Pictorial Primates: A Search for Iconic Abilities in Great Apes

Persson, Tomas

2008

Citation for published version (APA):
Persson, T. (2008). *Pictorial Primates: A Search for Iconic Abilities in Great Apes*. [Doctoral Thesis (monograph), Cognitive Science]. Cognitive Science.

Total number of authors:
1

Pictorial Primates: A Search for Iconic Abilities in Great Apes

PICTORIAL PRIMATES

A Search for Iconic Abilities in Great Apes

TOMAS PERSSON

1666

LUND UNIVERSITY

Lund University Cognitive Studies 136

Tomas Persson (2008)
Pictorial Primates: A Search for Iconic Abilities in Great Apes.
Lund University Cognitive Studies, 136

ISBN 91-974741-9-3
ISSN 1101-8453
ISRN LUHFDA/HFKO-1019-SE

Printed at Media-Tryck, Sociologen, 2008

Till mina föräldrar, och deras…

Contents

Preface

Pictures and other iconic media are used extensively in psychological experiments on nonhuman primate perception and conceptualisation. They are also used in interaction with primates in their everyday lives, and as pure entertainment. For most humans, interpreting pictures is an act of imagination as much as an act of face-value recognition. Pictorial competence is thus an intriguing area for investigating minds. But in what ways do nonhuman primates understand iconic artefacts? What implications do these different ways have for the conclusions we draw from those studies of perception and conceptualisation? What can pictures tell us about primate cognition, and what can primates tell us about pictures?

The present work has been conducted within the project "Language, Gestures and Pictures in the Perspective of Semiotic Development" (Språk, gester och bilder i ett semiotiskt utvecklingsperspektiv: SGB). The project included linguists, semioticians, and cognitive scientists at Lund University.

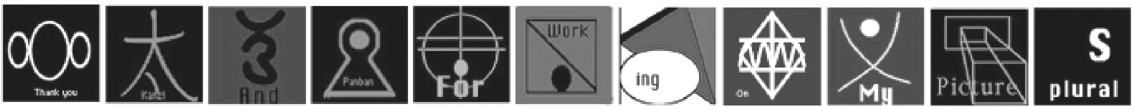
In SGB, "development" has meant to imply both ontogenetic and phylogenetic change. Therefore I took the opportunity to continue my nonhuman primate research interests within this project. Primates are prime models for the study of both evolutionary change, and change in the growing individual. When it was time to settle for a research topic my point of departure was that I wanted to conduct empirical work with great apes in a zoo setting. Language seemed not a feasible area for investigation at the time, but nonhuman gestures were still somewhat understudied. However, I saw an even bigger lack of research effort in the area of pictures. I had just finished an observational study and now wanted to experience an experimental situation instead. Pictures thus seemed the perfect choice.

Since picture understanding in animals, from a semiotic viewpoint, is indeed little explored, this thesis has taken an explorative form. Charting the land was a necessary step. Since time is always limited I simultaneously embarked on an empirical journey, perhaps prematurely from a strictly scientific standpoint, but not a moment too soon from an educational one. My priority has been the study of apes, and the study of studies of apes.

This is, to my knowledge, the first thesis in Sweden exclusively dedicated to nonhuman primate cognition. I have therefore taken the opportunity to introduce the field to hopefully a new readership. Occasional digresses from pictorial matters are therefore intentional.

Given the novelty of this line of research also at Lund University, written material has been the main source for this thesis. It would not have been possible without the Lund University library services. Acknowledging these is therefore in order.

Another major source for my development as a researcher is a group of gorillas at Givskud Zoo in Denmark. Unfortunately I do not know how to thank them, other than storing our moments in fond memory, but I can thank them symbolically through thanking a second family of great apes. The bonobos I have in mind reside at the Great Ape Trust of Iowa (GATI), USA, and have likewise made an impact on me. They might be able to read:



Some people in positions of power have been of importance for the realisation of this thesis. Director of Givskud Zoo, Richard Østerballe, my supervisor Prof. Peter Gärdenfors, Dr. Sue Savage-Rumbaugh, and the members of the Science Circle at GATI, all deserve my gratitude for their decisions to let me work with them and animals in their care. Fruitful discussions are included in this thank you.

In terms of direct help with testing I am especially grateful to Flemming Poder and the rest of the gorilla caretakers at Givskud Zoo. At GATI I am indebted to Takashi Yoshida, Jon Thompson, Ben Thompson, Elizabeth Rubert-Pugh, William Fields, and especially Dr. Sue Savage-Rumbaugh for involvement in testing. Indirect help was given by the rest of the GATI bonobo, orangutan, and administrative staffs. Thank you all. Ing-Marie Persson and Mathias Osvath were very helpful at Furuvik Zoo.

Other influential forces have been members of the project SGB and participants of the SGB and the Div. of Cognitive Science at Lund University (LUCS) seminars. Colleagues at LUCS and the Department of Philosophy at Lund University have naturally been supportive at all levels of work. I am most grateful. Although I decided on a collective thank you, an explicit one is due to Dr. Petra Björne who directed my attention to a particularly crucial piece of information for the understanding of my own work.

Funding has been received from the SGB project, Stiftelsen Fil. Dr. Uno Otterteds fond, Elisabeth Rausings Minnesfond, Landshövding Per Westlings Minnesfond, and Crafoordska stiftelsen. I thank them for their generous support and interest in my work. Without it I would have been stuck at the library.

Lastly, a heartfelt thank you is due to family and friends outside of work. For many years my parents have never discouraged me from my gorilla dreams. This thesis is therefore dedicated to them.

I

I made no further tests, as I consider it quite obvious that results are determined simply by the technical accuracy of the photographs and the difference of the objects they represent. Anyone who may take the trouble to experiment on other chimpanzees in the same way, will be able to demonstrate effectively and exactly, by means of larger and clearer reproductions, that the animals recognize and differentiate between such photographs. As a further variation – to meet possible objections – I would suggest, in the crucial experiments, the use of pictures of another food – say the very popular oranges or thistles – if bananas were used in the preliminary tests.

Wolfgang Köhler (1925/1957, p. 278)

Chapter 1

Introduction

The quote overleaf is the closing paragraph of the appendix to the second, revised, edition of Wolfgang Köhler's classic The Mentalities of Apes, published in 1925.[1]

Köhler was a German Gestalt psychologist who conducted extensive experiments and observations on chimpanzees (*Pan troglodytes*) at the Anthropoid Station in Tenerife in the 1910s. The studies focused on spatial problem-solving, perception and tool-use. Köhler's work stood in opposition to especially Edward Thorndike's associative psychology and claimed that chimpanzees were capable of solving tasks by insight, which at the time was believed to be a hallmark of human intelligence.[2] Together with the American Robert M. Yerkes, Köhler is generally considered to be the first to thoroughly study primate behaviour in order to draw conclusions about nonhuman great ape (henceforth ape) intelligence, and its relation to human thinking (Tomasello & Call, 1997).

But simultaneously in Moscow, Russia, from 1913 to 1916, Nadezhda Ladygina-Kohts took detailed notes on an infant chimpanzee that she raised in her own home (Ladygina-Kohts, 1935/2002). She addressed many questions still studied in comparative psychology today, and among other notable things developed the match-to-sample testing paradigm (Yerkes & Petrunkevitch, 1925). Matching-to-sample is an experimental setup where a subject is required to choose among an array of choice items the one that matches a sample item on a predefined dimension such as colour. Matching-to-sample (MTS) will play a substantial role in this thesis. Both in the literature review in Part II, and when it comes to my own empirical work in Part III.

Years after the original observations, Ladygina-Kohts made comparative psychology truly comparative by making detailed comparisons between her chimpanzee data, and data gathered on her own son. However, her comparison was not published until 1935, two years after the publication of Kellogg and Kellogg's (1933/1967) similar comparison between Gua, an infant chimpanzee, and their son Donald (*Homo sapiens*). Kellogg and Kellogg's work in turn was made in cooperation with Yerkes.

---

[1] The first German edition of *The Mentalities* appeared in 1917.

[2] Although a seductive account, Köhler's interpretations have been disputed in replications of his own experiments (see Chance, 1959; Harlow, 1951).

1.1 Pioneers

Yerkes was a biologically inclined psychologist with broad interests. He first studied the evolution and development of perception, learning and instinct in animals, but later turned to higher mental faculties and their behavioural correlates. He had for a long time, since his days in graduate school, nourished an interest in nonhuman primates (henceforth "primates") as mirrors and tools for studying humanity (Yerkes, 1943/1945).

Yerkes initially made plans to study primates at the Anthropoid Station in Tenerife, but hindered by the First World War he had to redirect his attention to America. Over the course of a few months in 1915 he thus investigated the problem-solving abilities of an orangutan (*Pongo sp.*) and some rhesus macaques (*Macaca mulatta*). This work resulted in the publication of another classic: The Mental Life of Monkeys and Apes: A Study of Ideational Behavior (1916b).

Yerkes' conclusions regarding the minds of great apes were comparable to, but independent from, Köhler's. They both concluded that ape minds contain something more than "mere" stimulus-response associations (Yerkes, 1916b; Köhler, 1925/1957). This most general conclusion lent itself to an infinite number of further studies. There were a lot to be discovered, a pressing interest that Yerkes also recognised in colleagues. But what were lacking in America were the proper facilities and animals to serve the cause. A centre for the study of everything primate was imminently needed, and Yerkes was set on creating one (Yerkes, 1916a).

In the 1920s Yerkes acquired a chimpanzee and a bonobo (*Pan paniscus*)[3] which he studied at his own home farm, but both animals died at a young age (Yerkes, 1926). He was granted the funds by Yale University to start a pilot laboratory with four new chimpanzees. During this time he also conducted extensive cognitive experiments with a circus gorilla (*Gorilla beringei*) (Yerkes, 1927a; 1927b; 1929). The running of the pilot laboratory impressed Yerkes' funders, and with the completion of new breeding and experimental facilities the Yale Laboratories of Primate Biology could be opened in 1930. This was well over a decade after Yerkes' public announcement of his intentions (i.e. Yerkes, 1916a).

Yerkes believed in broad scientific approaches and therefore his centre was divided into areas suitable for observations, as well as areas for psychological and medical laboratory work. Although an experimentalist, Yerkes also viewed naturalistic studies in the wild to be of utmost importance. He therefore commissioned two pioneering expeditions to Africa to study chimpanzees (Nissen, 1931) and mountain gorillas (Bingham, 1932) respectively. Field studies were among other things "an opportunity to check and correct the interpretation of experiments and the conclusions based upon them" (Yerkes, 1943/1945, p. 296). This attitude stands somewhat in contrast to modern primatology where a commonly held view is that questions should ultimately be settled in controlled experiments. What is observed in the wild is instead to be brought into the laboratory. Yerkes' position instead hints at a strong adaptationist view of cognition where no trait makes sense outside of its proper ecology.

---

[3] A few years before the official "discovery" of the species (Savage-Rumbaugh & Lewin, 1994).

Through donations of further chimpanzees to join the four from the pilot laboratory the Yerkes factory was soon in motion, eventually producing a wealth of data, methods and expertise. In 1941, when Yerkes retired, the Yale Laboratories of Primate Biology were renamed the Yerkes Laboratories of Primate Biology in order to honour Robert M. Yerkes' vast influence on the field of primate studies. Today this collection of facilities and colonies of animals are named Yerkes National Primate Research Center.

But Köhler, Ladygina-Kohts and Yerkes with colleagues were not really the first of their kind, neither as experimentalists nor field workers. Seldom mentioned in modern literature is Richard Lynch Garner (e.g. 1892; 1896), who was really the first to study the mentality of monkeys and apes in a systematic way against a Darwinian backdrop. Garner started out by studying monkey communication in America and developed with time the playback experiment using Edison's phonograph (Garner, 1892; Radick, 2005). Besides studying "monkey speech" he made experiments to investigate, among other things, colour and sound preferences, quantity judgements, and reactions to mirrors (Garner, 1892). Later, he was the first researcher to study chimpanzees and gorillas (*Gorilla gorilla*) in a field situation in Africa (Garner, 1896). Although heavily criticised for his unclear boundaries between observation and hearsay when it comes to wild animals (Candland, 1993), his descriptions of some chimpanzee and gorilla infants in his care are believable and informative. While some of Garner's interpretations are indeed difficult to take seriously, others were decades before their time. Both types no doubt lent to the critique he received. Among other things he took a strong position against anthropocentrism. "It is not a safe and infallible guide to measure all things by the standard of man's opinion of himself. It is quite true that, by such a unit of measure, the comparison is much in favor of man, but the conclusion is neither just nor adequate" (Garner, 1896, p. 61). This attitude he based on the fact that different ecologies and adaptive histories make direct comparison between human and ape mentalities difficult, if not meaningless.

With the apes in his care Garner performed several pioneering experiments. For example he tried to teach them to speak, but they could only learn to produce a couple of words. The same finding was made by Furness about a decade later (1916, partially reproduced in Yerkes & Yerkes, 1929/1953). The last attempt along similar lines was going to be by Keith Hayes and Catherine Hayes in the late 1940s (Hayes, 1951). After this, gestured sign language (Gardner et al., 1989; Miles, 1990; Patterson & Linden, 1981; Terrace, 1980), token chips (Premack, 1976), and printed symbols (Rumbaugh, 1977; Savage-Rumbaugh, 1986) proved to be the more successful modalities.[4] Since the use of pictures is a common ingredient in ape language training we will get to know most of the apes studied in these projects in this thesis. There is also reason to suspect that this training feeds back into pictorial competence itself in fundamental ways.

