpanzees: Enlarging genus *Homo*. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 7181-7188.
- Wimmer, H., Hogrefe, G., & Perner, J. (1988). Children's understanding of informational access as a source of knowledge. *Child Development, 59*, 386-396.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128.
- Woodruff, G., Premack, D., & Kennel, K. (1978). Conservation of liquid and solid quantity by the chimpanzee. *Science, 202*, 991-994.

- Yunger, J. L., & Bjorklund, D. F. (2004). An assessment of generalization of imitation in two enculturated orangutans (*Pongo pygmaeus*). *Journal of Comparative Psychology, 118*, 242-246.
- Ziv, M., & Frye, D. (2003). The relation between desire and false belief in children's theory of mind: No satisfaction? *Developmental Psychology, 39*, 859-876.
- Zucca, P., Antonelli, F., & Vallortigara, G. (2005). Detour behaviour in three species of birds: Quails (*Coturnix* sp.), herring gulls (*Larus cachinnans*) and canaries (*Serinus canaria*). *Animal Cognition, 8*, 122-128.

APPENDIX – ARTICLES BASED ON THIS THESIS

Krachun, C., & Call, J. (2008). Chimpanzees (*Pan troglodytes*) know what can be seen from where. *Animal Cognition*. Published online at

<http://dx.doi.org/10.1007/s10071-008-0192-x>. (print version to follow)

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (in press). A competitive nonverbal false belief task for children and apes. *Developmental Science*.

(accepted May 7, 2008)

Krachun, C., Call, J., & Tomasello, M. (under review). Can chimpanzees (*Pan troglodytes*) distinguish appearance from reality? *Cognition*. (reference number:

COGNIT-D-08-00420)

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (in preparation). A new nonverbal change-of-contents false belief test for children and chimpanzees.