The issue of pictures did not pass Garner by. "I kept a cup for a monkey to drink milk from, on the sides of which were some brilliant flowers and green leaves, and she would frequently quit drinking the milk to play with the flowers on the cup, and

---

[4] Garner had given token communication a try as well (Mitchell, 1999).

seemed never able to understand why she could not get hold of them" (Garner, 1892, pp. 25-26). But he also noted e.g. apes' common lack of attention to pictures (Garner, 1896). Both these kinds of responses will be central in the chapters to follow.

After Garner, but still before Köhler, Ladygina-Kohts and Yerkes, was Lightner Witmer, founder of clinical psychology. In 1909 he published an investigation of the capacities of a stage chimpanzee, Peter (Witmer, 1909, in Candland, 1993). Initially his expectations were low, but after subjecting Peter to various standard tests at his children's clinic Witmer was duly impressed by the chimpanzee (Candland, 1993). Of special interest for the present text is the experiment that also impressed Witmer the most. After observing a model the chimpanzee was able to perfectly replicate the writing of the letter 'W' on a blackboard. We will return to this episode in Chapter 10. Witmer concluded that the study of Peter's mind "is a subject fit, not for the animal psychologist, but for the child psychologist" (Hornaday, 1922, p. 90). Indeed, in today's primatology the line between developmental and comparative psychology is rightfully fuzzy. The ape mind and the child mind have been compared extensively since the pioneering works of Ladygina-Kohts (1935/2002) and Kellogg and Kellogg (1933/1967), and is one of the central themes in today's comparative psychology.

1.2 Rediscovery of cognition

During the middle part of the last century the study of ape thinking gave way to the study of behaviour in behaviourist and ethological frameworks (Tomasello & Call, 1997). In behavioural psychology one was looking for general principles of learning. The choice of test subjects was therefore ruled by convenience. Rats, pigeons and monkeys were used instead of the less manageable great apes, which was probably just as well for the latter. A notable exception was at Yerkes' laboratories where behaviourist regimes were implemented, apparently reluctantly, for a brief but productive time in parallel to the more traditional work (Dewsbury, 2003).

In the first decades of discrimination and matching tasks with primates, abstract rather than depicting stimuli were used. This is understandable given the control over single visual properties this allows. Naturally, interest in the animals' *conceptualisation* of depictions was virtually non-existent. The exception seems to have been Hayes and Hayes (1953), to be discussed in Chapter 5.

It was not until the "cognitive revolution" in psychology that the study of the primate mind, as opposed to behaviour, really got fashionable again in the western scientific world. It surfaced, after a slow start, in the 1970s with names such as Emil Menzel, David and Ann Premack, Allen and Beatrix Gardner, and Duane Rumbaugh among others (Tomasello & Call, 1997). However, drawing a sharp line between behaviourist and cognitivist approaches is not possible, especially not in experimental primatology. Both fields have made substantial contributions to contemporary studies of primate mentality.

As an example, Harlow's (1949) discovery of the learning set phenomenon in macaques was paramount for a later shift to cognitive approaches (Rumbaugh, 1997). Learning set formation can be described as “learning how to learn a kind of problem” (Harlow, 1949, p. 53). In a simple discrimination problem one of several choice items is rewarded, and the subject eventually learns which object it should chose in order to obtain food. Many primates[5], and non-primates (see Tomasello & Call, 1997), get better at solving discrimination problems the more problems they are subjected to, even when they do not get enough trials per individual problem to learn by trial and error. It seems as if they retain some form of hypothesis from earlier problems that they test on subsequent ones. They can thereby discover the “rules” of the task. With time the feedback from a single trial can be enough to inform the subject about the correct response on following ones. Learning set has been formed.

The “rediscovery” of cognition was also coupled with a renewed interest in an evolutionary perspective and of contrasting ape species. But many novel areas for cognitive comparison, such as deception, imitation, planning, use of pointing gestures, and linguistic comprehension, really predated the 1920s (Mitchell, 1999). Many issues were studied already in the post-Darwin 1800s by for example Garner above.

Primate cognition is truly back in business and is at present an ever growing field, adding new findings on a monthly basis. Numerous studies involve pictorial stimuli, but almost exclusively as a means to measure something else. Picture understanding itself has still not been thoroughly studied (Bovet & Vauclair, 2000; Cabe, 1980; Fagot et al., 2000). Both the indirect and the direct lines of study will be reviewed in Part II, which is the bulk of this thesis. Part I, in turn, explores the phenomena of pictorial competences, and Part III constitutes my own empirical work in the area.

1.3 Wolfgang Köhler’s picture experiments

Let us now return to the quote that introduced Part I. At Tenerife in the 1910s Köhler chose to look at chimpanzees’ performance with pictures after having observed their reactions to stuffed toy animals, cardboard face-masks and mirrors. He had noted that the chimpanzees became emotionally affected by stuffed toy animals that were placed in their enclosure, i.e. fearful. He also observed that it was necessary for the toy animals to have some likeness to real animals (“nearness to life”) in order to invoke such fear. Not any object would do. Confusingly, the stuffed animals invoked even stronger fear responses than did most real animals. He concluded that the stuffed animals, not being fully real, played on the imagination in a way that real animals did not, just as fear of ghosts or the play of shadows on a wall can be stronger than fear of real things. Uncertainty, as opposed to experience, seems to be key in both cases. Köhler also succeeded in frightening his subjects by wearing a demon’s face mask. Although he never tested, he imagined that a plain piece of cardboard in front of the face would not have the same effect.

---

[5] Humans included. Harlow (1949) found comparable effects in monkeys and children 2 to 5 years old.

With mirrors, the chimpanzees were interested from the very start in gazing at its contents: presumably the other chimpanzee inside, or behind it. Köhler describes how one of his subjects persists in trying to grab and hit the chimpanzee behind the mirror, and even throw surprise attacks at it. Such observations are common in the mirror literature (see section 9.4). Soon enough, however, the chimpanzees started to grimace and play with their own reflections rather than act towards them. Their play later extended to any shiny object they could get their hands on. Köhler reports that the chimpanzees never tired of playing with reflections, and in quite sophisticated ways. They used reflecting objects to look behind their own backs and they even used puddles of urine to look at things outside of their sleeping-room window. The reflective surfaces had turned into being *about* the world rather than being an actual instance of it.

Köhler was wondering if the chimpanzees' ability to recognise nearness to life in stuffed animals and mirrors would remain if the third dimension and colours were removed. He turned to black-and-white photographs. In the initial tests the chimpanzees intently studied the depictions of themselves and other chimpanzees, but only one of them showed suggestive signs that he recognised their content. He had extended his arm and chimpanzee-greeted a picture. The chimpanzee stopped his gesturing when Köhler turned the photograph around and showed its backside, but resumed when faced with the motif again. A second subject, a female, after having investigated her photographic self-portrait, tucked it into her groin and walked away with it. Yerkes (1943/1945) suggests that this could have been a case of expecting the little picture-chimpanzee to cling to her like a baby.

Köhler next developed a photograph of an empty crate and another photograph of a crate crammed full of bananas and pasted these on two boxes, both baited with fruit (see *fig. 1, left*). The star pupil of the previous test chose the box with the banana picture on 10 successive trials. However, being rewarded for any choice he soon started to choose randomly. When Köhler removed the baiting from the non-target box and only rewarded the subject for the correct choice (differential reinforcement) the chimpanzee's performance returned to about 90%.

By now Köhler wanted to control for rote learning and developed two new photographs: one of bananas and one of a stone (see *fig. 1, right*). The subject performed better with the new pictures than with the old ones. Köhler ascribed this to the superior quality and nearness to life in the second pair. The chimpanzee performed extra poorly on those trials where the old pictures returned after a series of trials with the new ones, which Köhler believed was a result of relaxation in attention after an easy bout with the new pictures.

When testing a second chimpanzee on the old pictures with differential reinforcement Köhler could not establish a permanent good performance. A third subject tested with the old pictures performed much better, but lost the ability when exposed to the slightest distraction. When confronted with the second pair of photographs, the stone and the bananas, she got focused and eager and made hardly any mistakes. These mixed results Köhler ascribed solely to the quality of the pictures. "I made no further tests […]" he reports (Köhler, 1925/1957, p. 278).

In the account of Köhler's brief investigation of picture perception in chimpanzees we can note several things that we will come back to: The (social) actions of one of the chimpanzees towards the photographs triggered by their life-like properties (although they in fact were in black-and-white), the different performance across individuals, the effect of food rewards, and perhaps most importantly, how little about picture understanding experiments of the type above can really tell us. The reason for this is that the tests assume that the perceived correspondence between picture and the real world can only be of one kind, and that it can only be present or absent. The possibility for different kinds of correspondences is neglected, and consequently alternative explanations are not controlled for. I will argue in this thesis that the issue is not merely a question of the "size and clarity of the reproductions," to paraphrase Köhler. However, to Köhler's defence it must be said that he was interested in chimpanzee perception of reality, and had he been more interested in how they perceive pictures as such he might have made further tests after all.

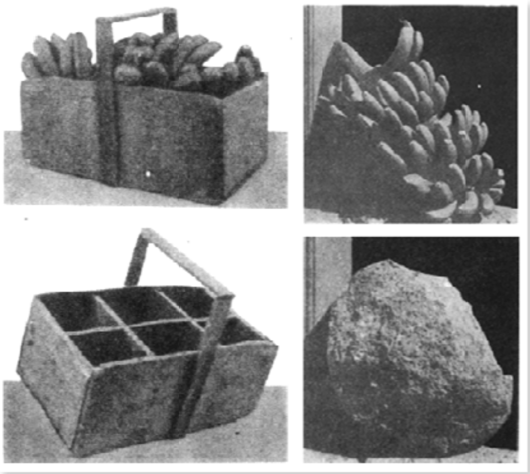


Figure 1. Wolfgang Köhler's picture stimuli: Crates, bananas and a rock. From Köhler (1922).

Köhler's conclusion was that the more like its referent a picture is, the easier it is for a chimpanzee to solve an object-choice task guided by picture information. This is an adequate observation. But why do we not see the same correlation in adult human pictorial competence? A much distorted photograph is not necessarily more difficult to understand than one that depicts reality more truthfully. Furthermore, we readily decode non-photographs that are quite far removed from the real perceptual world, like the cartoons in today's newspaper. We can even intend to see likeness where there is not supposed to be any, as when we look for faces in the clouds. Are the Köhler and human examples different points on a single competence continuum, with differences merely reflecting different experiences of picture qualities? I will argue in this thesis that it is not.

Köhler probably thought that his experiments after all said something about the understanding of pictures, but I would say, contrary to Köhler, that "anyone who may take the trouble to experiment on other chimpanzees" has a lot left to investigate. For example, why does an average human girl of 9 months try to drink from a depicted milk bottle when one year later she will instead point to and speak about the picture, and seldom confuse it with real objects to such a degree that she takes physical action towards it? Why can she at the age of 2.5 still not name a pencil drawing of a tiger when a gorilla in National Geographic Magazine is signing about cartoon kittens in a picture book? Why can a gorilla at Wolfgang Köhler Primate Research Centre at Leipzig Zoo use the motifs in photographs to guide her actions in a selection between two containers, while another gorilla, at Givskud Zoo, cannot? He is solely guided by their spatial placement. In a laboratory at Kyoto University a macaque can recognise its own species in photographs, and at Yerkes National Primate Research Center a chimpanzee can match pictures of chimpanzees with pictures of their mothers. Pigeons can discriminate depth-rotated table lamps in simple line-drawings. What are the processes behind these behaviours? Are they different? Are they related? The first step in answering these questions is to define different ways in which pictures can gain meaning for a viewer.

1.4 Three ways of looking at pictures

Many people have indeed studied primate cognition with the aid of pictures, which I have hinted at above and will review in Part II, but only a few have directly studied the understanding of pictures, and more importantly, the understanding of pictures as pictures. To make this distinction more clear I will now present the three forms of pictorial competence that is the basis for my further analysis and can account for primate (human and nonhuman) behaviours with pictures.

1. The first way in which pictures can get their meaning, and thus basis for acting upon, bypasses any estimation of what the picture might actually *depict*. What are perceived are rather the patterns, shapes, and colours, on the surface of the picture, and it stays at that. This form of picture understanding will here be called a ***surface*** type of picture processing. Besides perceiving local elements, seeing motifs in the sense of global forms is in theory possible, but they have only a learned connection to the real world, if any. Through association, i.e. rote learning, of specific picture - object relations, or generalisation based on invariant features, one can thus judge correspondences while circumventing *recognition*.[6] That is, one can sort pictures of e.g. apples on a level that does not involves realising that it is in any way apples that one sorts.

2. Pictures can also get their meaning from likeness to the real world, without being sufficiently differentiated from this, leading to the perception of pictures as part of reality and not *about* reality. Although the photographic image is perhaps the typical example, it is not necessarily limited to stimuli that seem realistic from a human perspective. Critical features in otherwise abstract depictions can likewise elicit a reality guided response. This second type can thus be called a ***reality*** based picture processing. An object is not seen as being anywhere else but in the picture, albeit perhaps in a stranger form than usual. With this type of understanding one can solve tasks that depend on categorisation, but it is not really different from categorising real instances of the depicted objects. If it is e.g. matched in an experiment with a similar object outside of the picture, it is at best an object relating to an object, not a picture of an object relating to an object.

---

[6] The term recognition is here used for categorisation on the level of objects, scenes etc. Responding to local features, such as colour, is only recognition to the extent that this is indeed how objects are categorised also in the real world.

3. A third way for pictures to be rendered meaningful is through likeness to scenes in the real world (and other pictures), but sufficiently differentiated from these not to be confused with them. In this manner pictures can be *about* the world. Reference thus lays in the specification of *what* the picture is similar to and *why* it is not itself this thing. Such a *stand-for* relation implies two types of expectations that can be said to have different relative weight depending on the type of picture that one encounters, although they are both necessary. In pictures where the likeness hits one directly in virtue of mirroring one's real-life experiences, an *expectation of separation* between picture and reality is crucial. But when it comes to pictures that require more of an interpretive stance an *expectation of likeness* enhances actual likeness. Many pictures would not be perceived at all without such expectations. The third type of picture use is distinguished from other referential instances by being called seeing pictures as ***pictures***.

It should be noted that the above distinctions pertain to a subject's approach to pictorial stimuli, and does not depend on the type of stimuli per se. For example, stimuli suitable for processing in a surface mode, such as abstract shapes in a discrimination task, are also possible to approach in reality or pictorial modes. Although such shapes are seldom found outside of laboratory stimulus sets, a red circle can still be discriminated as a "red circle," which can be a bona fide category, if nothing else within the experimental context. One can thus not exclude a specific mode of processing on the grounds of type of picture alone, but one can determine which mode suffices for adequate performance in a given task. A type of setup that would require seeing pictures as pictures and at the same time preclude processing in surface or reality modes would for example entail novel pictures (to counteract associative learning), that are impossible to confuse with reality, but still require a categorical response.

While one can solve tasks that only require a surface or reality mode competence with a pictorial ability, the vice versa does not apply. The reason for this is that as a pictorially competent individual one is both able to perceive similarities and marks on surfaces, but as a surface mode processor one is not necessarily able to perceive similarities between static versions and their real-life counterparts. The same overlap in competence can be seen from reality mode to surface mode. Being able to perceive motifs in pictures entails being able to perceive marks on surfaces.

For the above reasons the modes can be seen as hierarchical in relation to each other, but I would not go as far as to equal them with a developmental trajectory. The modes are not *general* competences but depend on interaction with specific pictures in specific contexts. Switching between modes is not best described as reverting to a previous stage in development, but to a different way of approaching a certain visual display. In this thesis we will find several examples of the modes competing with each other for one and the same picture.

My three-part division is very similar to that of Fagot et al. (2000) who propose the three modes "independence", "confusion," and "equivalence" modes of picture processing, in a review of bird and nonhuman primate picture experiments. I was not aware of their paper when I formulated my own distinctions and it thus seems to be a case of parallel reasoning. This is not that surprising since "[t]he proposed classification into three modes of processing is intuitively obvious" in the words of Fagot et al. (2000, p. 297). Indeed, Premack (1976) makes a similar distinction between forms of picture competences in animals, as do Cabe (1980) indirectly. Deregowski (1989) does the same for humans. The distinction also follows quite naturally from the picture semiotics of Sonesson (e.g. 1989) (see Chapter 4).

Fagot et al.'s (2000) notion of "modes" is very useful and I have borrowed that term for my purposes here. Mode alludes to *a way of approaching* specific pictures under specific circumstances rather than the possession of a boxed understanding about pictures in general. It is possible to switch between modes, not only as one develops a general picture concept, but for specific pictures and in short time spans. One can start out with the approach that what one is viewing is real, but suddenly discover that it is not, which instantly allows different actions, attitudes, emotions, perceptions etc. towards what one is viewing. Or one can learn as a child that a strange painting at home is of a horse, and even learn where in the picture the horse is supposed to graze. Then one day, at the age of 35, one suddenly sees the horse! It popped out after one isolated the head, after which the relation to the tail became apparent, and everything in-between fell in place. After that one has difficulties going back to not seeing the horse.

However, I will not subscribe to the terminology of "independence," "confusion," and "equivalence." Those terms refer to the relationship between the depiction and its referent, but in the independence and confusion modes there are by definition no referents. The relationship is only in the mind of the human observer who intends a referent. In this sense there is an important difference between the modes of Fagot et al. and mine since mine take the perspective of the subject, including what the subject can actually do with pictures in respective mode. For example, since I propose that it is indeed possible to make connections between pictures and objects in the *surface mode*, "independence" becomes a misleading term. Likewise, for the *reality mode* "confusion" is a misleading term because there is not necessarily any confusion from the perspective of the subject. (It would be to say that someone can perceive a real-life apple because it confuses it with other apples.) The word confusion in this case derives from an observable effect that the reality mode can have, such as grasping or tasting pictures, judged as confusion by someone who knows that the picture is not its referent. Furthermore, not all instances of picture processing in this mode give rise to confusion behaviours. There are cases where pictures and real-world objects are not confused although the picture is still seen as *a kind* of reality, but quite different from everyday reality, and also quite different from what would be a truly pictorial "reality." Sometimes reality mode results in confusion with reality, and sometimes it is differentiated from reality in the sense that the same behaviours do not apply both to real entities and their pictures. But pictures are still not seen as being *about* the former, and recognition is limited to pictures that share with reality enough of those properties that the perceiver is used to base everyday recognition on. Typically such pictures can be described as realistic pictures, or highly iconic ones.[7]

Fagot et al.'s (2000) behavioural definition of performance in the third mode, equivalence mode, is that an animal shall continue to exhibit actions pertinent to the depicted object even in situations where confusion is impossible, and they give the example of line drawings. I agree with this definition in the realm of behaviour. I also agree with their cognitive definition when they speak about an animal viewing the picture as a representation, and being aware of its difference from real objects. Similarly to them I want to stress that there is a differentiation between picture and referent from the point of view of the subject in this mode, but contrary to Fagot and colleagues, I do not think that "equivalence" is a suitable term for this purpose. First of all, regardless of how the term equivalence is used in learning theory, the word as such denotes the very opposite of what is implied by *differentiation*. A picture and its referent are not interchangeable. An equivalence relation is said to be, among other things, symmetric[8] (e.g. Sidman & Tailby, 1982), but the relation between pictures and objects is not symmetric. This seems in fact to be true for all similarity judgements, where one of two entities always takes on the identity of reference point (Rosch, 1975) usually in virtue of being the most familiar, or salient, of the two entities (Tversky, 1977). Only that which is less salient, or prominent, can usually stand for that which is more. Objects do not stand for pictures because the real world is more prominent than are pictures (e.g. Sonesson, 1989). If an actual object is to refer to a depiction it must selectively present features, among all those present in its rich real-world version, that pertain to the referred depiction and nothing else. This cannot be accomplished without extensive contextual scaffolding. The picture has to accomplish the same in regards to the object, but the picture has been prepared for this very purpose. Through a process of selective production or choice, features relevant for displaying a referent have been highlighted. This is a great advantage. In virtue of being more selective, or "simpler," than the real world, the picture more easily narrows downs reference.

Yes, viewed as behaviour in a test situation, matching objects to pictures equally well as pictures to objects, is symmetrical performance. However, matching of this type is on the level of matching an interpretation of a picture to a real-world object. Such interpretation can take two forms that both yields a symmetrical matching performance, but only one of them entails a symmetrical *view* of the picture – object relation. This occurs in reality mode, and not in pictorial/equivalence mode. In this mode *expression*, the actual markings on a surface, and *content*, the interpreted appearance of those markings, are undifferentiated by virtue of being seen as direct reality. Matching based on an almost complete identity can indeed be seen as being symmetrical.

In fact, Cabe (1980) also defines object – picture equivalence in a way more ascribable to reality mode. To make recognition pictorial, however, he adds that

---

[7] Iconicity is here used differently from the use in e.g. computer science where an icon is an abstraction, and a high degree of iconicity can be taken to mean a high degree of abstraction. Here a high level of iconicity rather means that the picture shares *more* properties with its referent than a picture low in iconicity does.

[8] The other two requirements for an equivalence relation are reflexivity and transitivity.

equivalence is not enough. There also needs to be "object – picture discriminability," i.e. differentiation. This is an aspect, he says, that "is rarely assessed" (Cabe, 1980, p. 335). He also points out the role of iconicity, in that recognition must always be spontaneous, and never learned.

Cabe's additions are necessary, because when approached as pictures, pictures and objects do not covary in the manner posed by the notion of "equivalence." The relationship between expression and content is not one of identity, as it is in reality mode, but one of similarity and differentiation. This results in an externalisation of the referent, i.e. the picture is *about* something else, and this new relation between the depiction and the depicted is not a symmetrical one. As argued above, when pictures are seen as pictures, they stand for their objects better than their objects can stand for them.

To truly include a conceptualisation of the picture as such in the definitions of the modes, and not merely base it on behavioural performance, the third mode will therefore here be called the *pictorial mode*. This does not mean that a picture – object relation cannot be one of stimulus equivalence, in terms of performance, only that this possibility should not define the third mode.

	Surface mode	**Reality mode**	**Pictorial mode**
Referent	None / Learned	None / Learned	Iconically perceived
Recognition	None / Invariant non-depicting properties	Categorical identity / isolation	Categorical similarity
Differentiation	*Not applicable*	None / Weak	Strong

Table 1. Three modes of picture processing. (*Categorical isolation* means that the content is perceived as being real but is not recognised as a category from the world at large.)

When applying the definitions above (*table 1*) on Köhler's picture experiments (we leave the mirror ones aside for now) it is clear that what he tested, and found, in his subjects was a competence based on reality mode. The greeting gestures and the performance with pictures judged high versus low in realism points to just that. The beauty of the pictorial mode, on the other hand, is that likeness to reality is sometimes secondary to the appreciation that there exists such a likeness. In reality mode there is only likeness and no relational suspicion at all. Thus, with a bit of experience, the crude photographic displays would not have posed a problem if a pictorial mode of processing had been used instead. Köhler's conclusions are thus accounted for by the fact that the chimpanzees saw the photographs as objects in themselves. Had they seen the photographs as *views* of objects, they would not necessarily be confused by a slightly distorted photograph, just as adult humans are not. It is analogous to viewing objects through a dirty window when you know that you are looking through a dirty window. Seeing the same view without knowing that you are looking through a dirty window would confuse you more.[9]

Not making the distinction between different forms of views on pictures is thus a mistake. Assumptions that pertain to a pictorial mode competence from data that can be explained by a reality or surface mode are unfortunately abundant in the animal literature. If we are interested in seeing whether animals can understand reference in pictures, and we should be because pictures can be a window to the mind as well as any language, we must look for experiments designed specifically for testing pictorial mode performance at the expense of other modes. The ambition of my own empirical work, which will be presented in Chapters 12 and 13, is just that. I have so far come across only a couple of attempts in the primate literature to directly address this question. Little has thus improved since a similar call for research by Cabe (1980) 30 years ago.

1.5 Outline of the thesis

Before I continue with the analysis of nonhuman primate picture data, let me first turn to human developmental psychology to make clear what is meant by understanding pictures as pictures, when and how the ability tends to develop in humans, and what types of pictures are easy and what types are difficult for small children to interpret. This will follow shortly in Chapter 2. Then I turn to cross-cultural research (Chapter 3). The focus of these chapters is on pictures as conceptual and perceptual problems. Semiotics (Chapter 4) lends valuable theoretical insights into speaking about pictures, which is lacking in much empirical work. The chapter introduces some novel definitions, which is the reason I present it comparably late. I do not want to change other researchers' terminology too much. However, for my analysis of the primate literature and my own data, a refined terminology will be useful. Part II, beginning with Chapter 5, reviews direct and indirect tests of picture competences in primates and some birds. Chapter 8 addresses the intriguing case of enculturation in apes and reviews picture use in such projects. In Chapter 9 the three modes of picture comprehension are applied also to sister competencies to pictorial comprehension, such as the use of scale-models and mirrors. More research seems to have been conducted on the referential use of these iconic media than on pictures. Since referential claims have also been made for ape painting and drawing, a foray into productive ability is inevitable (Chapter 10). Lastly, Part III is introduced by a methodological background to my own empirical work and a review of primate categorisation and its relation to pictorial competence (Chapter 11). Very different findings from Givskud Zoo (Chapter 12), and the Great Ape Trust of Iowa (Chapter 13), are then presented and discussed. Overall conclusions and suggestions for possible future directions for pictorial work with apes concludes the thesis in Chapter 14.

---

[9] When they have a choice, chimpanzees prefer to look at objects through clear windows as opposed to distorting ones (Menzel & Davenport, 1961).

Chapter 2

Child research

In most human cultures today we live a life saturated with pictures and pictorial information. Moving pictures, still pictures, photographs, drawings, computer generated 3D-art, road signs etc. We seem to learn to decode these objects as readily as we learn to speak. Pictures are generally seen as something simple, while language is a baffling achievement. However, this may be a premature assumption, evident even in a superficial glance. Language can in fact be argued to be exceedingly simple, since all average children learn it (Patterson & Linden, 1981).[10] Pictures, on the other hand, are not integrated into human biological adaptation, but are marks on surfaces with an inferred significance (e.g. Ittelson, 1996). Still we act, as caregivers and scientists, as if it is pictures that are intuitive to the human (and animal) mind.

When we are given our first picture books, the purpose is not that we shall learn to look at pictures, but to practice verbal naming. Perceptive caretakers have noted, though, that young children are quite tactile with pictures, and they have no trouble what so ever to browse a magazine upside down (Pierroutsakos et al., 2005). Concerned caretakers naturally turn the books back so that the right side is up, and tactile exploration they explain away as motor development: “You know children, they like to touch everything.”

Developmental psychologist DeLoache and her colleagues took these observations seriously. It is true that infants like to touch things, so why are they touching *pictures*?[11]

2.1 Grasping pictures

In the developmental psychology literature there is little focus on alternative views of the picture. It speaks about understanding the “symbolic” nature of pictures. It seldom acknowledges that you can actually do something with for example a reality mode competence and that it is not entirely trivial. They often seem to see everything but a fully fledged pictorial competence as just a lack of something: symbol-

---

[10] Although a complex phenomenon to describe, language is clearly not an advanced *competence*.

[11] Similar reality responses have been investigated by e.g. Beilin & Pearlman (1991).

ism.[12] Focus is thus on the referential use of pictures, which is also the reason small infants grasp at photographs: They perceive it as an object in itself, and not referring to something else (DeLoache et al. 1998a). They prevail in this behaviour despite cues like two-dimensionality and size transformations. DeLoache (e.g. 1987) calls the ability to simultaneously treat pictures as objects (e.g. a piece of paper with marks on it) as well as depictions "dual representation." This term has also spread to primatology (Boysen & Kuhlmeier, 2002; Tanaka, 2007b; Tomasello et al, 1997).

In a study on 9-month old infants using "highly realistic" photographs of familiar objects, all infants felt, rubbed, patted and grasped the depicted objects (DeLoache et al., 1998a). This occurred although infants at this age are fully capable of discriminating between real three-dimensional objects and depicted two-dimensional ones, although the exact nature of this perceived difference is not clear. Bower (1972 in Bovet & Vauclair, 2000) found that neonates stretched out for objects but not their photographs, and Slater et al. (1984 in Bovet & Vauclair, 2000) similarly found that infants as young as two days preferentially looked at objects rather than pictures of objects. However, Dodwell et al (1976) found that infants under 23 days did not seem to differentiate between objects and their pictures.

A second part of the DeLoache et al.'s (1998a) study thus aimed at investigating whether the grasping response was just standard infant investigatory behaviour at play or if they really attempted to act on depicted objects. Nine-month old subjects were therefore presented with objects simultaneously with life-size colour photographs of those objects. The infants preferred to reach for the real objects and ignore the pictures. This condition showed that depicted objects did not have the same status as real objects, but when presented in isolation they approximated the real thing in terms of manipulation. In addition DeLoache et al. (1998a) had noted that infants never showed signs of surprise when the objects they were aiming at would not come off the page, further suggesting that the infants accepted that the pictures were not identical to real objects. However, they were real enough to elicit investigatory actions. The researchers watched in bemusement as infants not only tried to pluck objects off pages but also e.g. tried to drink from depicted milk bottles. They concluded that the infants seem to investigate because they are unsure of the nature of the depicted objects. They literally tried to grasp the nature of pictures (*fig. 2*).

That infants really target depicted *objects*, and not just interesting patterns, was shown more clearly by Pierroutsakos (1998) who reported that 9-month old children who were allowed to play with objects that were subsequently depicted in photographs grasped less at the pictures of those particular objects than did children who were presented only with pictures. Furthermore, when grasping at pictures the group experienced with the depicted objects targeted those details of the picture that had an effect on the object, e.g. buttons that triggered sound.

For those who are still not convinced, a complementing control for the alternative explanation that infants just reach for interesting stimuli and not objects, is a condition in Pierroutsakos and DeLoache (2003) in which they put the photographs inside a darker oval to see if the infants' grasping was directed towards the highest

---

[12] "A symbol is something that someone intends to represent something other than itself" (DeLoache, 2004). See Chapter 4 for a semiotic view of this term.

contrast in the picture, which was the border between oval and background, or the depicted objects. Not surprisingly children grasped at the objects and not the surrounding ovals.

In a third study in the (1998a) paper DeLoache and her colleagues gave the same pictures as above to children from a community in the Ivory Coast, who had very little exposure to printed pictures. Infants 8 to 18 months old responded in a similar way to the depictured objects as the American infants did. Objects that were not familiar from the local culture were grasped as much as familiar ones.

The same kind of grasping behaviours that have been reported for photographs are also elicited by objects on a TV-screen for e.g. 9-month old infants. When movement is added the infants follow the object with their hands (Pierroutsakos & Troseth, 2002).

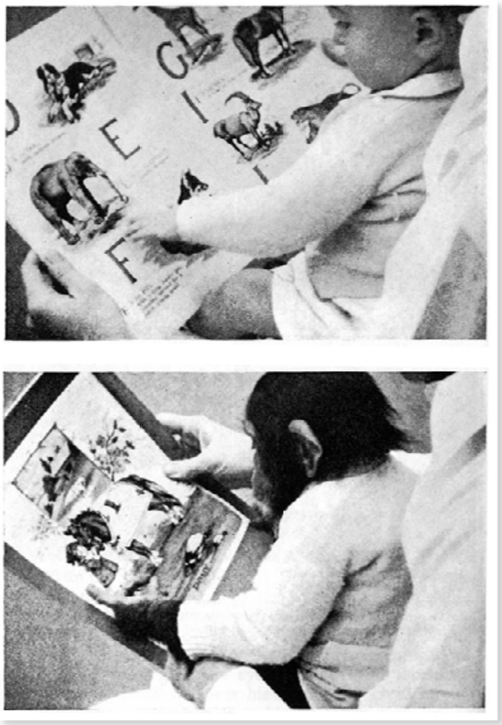


Figure 2. Donald (*Homo sapiens*) and Gua (*Pan troglodytes*). "When looking at colored pictures each will point to them and scratch the surface of the paper as if to pick them up" (Kellogg 1933/1967).

When DeLoache et al. (1998a) compared 9, 15 and 19-month-olds in the experiments above they found a notable change with age. Grasping was very rare in 19-month old children, who instead had a higher number of pointing to and vocalising to the pictures in communication with the adults present. The pointing of the older children suggests that a central aspect of what children learn about pictures as they grow older is that pictures are something to communicate about with e.g. caretakers. This has lead researchers to propose on the one hand that children's ability to attend to the actual visual content of pictures might get impaired (see section 2.3), and on the other hand that it is the very source for pictorial competence (see section 2.4).

In the TV-condition infants likewise replaced physical actions on the screen surface with pointing and vocalising when 15 to 19-months old (Pierroutsakos & Troseth, 2002). However, such changes in behaviour does not on its own imply that young children have started to understand the referential potential of pictures, only that there is no use grasping them, and that sharing attention to them with adults can be joyous.

In an attempt to show that it is the realness of photographs that affords investigation Pierroutsakos and DeLoache (2003) compared 9-month old infants' manual investigation of colour photographs, black-and-white photographs, colour line drawings and black-and-white line drawings. Line drawings were based on the photographs by tracing the objects in ink.

There was only a significant difference in grasping between the colour photograph condition and the black-and-white drawings, not between the intermediate conditions. However, because there was a linear relationship between what they judged to be a degree of realism (the four conditions) and the amount of grasping, they concluded that grasping varied with degree of realism. But the results could also mean that anything less than photographs were not perceived as objects at all, and thus there was no continuum in the perceived likeness to the real world to speak of. It could be that the children grasped at objects in the photograph conditions, or perhaps only in the colour photograph condition, but grasped at interesting patterns in the ink conditions. That would mean that there was never a relation between degree of realism and grasping, but a comparison between two forms of interest: That of investigating objects and that of investigating interesting patterns. The least interesting was the black-and-white ink drawing, but to say that it was the least interesting because it was the least real looking object is an assumption awaiting more clear support.

What requirements do pictures have to fulfil in order to appear real to the subject? Why is not a line drawing a very real but unknown object, and thereby actually being even more in need of investigation than a photograph? That is, is recognition necessary to elicit manual investigation?

In Pierroutsakos and DeLoache (2003) the researchers also, as mentioned above, placed photographs inside an oval to see whether the infants' grasping targeted the area of highest contrast or the objects. They targeted the pictures of the objects twice as much as the edge of the oval. What I wish they would have tested as well, to address my concerns above, is to have replaced the photograph with a nonsense shape and see whether the grasping behaviour would compare to black-and-white line drawings, colour line drawings, black-and-white photographs or colour photographs.[13] If they would not grab at all at a nonsense shape or an empty oval, I am willing to admit that perhaps they did see something real in the ink drawings.

One should keep in mind that the above experiments do not prove, or require, that the infants can adequately categorise the objects in the pictures, just react to the realness of them. However, recognition is not trivial. When children are better able to categorise the depicted objects it ought to be easier to start differentiate, and with time realise the referential nature of pictures.

2.2 The Dual Representation Hypothesis

It is still often assumed that learning is not necessary in order to interpret highly iconic pictures. A reason for this belief is that several studies have shown that an infant can recognise familiar three-dimensional objects in two-dimensional photographs from a very early age. But, as have been stressed in this text several times already, recognition does not equal understanding pictures as pictures (e.g. DeLoache

---

[13] They seem to have grasped twice as much at the edge of the oval than at the black-and-white ink drawings, but less than at colour ink drawings for example.

& Burns, 1994). If the latter ability is to be demonstrated, the child must show that it is able to relate pictorial information to the real world, but without confusing the two.

In development, perception of pictorial information precedes conceptual understanding of what a picture is (Pierroutsakos & DeLoache, 2003). One must learn that a picture is both similar and different from what it depicts. Developmental psychologists agree that experience leads to the concept of picture, a concept that according to DeLoache and colleagues includes that a picture has a double nature and that there are culturally appropriate uses of pictures. The former requirement is dubbed the *dual-representation hypothesis*. "To understand and use a symbol, *dual representation* is necessary - one must mentally represent both facets of the symbol's dual reality, both its concrete characteristics and its abstract relation to what it stands for" (DeLoache, 2003). This notion is inspired by the "duality of pictures" á la Gibson (1979) and the "double reality of pictures" from Gregory (1970), both as referred to in Pierroutsakos and DeLoache (2003).[14]

DeLoache and colleagues assume that it is difficult for the child to hold these two aspects in mind simultaneously: that a picture is an object in itself, and that it is about something else. Does the dual-representation hypothesis explain reference? No, only what is necessary: an ability to simultaneously entertain two views of the "symbolic" artefact which can otherwise compete with each other. They believe that the physical aspects can obscure the referential part, which make infants' interest in depictions as real objects an initial problem for true pictorial competence. This is attributed to a hierarchy of prominence in e.g. Sonesson (1989). In virtue of being in closer connection to the perceiving system some objects are more attention grabbing than others. A piece of paper, or a model, as objects in themselves can be very prominent indeed in relation to a referent removed in time and/or space, not to say the specific relation between these two entities.

DeLoache first became aware of the problem of dual representation when she found that 3-year-old but not 2.5-year-old children could realise the relation between a scaled down model room and a real room (DeLoache, 1987). The task had been to first observe a hiding event in a scale model, or one being indicated on a picture, and then find the full-sized equivalent to the object in the real room, or a larger scale model. That it was a problem of dual representation became apparent when it was found that 2.5-year old children who failed in the scale-model task could succeed in a similar task when utilising photographs instead (DeLoache, 1987; DeLoache & Burns, 1994). Photographs have less object properties than do a model room with miniature furniture. The latter affords play and investigation in its own right.

2-year-olds failed also in the photograph condition despite various modifications to simplify the task, forcing the researchers to conclude that "24-month-old subjects did not interpret pictures as representation of current reality" (DeLoache & Burns, 1994). The reason they added "current reality" was that children were surprisingly good at *placing* toys in a real room in locations that had been pointed out to them in a photograph. These experiments are thus not a black-and-white test of referential understanding of pictures, but about the nature and use of this reference.

---

[14] Haber (1980) calls the same phenomenon the "dual reality of pictures."

A retrieval task involving a picture or model requires much more of the subject than just realising *that* there is a relation (DeLoache & Burns, 1994). As with the simpler task, recognition of the objects and their spatial relationships is necessary, as well as the formulation of this relationship into a proposition ("the object is behind x"). This should then be stored in memory.[15] The subject must then discriminate between picture/model and reality, and also understand the relation between the two spaces. This relationship can involve many things, of which understanding that the picture depicts *current reality in an adjacent room* seemed to be specifically problematic in the retrieval tasks of DeLoache and Burns (1994). In addition the task requires more than global reference from model to real space. There is also a need for specific mappings between points in the scale model or photographs, and points in the room referred to. It is thus a task that requires e.g. parsing of part-whole relationships, as well as memory for those relationships.

A similar suspicion can be raised regarding the performance in tests where children have been asked to place stickers on a picture or a doll to indicate where they have stickers on their own body (Smith & DeLoache, 1996). Children 2.5 to 3.5 years old place stickers in the right place on a photograph only 55% of the time, with younger children making the most mistakes. In the model-search task they might thus have understood the overall reference between the two spaces, but been unable to recognise the significance of local mappings. Recognising such mappings might have been more salient in the photographic condition.

Finally, but crucially, the subject must be able to act on the knowledge gained above without interference from conflicting memories.

Despite all the above factors, the main reason for the failures with scale models for the 2.5 year old children was hypothised to be that the model was just too much an object in itself to serve a referential function: i.e. a problem with dual representation. DeLoache (2000) report that when the scale model was placed behind glass or when children were not allowed to interact with it the function of object-in-itself diminished and the referential nature became more apparent, leading to more successful searches by the children. This discovery could also be helped by instruction, focusing on the intent behind the scale model (DeLoache et al., 1999).

The workings of dual representation also became apparent for video stimuli in Troseth and DeLoache (1998). A hiding event taking place on video could not be solved in real life by most 2-year-olds but by most 2.5-year-olds. However, when seeing the same event through a window all 2-year-olds were successful at later retrieving the object. Furthermore, when being *led to believe* that they saw the event through a window, when in reality they were watching a video tape, performance approached the window condition.

Similarly, when infants are made to believe that an actual room has been shrunk to a miniature model by a magical shrinking machine, and then subsequently enlarged again, children (2.5 years) are more than able to find an item in a room that was hidden as a miniature item in the model (i.e. shrunken room) (DeLoache, 2004).

DeLoache and her colleagues have in the research described above clearly shown that it is the correspondence between two entities that seem to be the problem for 2

---

[15] Delays are detrimental also to 3-year-olds' attempts at object retrieval (Uttal et al., 1995).

year old children. By 2.5 years they seem to have overcome this for pictures but not for scale models. Interestingly, a reason Troseth and DeLoache (1998) expected also younger children to be able to successfully retrieve an object hidden on video was that nonhuman primates had seemed able to (i.e. Menzel et al., 1978). However, the data from the children supports the interpretation that the apes succeeded precisely because they took the video events to be some kind of events through a window (see section 9.3).

Suddendorf (2003) has pushed back the age for succeeding with both photographs and video cues in retrieval tasks to 2 years. Instead of four trials on the same room he has used single trials on four rooms. According to Suddendorf the reason for the low performance in earlier studies was that small children are subject to strong perseveration effects, which means that they tend to search at the location where they last found a hidden object. When counteracting this by only letting them do one search per room they perform as well as 2.5-year-olds. However, the respective experiments were conducted in different laboratories, which is always a complicating factor. Sharon & DeLoache (2003), however, maintain that perseveration errors are a consequence of failing to adequately realise the symbol-referent relationship, since it in their study is apparently not a consequence of poor inhibitory control. Further research is clearly needed to settle this.

But succeeding in a search task does not mean that all types of iconic reference is understood. In a simple test by Callaghan (1999) children were shown a drawing of an object and were then required to put a similar object down a tube. But the task was framed in an ambiguous way. Given a choice of the referent object or a second picture of the object, 2, but not 3-year-olds, often put pictures down the tube. The youngest children in Callaghan's study also failed with the simplest types of discriminations, that of a straight line versus a circle representing a stick or a ball.

Callaghan (2000) even insists that the appreciation of the "symbolic nature" of pictures (and objects like scale models) per se, does not actually appear until the middle of the third year of life, and for some mediums and domains even later. She suspects that the performance in DeLoache and others' studies were biased by the fact that they used pictures of familiar objects with verbal children, and that verbal labelling would affect the use of pictures in guiding action. It was rather the symbolic function of language, and not the symbolic function of pictures, that guided these children.

There are further findings that suggest that the referential nature of pictures can be grasped much earlier than at 2.5 years of age. Preissler and Carey (2004) report that after associating a new word with a picture of an object, subjects as young as 18 months choose the object rather than the associated picture when asked to indicate the novel entity in a choice between the two. This is contrary to the result of putting pictures down tubes in Callaghan (1999). They had learnt that the new word referred to the object in the picture and not the picture itself. This is argued to show that the children understood the referential nature of the picture. Importantly, these were not photographs, but black-and-white line drawings, albeit prototypical views of simple objects. This would preclude that the children worked in reality mode and just chose the three-dimensional versions because they were seen as more proper exemplars of the same thing. Note, however, that this is a weak test of iconicity. No interpretation of novel pictures at the expense of other novel pictures is required. And again, language is an integrated part of the task.

A follow-up study by Preissler and Bloom (2007) address both these issues. In a setup where 2-year-olds were shown two novel objects and two novel drawings, children generalised from drawing to object if the drawing had been named (e.g. "this is a whisk, can you find another one?"), but from drawing to drawing if no label had been used (i.e. "can you find me another one of this?"). Depending on context the children thus attended to either the drawing or to the picture as a piece of paper. Thus, the dual nature of pictures, as well as interpretation of simple line drawings, can be grasped by 2-year-olds if the problem is properly framed (by language).

2.3 Language and iconicity

"[...] the symbol is always different in some way from that which it represents. What makes something symbolic is human intention; an entity becomes a symbol only as the result of a person using it to denote or refer to something," writes DeLoache (2003). When talking about pictures as symbols this argument is extended also to pictures, and reference in pictures likewise equals the *use* of pictures. But for pictures this use is of course not independent of iconicity. It is therefore not *only* a case of intention.

Callaghan (2000) investigated the effect of iconicity on picture recognition, together with its interaction with language. The hypothesis was that language facilitates the interpretation of pictures that are low in visual realism.[16] Language in Callaghan's (2000) study was the availability of verbal labels for the referents, in terms of familiarity or applicability. Children tested were 2.5 and 3 years old. A matching-to-sample (MTS) paradigm was used where the picture was the sample and two objects the match and non-match respectively. That will say, the experimenter held up a picture (the sample) and the child was to match this sample by indicating one of the choice objects.

In the first study visual correspondence (iconicity) between picture and referent was varied and the match and non-match had the same verbal label (e.g. two types of dogs). The types of pictures were, from "abstract" to "realistic": black-and-white (cartoon-like) graphic drawings, black-and-white perspective pencil drawings, "realistic" acrylic colour paintings, and miniature replicas of the referents (*fig. 3*). 2.5-year-olds failed to match any pictures (and replicas) to the referent. The 3-year-olds performed well with all types of pictures but best with the "highly iconic" ones. Only the graphic picture was significantly worse than the others. This is interesting since both the perspective drawing and acrylic painting were made from a template while the graphic one was highly conventional. It seems like the choice of stimuli was not on a continuous scale at all, just as in the Pierroutsakos and DeLoache's

---

[16] Level of iconicity is a subjective measure, if it at all can be quantified. In fact, I would myself not equal degree of iconicity with degree of realism although they often covary (see Chapter 4).

(2003) study. Difference in style between stimulus pictures might not equal a similar difference in appreciated iconicity. Perspective drawing and acrylic painting seemed to have mediated a comparable amount of relevant cues despite being very different.

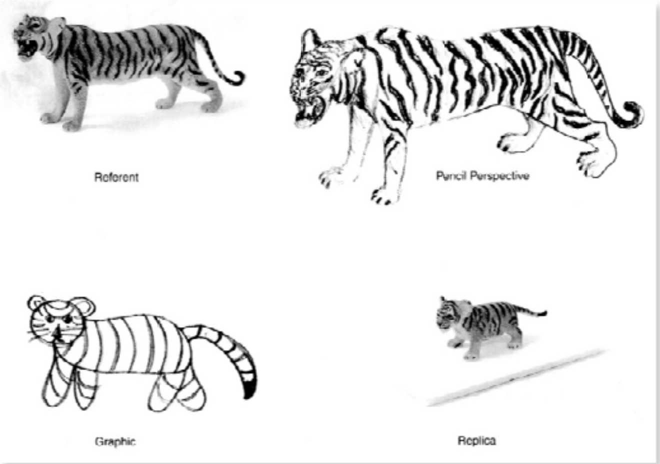


Figure 3. Picture and replica stimuli used in Callaghan (2000). (Note that the "acrylic colour painting" version of the stimuli is not included.) From Callaghan (2000).

In the second study only black-and-white pencil perspective drawings were used as sample. Match and non-match had very different verbal labels, familiar and unfamiliar to the children. 2.5-year-olds performed above chance when labels were available but at chance level when they were not. 3-year-olds were above chance in all conditions, but worse when labels were unavailable. The second study, together with the first, show that while older children match by the visual correspondence in the pictures, younger infants need to label pictures and perform on a verbal level.

In a third study the 2.5 year old children were equipped with the verbal labels when presented with the picture samples, and had been given the labels of the choice objects in beforehand. Thus, in this condition the pictures potentially played no role at all. Labels used were below basic category level (e.g. tiger cat vs. leopard cat). Performance went up, even when choice objects were from the same category.

By combining visual similarity of the choice objects (e.g. button vs. coin) and verbal labels in a fourth study, it was found that when labels could be used, and match and non-match were dissimilar, the performance was much better for 3-year-olds than when choice objects were perceptually similar and there were no labels available. In the latter case they failed miserably. However, in the intermediate conditions, where either choice objects were dissimilar or labels were available, they performed above chance. 2.5-year-olds on the other hand failed in all conditions except when labels were available and choice objects were very dissimilar.

The conclusion was thus that young children needed two types of information to succeed, both verbal and perceptual, while older children could do with one or the other. Among these, verbal information seemed to be the most important factor for the younger children. Callaghan (2000) speculated that this finding is hardly surprising considering the practices adults engage children and pictures in, which is mostly naming games. Young children are just not used to using pictorial information in the absence of linguistic scaffolding. It might be that picture book "reading" actually hinders children from using pictorial information independently of language, and instead drills them in replacing pictures with verbal labels. Not until they are older does the iconic route catch up.

In a longitudinal training study Callaghan and Rankin (2002) showed that pictorial competence could indeed be improved by playing picture games with young children (28 months old at start of study), where visual rather than verbal matching was the method. The pictures in these studies were simple pencil drawings. The control group, who got placebo training, lagged two months behind the experimental group in their comprehension of graphic symbols, and one month behind in their production. No children in the study had showed any comprehension at the onset of training.

This dependence on naming, and thereby inattention to visual correspondence, may help explain why young children fail in scale-model tasks and when placing stickers on dolls in the absence of instruction.

Language interacts with the use of iconic information in more respects than as labels. It can for example structure a situation so that necessary relations gain salience, and it can make iconicity redundant. As mentioned above, to use pictorial information in model-search tasks one must be able to appreciate the relation between picture and referent in at least three different ways (DeLoache et al., 1998b). Firstly, *that* there exist a symbol-referent relation (e.g. that a map shows a country). Secondly, *how* the symbol is related to the referent (e.g. that dots on the maps are cities). Thirdly, one must be able to *compute specific relations* between symbol and referent in order to use the symbol's information about the real world (e.g. that specific dots refer to specific cities). By 3 years of age children can do this in a model-search task, but there are interesting interactions between the three levels. If children have been explicitly instructed that there exist a global model-room relation they can find a toy in a room where the furniture does not correspond to the scale model, but with less instruction it is crucial that the internal elements (i.e. the furniture) between the two spaces correspond in order for the *stand-for* relation to be discovered and the toys be successfully retrieved (Marzolf et al., 1999). Again it seems that language relieves iconicity from its duties.

There are other means than instruction to heighten correspondence relations between items. DeLoache et al. (2004a) found a robust transfer effect *between* symbolic mediums (pictures to scale models), suggesting that training affects a general ability. However, instruction still plays a significant role in these experiments and instructions per se, not only experience with representations, also transfers between tasks (Peralta de Mendoza & Salsa, 2003).

Troseth (2003) attempted to train 2-year-olds to understand the connection between video and reality in an object-retrieval task by showing hiding events simultaneously live and on a video monitor, or by showing an adult model finding objects through video. The 2-year-olds did not become better at using video information on subsequent trials. When viewing the hiding events through a window, though, they could generally retrieve the objects.

I wonder how much faster it would have taken to teach the referential nature of photographs instead of line drawings in the Callaghan and Rankin (2002) study above. They only address one of the ways into pictorial competence. By using abstract drawings the scientists target the pictorial concept from the "referential side," while it is fully possible to first use pictures that the child already can decode, i.e. highly realistic photographs, and then add reference onto that.

However, child researchers argue that iconicity plays a small part in acquiring the symbolic function of pictures. “One reason that iconicity should not be considered criterial in thinking about symbols is that even the most realistic color photograph expresses a point of view regarding its referent” (DeLoache, 2004). That is, there is a symbolic aspect in even highly iconic pictures. This is certainly true, but more from an outside perspective. For the child there is definitely an unambiguous “referent,” which is the perceived object in the photograph, regardless of “expressed points of views.” Iconicity is of course central in this process and is necessary for learning the referential nature of pictures in terms of similarity *and* differentiation.

DeLoache et al. (2004a) for example brought differentiation into the picture when they made 2.5-year-olds train on photographs and subsequently perform better than 3-year-olds on scale-model tasks. This was a very different solution from putting the scale model behind a window to minimise its object properties. However, it served a similar differentiating role. Reference was possible in the photograph condition just because it was sufficiently differentiated from the real world, but still had a striking likeness to it in virtue of being photographic. Would reference have turned up as effectively if drawings had been used instead?

Although I admit that realistic pictures are perhaps in minority in an infant’s upbringing, and probably not the typical path towards reference, I would not exclude the possibility that iconicity can be a way into reference until it has been tested more thoroughly. Pictorial competence is after all an ability, or abilities, with a cultural foundation, and as such there can be many paths to the same end state. The typical route might not be the most effective one.

2.4 Intentions and context

To explain the appearance of pictorial reference child researchers tend to invoke two of the hottest topics in current developmental and comparative psychology: understanding intentions and social learning.

In an attempt to connect pictorial competence to social learning, Callaghan et al. (2004) studied how infants ranging from 6 to 18 months responded to a model’s actions towards objects and pictures. The objects were children’s toys and the pictures were colour photographs of those objects. They had two conditions. The first condition was acting towards the objects in a manipulative fashion and the other in a “contemplative” fashion. The contemplative stance was to point to and engage the infant’s attention towards the pictures and objects, and the manipulative stance was just to handle them in front of the infant. Note that pointing occurred towards the pictorial content of the pictures, and not to the flimsy sheet of laminated paper.

Older infants, i.e. 12 months and above, did copy both manipulative and contemplative actions towards pictures, but not towards objects. Younger infants did not reproduce any of the adult’s actions. Callaghan and colleagues (2004) concluded that infants use imitation to learn conventions regarding pictures, and that these ways of acting towards pictures give the infant firsthand experience that is necessary for learning the referential nature of pictures a year or two later. They stress that imitation of a contemplative stance does not mean that the infants can use pictures as referential, although it can certainly look like it.

That older infants model only actions on pictures and not on objects Callaghan et al. (2004) attribute to social uncertainty about the proper actions towards strange objects. Ways of handling ordinary objects were already familiar by the time the subjects came to the study, but conventions with laminated pictures handled by experimenters were not. From this reasoning follows that pictorial media must always be new and strange if infants are to imitate conventional actions towards it. Indeed, whenever a pictorial medium *is* new to an infant in the course of its upbringing, adults around it act towards it in a contemplative rather than manipulative fashion, since that is the common stance towards pictures. Thus, from early on the infant is exposed to differential adult attitudes towards objects and their pictures, and are furthermore motivated to imitate contemplative stances towards pictures if unsure about their nature. As indeed shown by Gelman et al. (2005) objects and pictures do elicit different interactions between adults and infants. Objects for example generate talk about individual items, while pictures generate talk about categories. This pattern is seen in both children (2-3 years old) and their caretakers.

Callaghan et al. (2004) attributed the younger infants' failure to copy the contemplative stance to the insensitivity to communicative intentions that infants hold before 12-24 months of age, clearly inspired by the research of developmental psychologist (and primatologist) Tomasello (e.g. 1999). However, the infants did not copy the manipulative stance either, perhaps saying more about copying abilities than anything pictorial. As mentioned, the contemplative stance was directed towards the motif of the pictures, why it is not surprising that photographs and objects are treated in an identical fashion. The assumption that intention-reading and pictorial reference is intimately linked is a plausible one, but still an assumption. One must separate the intentional *use* of pictures in social interactions of various types, from the interpretation and meaning of pictures as they stand on their own. If one puts all the focus on uses, one denies the fact that pictures are connected to the world in ways which symbols (conventional signs) are not, i.e. through iconicity (see Chapter 4).

The following is an example of an alternative view. As a group, both children diagnosed with autism and children with mental handicaps passed an adapted version of the object-retrieval test, both with photographs and scale models, but just as average children they were somewhat better in the photographic condition (Charman & Baron-Cohen, 1995). We have been told that people with autism are particularly insensitive to the intent of others, but still they pass DeLoache's object-retrieval tasks.

The main reason for suspecting that the role of reading referential acts, e.g. pointing, (which should not be confused with reading minds) is helpful in developing a picture concept is the proven role of those abilities in word learning, which is in place at 19 - 20 months (Baldwin, 1993). It seems unlikely that those resources are not recruited in learning about pictures as well. However, from this one should not conclude that there is therefore a "single symbolic competence" that is unveiled across media.

Intentional use can indeed help to specify a pictorial referent. For example, the same photograph can refer to the particular entity that is depicted, the event the entity is taking part in, to a category of which the entity is a member etc., and the intentional use of the photograph can help pinpoint this. Recognising that the use *is* intentional might be crucial in this process. But intention is in this case only a specific form of context. The same referential judgement can be made from other contextual cues, such as exclusion, salience, or from plain experience.

It has for example been thought for a long time that children learn new words only if they have reasons to believe that the person using the word is naming something (e.g. Baldwin & Moses, 2001). However, in for example a test using referential ambiguity, children with autism performed as well as average 2 year old children when mapping new words to unnamed objects and pictures of objects, including line drawings (Preissler & Carey, 2005). The children with autism were described as “impaired in monitoring referential intent,” but still they had no problem to infer that what was called a novel name did not pertain to an object with a familiar name.

Children 2.5 – 4 years old interpret the same picture differently, as shown in their naming, if they are told that the picture is an accident with paint, than when they are told that someone has worked on the picture (Gelman & Ebeling, 1998). But by this time children already know that there are pictures and non-pictures. To know whether an object is intentionally crafted or not just helps them to apply this distinction. Similarly, objects that are presented with a story about intentional creation are named as artefacts (e.g. knife) while objects that are presented as accidental get named based on its physical properties (e.g. steel) (Gelman & Bloom, 2000). But this is also just a test of how naming works, not why naming is possible and whether “reading intentions” plays a necessary role in starting up that ability.

To return to Judy DeLoache’s research. 2-year-olds that fail in scale-model tasks, or fail to see video information as displaced in time and space, can solve tasks by being made to believe that the room has been shrunk (DeLoache, 2004) or that the TV is actually a window (Troseth & DeLoache, 1998). They can also be helped by instruction. There seems to be a “need for the experimenter to make the intentional basis for the symbol - referent relation clear by explaining everything about the task” (DeLoache, 2004). Why is it not enough to explain only the intentional basis? Everything else about the task has to be explained too, including the spatial similarities between two spaces.[17]

On a different note, the beneficial effects of instruction do not mean that all other paths are closed. The transformation of video information into window information also helps young children. Then differentiation, and not intentions, seems to be the central factor. Is for example knowledge about what cameras do in the world knowledge about intentions? Can one learn differentiation by learning what cameras do? (Although not a *typical* development it is not implausible.)

When interpreting photographs it has been found that 3, 5 and 7-year old children attribute changes in depicted viewing angles to changes in the referent, not movement of the photographer or camera. And if they do not detect a change in the

---

[17] However, Salsa & Peralta de Mendoza (2007) claim that 2.5 year olds benefit more from having the intentional basis for a relation explained than having the correspondence pointed out.

referent they can even have difficulties separating pictures that depict the same scene but from strikingly different angles (Liben, 2003). This effect decreases with age. These phenomena can be an expression of still having one leg in the grips of a reality mode understanding despite having a pictorial mode conceptualisation regarding photographs. Or to relate it to the dual-representation hypothesis, instead of performance being obscured by the physical properties of the picture, the properties of the medium is obscured by the referent. The children are somehow inside the photographs, and when inside, when something changes it must be the referent. The inability to take the photographers viewpoint and move around with it may also be attributed to a lack in perspective taking and not at all confined to the photographic domain. If that is the case it might not be photographic concepts as such that develop from 3 to 7 years, but social-cognitive ones. Whatever the case, and this is my point, first-hand experience with photographing ought to affect the ability to understand photographic viewpoints.

Callaghan (1999) postulates, inspired by the research of DeLoache and colleagues, that the developmental trajectory for pictorial competence is that first a picture is seen as equal to its referent (picture-as-referent), followed by the picture being an interesting object in its own right (picture-as-object), presumably because of how adults act with pictures (see above), and lastly pictures become symbols (pictures-as-symbols).

Perhaps Callaghan's trajectory is indeed a better order than the one I propose, if I would claim that pictorial mode stems from reality mode or surface mode. However, I am not that concerned about order, and my hierarchy is not primarily about the *development* of picture understanding, but a way to categorise different modes of performance with pictures. I imagine that pictures can appear as interesting objects regardless of a phase of pictures-as-referents, but I do believe that pictures-as-objects must precede pictures-as-symbols, but not necessarily as separate steps in development. I think of it in terms of attention. The subject must attend to pictures for the referential nature to reveal itself at all, and social processes help to guide attention towards pictures as an object category, as well as appropriate parts of pictures. However, I believe that the scientific focus on intentions, as in mind reading, exaggerates its role in this process. Intention reading is still poorly defined, and it does away with the role of iconicity in a worrying way. I do believe that "reading" intentional behaviour is helpful, again for attention directing purposes. I think differentiation is key in the process, but differentiation can take many forms, of which the socio-cognitive route is only one, albeit possibly an effective one.

Can pictorial reference be discovered on one's own or must it be pointed out, and how "intersubjective" does this process has to be? This issue has partly been a driving question behind my own empirical investigations, where picture naïve apes have been given pictorial tasks (see Part III).

2.5 Pictures as cultural artefacts

Let us return to caretakers and their concerns about the way in which the newest member of the family reads her picture books upside down. Pierroutsakos and colleagues (2005) became interested in the fact that many infants are insensitive to the orientation of pictures, often preferring to look at pictures upside down. Similar behaviours can be seen in apes that browse through books and magazines, e.g. the gorilla Koko (see e.g. Brennan & Visty, 1999).

One-and-a-half to 2.5-year-olds that are handed a picture that is positioned upside down do not turn it around, but continue to study it in its inverted orientation 71% of the time. Despite this they can still identify the motif equally well as when the picture is in an upright orientation. The same takes place if an adult reads to the child in an upside down picture book. The child does not correct the adult.[18] When it comes to objects children tend to choose objects held upside down and objects held upright equally often, but most of the time (85%) they reoriented the inverted objects before interacting with them. The re-orienting behaviours with objects versus pictures are thus quite different.[19]

Because discrimination of orientation has been shown to be present from at least 6 weeks in habituation studies, and 2 to 4-month old infants look longer and smile more to photographs of faces presented in an upright fashion than inverted ones, the conclusion of Pierroutsakos et al. (2005) was not that the children were insensitive to orientation, but simply that they had not yet acquired a cultural convention for picture orientation. One can speculate that since it is equally efficient for young children to look at pictures that are upside down as are upright, the adult way of looking at pictures is not a very transparent convention and thus is acquired comparably late.

Adults, on the other hand, are used to viewing pictures in a canonical orientation, where heads are up and feet down and objects rest on surfaces as in daily life. Adults are more skilled at remembering pictures in their canonical position than pictures that are upside down. The hardest to remember seem to be inversed pictures from domains in which the individual is very experienced (Pierroutsakos et al., 2005). Children approaching school age also show better processing of upright pictures than inverted ones, perhaps due to a few years of always looking at pictures in a canonical fashion. (More on picture orientation, but for nonhumans, see sections 6.3.4 and 7.1.)

It seems that picture orientation conventions actually narrow down our competence, since we become increasingly poorer at processing inverted images as we grow older. What other types of picture conventions might develop independently from picture perception, but feed back into picture recognition? We saw earlier that for example the accurate attribution of points of view in photographs might be learnt independently from perceiving photographs as referential. With time we learn to expect specific such points of view. When we experience a very rare one we can

---

[18] However, in a previous study 2.5-year-olds readily corrected adults with picture books (DeLoache et al. 2000, cited in Pierroutsakos et al., 2005).

[19] This might reflect a difference between objects for manipulation and objects for inspection.

therefore easily get confused. Every angle that we have seen of objects in real life should in theory be equally easy to recognise in pictures, had it not been because we have specific expectations for what should be presented in them. That said, contextual information is scarcer in a photograph than in an experience of physical objects in a familiar world. Contextual information is important in order to make sense of any scene.

A case against the role of learning to perceive pictures through a communicative use, instruction, or even experience, is an odd but important study by Hochberg and Brooks (1962). They wondered whether a child, in this case a boy, could recognise objects in photographs, and furthermore line-drawings, if he had never received any instruction or training whatsoever in pictorial representation. His vocabulary was to be practiced solely with real objects instead of pictures, he was not to be given picture books, and no comments in regards to incidental pictures were allowed in his presence. He was also to be kept away from television. Although occasional pictures were unavoidable the boy's attention was at such occasions always redirected from them, without commenting their contents. Billboards and labels on jars of baby food were particularly tricky to avoid, however. There was also a decal that depicted babies on the boy's high chair. When 19 months old he turned around in his chair, pointed to the decal and said "baby." At around the same time he also caught a glimpse of a horse on television and uttered the word "dog." He then started to spontaneously seek pictures out, and it was judged that he could not be kept away from them any longer.

Thus, at the age of about a year and a half the boy was subjected to two tests. In the first one he was given 21 pictures (of which at least 8 seem to have been line-drawings) of objects in his spoken vocabulary, e.g. cars, dolls, shoes, keys etc. (*fig. 4, left.*) He was given the drawn exemplars before their photographic counterparts in order to minimise generalisation from photographs to drawings. The process and way of response is unclear, but it is likely to have been verbal. The sole description reads "The drawings were handed one at a time to *S*, a somewhat unsuccessful attempt being made to convert the test to an interesting game. Responses were obtained by tape-recording" (Hochberg & Brooks, 1962; p. 626). The boy was judged to have responded incorrectly on only four pictures, of which three were photographs. The only drawing he failed on was one of a duck.

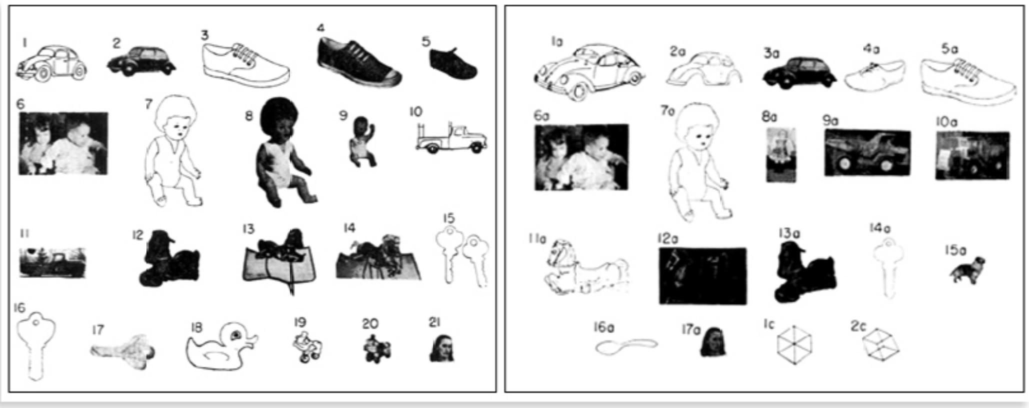


Figure 4. The drawing stimuli used in the first (left) and second (right) experiments in Hochberg and Brooks (1962). From Hochberg and Brooks (1962).

Then the boy was given unlimited access to pictorial materials (but still no television) for a month, while instruction or naming games were avoided at all cost. The boy himself, however, frequently named pictures. A second set of 19 pictures, of which 10 were line-drawings, were given in the same procedure as before (*fig. 4, right,* p. 32). This time a drawn shoe and a spoon proved difficult, as well as two photographs, of which one had proved difficult in the previous set as well. This was a small photograph of his mother's cut-out head. Identification of two perspective drawings of transparent "boxes" yielded a mixed interpretation from interraters.

Although a small dataset was used, the pictorial innocence of the child uncertain, and the procedure unclear, the experiment shows that a child that has not been explicitly *interacted* with in a pictorial context can still quite readily identify both photographs and drawings of a certain kind. Adult competence might just be a larger set of resources and experience in this regard.

It is important to note that the line-drawings used in the study have been traced from photographs and thus retain the silhouettes of real objects. The exception is perhaps the duck, unless it is traced from a photograph of a bathtub duck, in which case it is a spot-on carbon copy. These line-drawings highlight both shape and salient interior properties. There is no report on whether the child manually investigated the pictures, hinting about his ideas about their nature, but given his age it is not unlikely that he did (Callaghan et al., 2004). There are many other types of pictorial abstractions that would perhaps not be as accessible to a naïve picture viewer. A larger dataset, with a bit of variation, would have been most enlightening, although it would of course be impossible to say exactly which transformations were problematic. This is because of the different impact of different transformations in different pictures. A shadow might mean everything in one particular picture, for a particular viewer, while being redundant in another picture or for another viewer.[20]

The choice of response, i.e. the reliance on naming in a young child, also makes straightforward conclusions difficult. The fact that dolls, cars and keys were chosen as stimuli was probably because the boy could name these objects. If the experimenter asks "what is this?" regarding a picture, this is a drill similar to one that the child has gone through countless of times with objects. The child perhaps realises the similarity between these experiences and maps the pictures to a limited set from its vocabulary, rather than to the visual universe at large. Keys, cars, and dolls, among other drilled objects, are then in a privileged position for comparison with the pictorial versions.[21] The Hochberg and Brooks (1962) study demands replication. If it is unethical to deny human children social interaction around pictures, perhaps nonhuman animals can help us shine some light on these issues. But let us first take a look at cross-cultural research, with a special focus on inabilities to recognise depicted objects and events.

---

[20] Iconicity is always subjective.

[21] The same potential problem occurs when a child is required to match a picture to an array of objects when objects and pictures are visible at the same time, or when verbally asked to find a particular picture among other pictures. The situation sets up a context that narrows down the number of interpretations that are probable.

Chapter 3

Cross-cultural research

Claims are sometimes made that natives in exotic lands cannot recognise pictures, flee in terror when watching a film, or wonder how their child got into a piece of paper. What is the factual basis for such stories? What does *adult* pictorial competence look like when the person has no previous experience of pictures? This question is relevant from a comparative perspective because apes are not human children. All sources that can help us see what pictures mean to the naïve eye are potentially helpful in understanding the requirements for pictorial competence. A picture is a constructed object. The image has travelled e.g. thru a camera lens, or thru the muscles of a painter, so to speak, and ended up on a surface. Sometimes one must know something about this process to be able to perceive the content of a picture with precision. Techniques for perspective rendering, shading, etc., often contain conventional aspects, which one gets familiar to with experience. That said, some conventions are superfluous for recognition, and some techniques that would seem conventional at first glance are in fact based on everyday perception. Nevertheless, the process of recognition can look very different depending on one's experience with a particular type of picture.

If indeed there is cultural variation in picture perception we can conclude that picture specific experience, and not language alone or general human intelligence, is all-important for recognition of the content of a picture, and furthermore that a picture is a picture. However, few studies have been primarily concerned with the ability to perceive pictures as pictures. When it has been found, for example, that some people can decode photographs but not drawings, it has been deduced that drawings are just poor examples for naïve subjects. Yes, they are poor examples, but I will argue below that this is not only attributable to experience of conventional techniques, but is in some cases due to the fact that such pictures require a different mode of picture processing than does the typical photograph.

3.1 Drawing conventions

The less like its referent a picture is, the more it depends on techniques to represent properties of the referent, rather than *actual* likeness. What kinds of experience with these techniques are necessary? In a review of cross-cultural picture research, Miller (1973) suggests that less amount of experience is needed to perceive an object in a colour photograph than in, say, a line drawing that relies heavily on techniques. A reason for why Miller stresses this point is that arguments have been made that differences in portraying for example the human form, reflects differences in perceiving humans in the world. To find such differences was in fact the original mission for cross-cultural picture research (Miller, 1973). As an example, do an ancient Egyptian and a modern westerner perceive Egyptian tomb drawings in the same way?

The focus of cross-cultural research has thus been on depth perception, optical illusions, and consequently relations between entities in pictures, rather than recognition of entities as such. In experiments, actual recognition and identification has therefore mostly been a *requirement* rather than targeted in controlled testing.

An example is embedded figure tests, reviewed in e.g. Jones & Hagen (1980). In cross-cultural research performance on such tasks shows a clear correlation with exposure to western culture (and gender). But most of the time embedded picture tests are made by showing a pictorial target rather than verbally asking to identify a specific object. Identification of real-life referents can thus be bypasses and pure shape matching can take place. This can be one explanation for the cultural differences found, although a stronger "field dependence" in non-western cultures (and in females) is typically the preferred explanation.

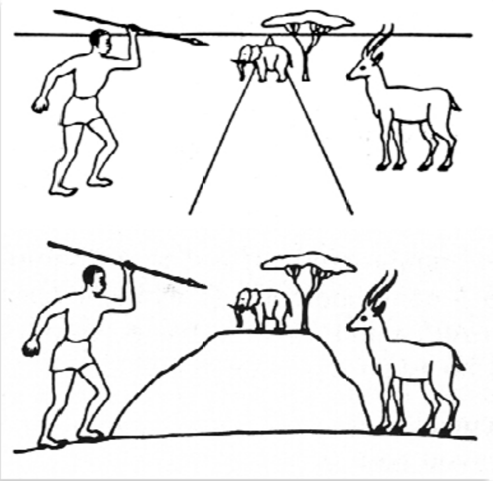


Figure 5. Examples of outline-drawings from Hudson's (1960) pictorial depth perception test. From Jones and Hagen (1980).

The most common case where recognition has been presumed is Hudson's classic studies of depth by presenting line drawings of hunters, elephants, trees, and antelopes in different positions and sizes relative to each other, as well as to hills and roads in the pictures (*fig. 5*) (Hudson, 1960, in Deregowski, 1989). But in order to appreciate that the elephant is further away than the antelope, only because it is depicted as smaller, the entities must be recognised as elephants and antelopes in the first place. Furthermore, they must not be interpreted as e.g. baby versions of the animals. Not surprisingly, the task of judging depth was far from simple.

Hudson (1960, in Miller, 1973) reports a great variability in performance of over 500 South Africans on decoding drawings that utilise depth cues. Factors that implied exposure to western forms of drawing correlated highly with ability to perceive depth in the material. It also seems that subjects performed a bit more three-dimensionally when a control photograph of models was used instead of drawings (Jones & Hagen, 1980). Kilbridge and Robbins (1969) replicated Hudson's findings with Ugandan Baganda subjects, including the conclusion that exposure to western culture was the major factor for the subject's performance. Other replications found great cultural biases in Hudson's material, as well as strong effects of the wording used when questioning subjects (Jones & Hagen, 1980). This was found for both picture experienced and more picture naïve subjects.

Misidentification of the animals and landscape features in the Hudson's pictures were common in both rural and urban subjects, but targeted at different entities, presumably mirroring their real life experience of the referents. Consequently, inferred depth in the pictures was heavily confounded with recognition of the individual constituents of the scene. Omari and MacGinitie (1974) compared Hudson's original drawings with drawings that used the same depth cues but contained entities more familiar to the subjects. With Hudson's pictures Tanzanian school children in all ages performed poorly, but with the adapted material they performed significantly better, and furthermore showed an increase in performance with age.

In a subsequent study on encoding depth in Hudson's material, Indian students, who had experienced a rich pictorial culture since birth, also performed poorly (Hudson, 1962, in Miller, 1973). This was ascribed to the fact that oriental art depicts depth a bit differently from the western conventions used by Hudson. But although pictorial styles differ between cultures there is also considerable overlap. Most of the times scenes and objects can be identified, but often they look distorted from the point of view of an observer from the other culture. The inability to infer depth in Hudson's pictures is probably an effect of this. However, such effects seem to wear off with exposure to the new style (Deregowski, 1989).

But Miller (1973) finds the finding most puzzling, since oriental art uses at least superposition and size to depict depth, just as Hudson's drawings. Miller thus makes an assumption that seems common in the literature, which is to believe that recognition is about learning general rules of transformation.[22] Pictorial techniques can be described as such rules, and because of this the ability to decode pictures can be studied as the ability to read such cues, or the sensitivity for such cues. Some of these cues are believed to transfer from the real world to the pictorial one. Segall et al. (1963) for example tried to show that people's experience of sharp angles in the real world makes them more prone to certain visual illusions in pictures that depend on relating lines to each other. People in "carpenter societies" were thus more susceptible to such illusions than were people with less angular environments.

However, that there are rule-bound cues in pictures does not mean that picture perception, and furthermore learning to perceive pictures, is a matter of learning general rules, isolated from context. Rather than being a case of applying decontextualised rules, cues might rather be learned as parts of typical pictorial scenes, and what goes on is relating one scene to another. We should not underestimate human recognition memory and take for granted that encoding general rules, however it is done, is efficient and therefore more plausible as an explanation. For example, a general principle is that a relative size reduction signals depth, i.e. that something is far away. The same general technique can be used to depict a boat that is far out at sea, as for depicting an acacia tree far away on a savannah. But one must know something about boats and acacia trees to infer depth from the fact that the boat and the tree are painted as very small in relation to something else in the pictures. One can learn all and everything about depth in sea paintings, but not necessarily

---

[22] See also the research of e.g. Segall et al. (1966, reviewed in Deregowski, 1989)

transfer this experience to savannah paintings. The same is true for perspective drawing. Seeing a box drawn in perspective is not necessarily automatically generalisable to a garden path that narrows as it disappears into the garden, although the general principle is the same, i.e. that parallel lines converge in the distance.

Furthermore, these examples are not conventional principles, but are derived from the everyday perceptual world. Still many picture-naïve people have shown poor ability when it comes to decoding cues that define depth, even though the ability to understand *that* a flat image can depict depth is not lacking. Even though the principles are the same, the viewer's contact with this information is different in a real scene and in most depicted scenes. Perspective in real life has its convergence at the level of one's own eyes, while a picture, if it is at all accurate, is constructed from the position of the producer's eyes, which is only in line with a secondary viewer's under special circumstances. Since objects in pictures thus do not change perspective when the viewer moves its head in the same way as real objects would transform, valuable feedback about the third dimension is lost. Such information must be filled in by other means. Visual effects that can counteract this loss are for example shadows, occlusion, size reductions etc. Often they are sufficient, but sometimes they do not seem to fill the gap.

Even in western cultures children are not able to readily infer depth in perspective drawings until a considerable age. To Hudson's drawings children in their first school year were likely to respond to the drawings as picture-naïve adults, and Newman (1969) did not find robust depth perception in drawings until children were 10 years old. Among 6 year olds, only a quarter of the subjects ascribed depth to Newman's material. But, surprisingly, they were sensitive to illusions created by depth cues. So while a part of them took depth cues into account, otherwise there would be no illusion, the part of them that interpreted the scene ignored the very same cues. In this specific case a correct answer would be something like "a corridor," or "a tunnel," but two-dimensional viewers instead answered e.g. "a television aerial" or "a pattern." However, rather than a failure to apply certain principles the younger children can just have had very little experience with abstract corridor pictures. Had a door been placed at one of the walls, or a painting been hung somewhere, the "corridorness" of the picture, *including depth*, might have popped out. Unfortunately Newman (1969) only focused on one type of picture and did not explore the effect of additional cues. Another help could have been offered by giving the children an array of possible answers. Again, context is vital for interpretation.

A significant difference between two-dimensional and three-dimensional perceivers was found by Deregowski (1969, in Miller, 1973) using a drawn illusion that made an impossible figure, but only if you recognised its three-dimensional nature. Just attending to the patterns of lines making up the figure did not bring about an impossible figure. Thus, after a 10-second delay after having seen the figure, two but not three-dimensional perceivers could reproduce the drawing. The reason for this was that the three-dimensional perceivers tried to remember an impossible object, which taxed their capacities differently from remembering merely a set of lines. A difference between viewing a geometrical drawing as two or three-dimensional also becomes evident when letting subjects construct a model of a drawing. They only include a third dimension if the picture has been interpreted likewise (Deregowski, 1989).

Deregowski (1968, in Miller, 1973) also found that children (mean age about 4 years) who were categorised as two-dimensional perceivers by Hudson's method with drawings, could easily model three-dimensional representations in clay from geometrical shapes that were presented as drawings. The conventional aspects in Hudson's material thus seemed quite high, and some of the problem might lay in the ability to recognise the entities that were supposed to have a spatial relation to each other in the scene.

3.2 Prominence

In his review Miller (1973) indeed finds evidence for the necessity of experience with pictorial stimuli in order to perceive the content of many pictures. However, rather than targeting the picture as a referential entity, the criticism of studies, such as Hudson's above, has mainly focused on the stimuli used, instructions given, and the techniques used for questioning the subjects. These are valid concerns, but there lingers an assumption in this focus, which is probably fuelled by our own observations of children and pictures, namely that pictures are easily seen *as pictures*. Most people from different cultures report that they see something in pictures, although what they see can vary. Are there cases where people fail to see anything in pictures that are easily decoded in other cultures?

For a naïve picture viewer there is at least one central problem, from the perspective of perception, which one must deal with before an analysis of a picture *as a picture* can occur. One must get to grips with the novel medium as such. In virtue of being e.g. a print on a flat surface, perhaps of a certain glossiness etc., a picture seldom reproduce all the properties of a real object, as well as adding some transformations of its own.

Herskovits (1959), cited in Miller (1973) report: "[…] a Bush Negro woman turned a photograph of her son this way and that, in attempting to make sense out of the shadings of greys on the piece of paper she held. It was only when the details of the photograph were pointed out to her that she was able to perceive the subject" (p. 136).

Segall et al. (1966, in Miller, 1973) attribute this trend to the fact that the more dominant aspects of a stimulus draw the attention of the subject away from more modest details. Thus, the sharp edges of a photograph, its feel in the hand or reflexions in its glossy surface, can outcompete the shades of greys within, in this case, a black-and-white photograph itself. Attention wise a captivating medium obscures the message. The effect has been called "prominence" elsewhere (e.g. Sonesson, 1989). Because of how we tend to compare things in a non-symmetrical manner (see e.g. Rosch, 1975), it seems that a less prominent object can stand for a more prominent one more intuitively than vice versa. As long as the surface is the more prominent part, one will not attend to the markings. This can be an additional reason for why pointing out the features in a motif generally helps.

Deregowski et al. (1972, in Deregowski, 1989) give a further example: "When they were given pictures printed on paper they attended to the paper – a strange material to them – and not to the surface pattern. They felt the paper, sniffed it, crumpled it, and listened to the crackling noise it made; they nipped off little bits and chewed them to taste it" (p. 57). When the same pictures were instead presented on coarse cloth, a more familiar material, the subjects did attempt to make sense of the patterns, although they were not always successful at this. Learning about pictures is thus also learning about the characteristics of pictures as physical objects. Miller (1973) explains that as soon as the referential nature of pictures is pointed out, attempts at decoding the motif will take place quite naturally. However, this is not always the case; "Some see a picture instantly, while old men fail to see anything at all no matter how long and patiently one tries to explain the matter to them" (Kidd, 1904, pp. 282-3, in Jones & Hagen, 1980, p. 195).

When, on the other hand, the cues that inform of the physical properties of a picture is removed or reduced, as in a projected slide show, the effect can be dramatic to the other extreme. Lloyd (1904, in Deregowski, 1989) report how his Ugandan audience fled in terror when the first picture in such a slide show happened to be a photograph of an elephant. A brave individual looked behind the projection sheet to see if the animal had a body there.

That said, imbuing photographs with properties of the real can also occur even though a paper picture is used. Thomson (1885, in Deregowski, 1976) describe how Wa-teita women recognised white women in photographs and did not object when they were being led to believe that the pictures were living, but asleep beings.

Deregowski (1968, in Miller, 1973), had adults and children of the Zambian Bisa tribe match photographs of model animals to these models. Familiar and unfamiliar animals were used. As groups, both children and adults passed the test. However, both groups had more difficulties with the unfamiliar animals than the familiar ones. Between groups children were better than adults in matching unfamiliar animals and adults were better than children at matching familiar ones. This alludes to Callaghan (2000) and the potentially detrimental effect of labels on iconicity judgements. However, Miller (1973) argues that the reason for the difference with unfamiliar model animals was that the children were familiar with picture books from a school that had just opened in the community. They were thus more experienced with picture transformations. Adults, on the other hand, had more experience with the real animals and consequently recognised familiar models with more ease than did the children. Whatever the reason for the difference, the fact that familiar animals were overall recognised more than unfamiliar ones support the conclusion that the models were indeed treated as standing for real animals. It was not merely a case of matching any old objects to their photographic counterparts.

That said, in a second study Deregowski (1971, in Miller, 1973) tested another Zambian group of people and found that adults found it easier to match models to models, or photographs to photographs, rather than across medium. There seemed to have been a problem of "translation" from one abstraction to another. It also suggests that local features, rather than recognition on a global level, could have accounted for at least some of the performance. Photographs have more likeness to other photographs, and models have more likeness to other models. If the referents of both the photographs and the models were recognised, though, translation between media should not decrease performance.

3.3 Successive approximation

Sometimes parts of an object in a picture can be recognised while the object as a whole takes considerably longer time, or cannot be identified at all. An example would be to recognise a foot and a tail and then conclude that one is viewing an animal, although one cannot really get to grips with the entity as a whole. Deregowski (1976) describes several episodes of this phenomenon. A case in point is Fraser (1932, in Deregowski, 1976): "She discovered in turn the nose, the mouth, the eye, but where was the other eye? I tried, by turning my profile to explain why she could only see one eye but she hopped round to my other side to point out that I possessed a second eye which the other lacked" (p. 20).

Partial recognition is not only ascribable to an inability to recognise a particular rendering of an animal. For a picture-naïve subject an integrative analysis of a picture can also entail novel *attentional* demands. Thus, attention focused on only isolated features is not uncommon (Deregowski, 1989). As a result the same picture of e.g. a tortoise can be described as a snake, an elephant or a crocodile depending on what parts of the animal one attends to, and which ones one fails to attend to, or integrate in the complete view (Shaw, 1969, in Deregowski, 1989).

Some objects are of course more recognisable than others. Deregowski et al (1972) found, with the subjects who could not recognise pictures until they were printed on cloth instead of paper, that drawings of leopards were more recognisable than were buck antelopes. Recognition of the critical properties displayed by a leopard picture seemed to more easily lead to more complete recognitions.

A slow and stepwise recognition of a motif by picture naïve subjects, e.g. that goes something like: "that is a tail, this is a foot, that is a leg joint, those are horns… it is a waterbuck," is according to Deregowski (1976) similar to the struggles that for example picture competent medical students have with decoding their first X-ray plates. It is the same phenomenon. In one case picture-naïve subjects struggle with interpreting "simple" pictures, in the other picture-experienced subjects face the same problems with "complex" pictures. Note, though, that these examples do not imply a process where a picture is recognised solely by piecing together constituent parts in a manner reminiscent of e.g. Biederman's (1987) theory of object recognition by piecing together "geons." "Tails, feet and leg-joints" are already recognised on the level of identifiable entities.

The most parsimonious explanation for the process is rather the one described by Gregory (1973, in Deregowski, 1976) whereby the perception of a picture occurs in a series of "hypothesis."[23] A set of properties in the picture is the basis for a hypothesis which is then verified against further properties of the same picture. If necessary,

---

[23] Gregory attributes this process to visual perception at large.

the hypothesis might be modified and retested against the features until a stable identification has settled. One can say that parts and wholes define each other with continuous feedback.[24]

If there is a chicken and egg situation here I would suggest that some kind of whole is the first link in the chain, e.g. a set of properties that are perceived as a set, perhaps due to gestalt laws, stimulus salience, or in virtue of forming a prototypical display. However, this first whole is potentially very different from the whole that will be the outcome of the recognition process. Since picture interpretation is a constructive process each picture is unique in the way it interacts with the perceptual processes of the viewer. Some pictures might give only a very small "whole" to start off the process with, say an eye or some other feature with high saliency from everyday life. On other times a more encompassing but ill-defined whole can catch one's attention. For example, one of the earlier recognitions that started the process of identifying the waterbuck might *not* have been a discrete element, like its tail, but that it was some sort of animal. Only after this recognition, or perceptual hypothesis, could recognition of a tail, feet and horns occur. This in turn, in their new configuration as a whole, led to the recognition of the animal as a waterbuck. Experience speeds up this process of "successive approximation" (Deregowski, 1976). Experienced picture viewers, like medical students, would therefore recognise the waterbuck instantly (but perhaps be unable to name it because they are waterbuck-naïve).

That parts and wholes define each other is what Sonesson (e.g. in press a) calls resemanticisation. It explains why attention to a new detail can change the recognition of another. Deregowski (1976) gives an example of this when he found that a window in a drawing was interpreted by his subjects as a four-gallon tin on the head of a woman. This occurred because the subjects did not pay attention to a particular shadow that defined a crucial corner that turned the picture into an indoor scene.

Successive approximation, or resemanticisation, is also the reason that we can perceive, by iconic means, novel pictures that have very little in common with real-life experiences of the world. The *combination* of features makes individual features meaningful, and these in turn feedback to the whole. Without this constructive process pictures that are not possible to interpret in a reality mode would fall flat.

In fact, the real world would likewise fall flat. The reason that we apply successive approximation to pictures in the first place is that "pictures are not unique in being ambiguous and incomplete" (Hochberg, 1980, p. 59). It seems to be true also for objects in the real world. At each momentary glance only parts of an object is informative to our brains. Identifying an object is thus a question of using attention electively to complete the picture, so to speak. "Elective use" means that eye and head movements are not random, but are dependent on the viewer's "perceptual purpose" (Hochberg, 1980). This process will make us perceive that which is most probable in comparison to our expectations, in relation to the stimulus patterns that we attend to. If we expect to see an array of lines and colours we will consequently not see

---

[24] For a neurological perspective on parallel but interacting top-down and bottom-up processes in visual attention, see e.g. Corbetta and Shulman (2002). Also Bar (2004) reviews findings on the interaction of parts and wholes in object recognition, but in terms of "features" and "context."

a waterbuck. What one needs to have in order to identify the content of (non-realistic) pictures is thus the intent to identify objects in a picture, and successive approximation will do the job[25].

Successive approximation is usually a subconscious act, but when confronted with exceptionally ambiguous scenery the process slows down and can be experienced, at least the later parts of the approximation chain. If nothing else one notices that one stares longer than usual at a particular entity. Everyone that has tried to identify a shape in the dark has probably experienced an extraordinary effect of this process; that of switching between complete object identities as new information re-defines the previous. What e.g. is first seen as an animal, perhaps complete with movement, suddenly turns into a dead branch in front of one's eyes as one "takes a second look," or gets closer. In an instant animate movement is redefined as wind movement, or shadow play. In addition it will be difficult to go back to a state of perceptual "limbo" after recognition has settled.

An episode like this was used by Köhler (1925/1957, see pp. 274-5), not a constructivist but a gestalt psychologist, to explain, already in 1921, why his chimpanzees reacted to stuffed toys, facial masks, mirror images and photographs, as to their referents. The chimpanzees were not quite sure of what they experienced and were therefore likely to perceive it as that which it was most similar to. In virtue of containing overlapping information, one object can take over the identity of the other.

Importantly, elective strategies are also required to attend to the *differences* between a depicted scene and a real scene (Hochberg, 1980). Nothing, except one's nervous system, forces one to attend to anything. But again, the nervous system does not do this randomly, but according to where relevant information is likely to be found. If one (or one's brain) does not have a theory about the e.g. realness of what one is viewing, one would not attend to cues that give off the required information to confirm or reject that theory. Without a "perceptual purpose" in this direction, picture specific cues, such as flatness, do not have any relevance in one's identification of what one is looking at. The reversed case is also possible, i.e. that *too many* difference-cues are attended to because content recognition was never expected in the first place. This adds to the probable occurrence of prominence effects. When one tries to make sense of a new object, i.e. the picture medium, one is working with very different theories than those required to decode the actual pictorial elements of the same medium. Consequently attention will single out salient properties differently. Both when differences are under and over attended can they be said to result in picture blindness, or reality and surface mode processing respectively.[26]

Maintaining the view of picture processing as both a direct and a constructive process Deregowski (1989) describes picture processing, on the one hand, as the extension of three-dimensional spatial experience from the real world into the pictorial

---

[25] Contemporary support for top-down processes in attention to visual stimuli on the level of for example eye saccades can be found in the works of e.g. Theeuwes and colleagues (e.g. Van der Stigchel et al., 2006).

[26] Under or over attention to differences between depicted and real material is of course not the sole cause of non-pictorial modes.

realm, and on the other, the application of picture-specific experience that does not have anything to do with everyday spatial principles. This latter area he calls "representational skills." Different people, in Deregowski's research defined by different cultures, combine these two areas of experience differently. On one extreme are people who have three-dimensional spatial skills but who cannot attribute these to pictures. These are not sensitive to *any* pictorial phenomena, not even low-level illusions. Then there are those who can extend real life spatial experience also to picture surfaces, but without seeing them as representations. They are sensitive to some of the optic principles derived from the real world also when they appear on e.g. a piece of paper. (Here we would find reality mode perception.) They might also realise that a picture is a picture, but their lack of experience with pictures makes their ability to make sense of what they see limited and highly variable. (Now we have switched to a pictorial mode competence.) When further tipping the balance towards "representational skills," people will start to add conventional experience to their picture processing. Such people thus display spatial skills derived from perception of the real world, and also skills that have been learned from other pictures. This would be where we would find most readers of this text. Lastly, at the representational extreme, are people that display only learned recognition. They can see that a stick-man represents a human being, but only because they have learned this from other pictures of stick-men.[27]

The complex dynamics of this model describes why cross-cultural data is inconclusive. Different pictures and different tasks require different combinations of numerous spatial and representational skills. However, Deregowski seems more concerned with *what* subjects perceive than *how* (Caron-Pargue, 1989) or indeed *why*. The surface, reality, and pictorial-mode framework, on the other hand, takes into consideration that the way a picture is approached in the first place is very much responsible for how it can be interpreted, and that this in no way is fixed within the individual but can vary across contexts.

Deregowski (1989) makes another important point: "Pictures should not be regarded as forming a unified category in which individual instances differ merely in the quality and quantity of the monocular cues; rather there exist two distinct kinds of pictures. One kind is responsible for [inferred three-dimensional] perception and includes such forms as stick figures; the other is responsible for [direct three-dimensional] perception and includes figures that are immediately seen as three dimensional. The two kinds of representation seem to involve different processes" (p. 73). As cultural products, the first type attempts to *describe* nature, the second to *imitate* nature. Most pictures blend the two characteristics, Deregowski adds.

This division is reminiscent of Sonesson's (e.g. in press a; 1989) notion of secondary and primary iconicity, as well as the idea of pictures simultaneously comprising degrees of iconicity and conventionality. Let us now turn to semiotics in order to refine our notions of picture, iconicity, content, and referents.

---

[27] People at the two extremes are probably only hypothetical ones.

Chapter 4

The semiotic picture

The view of pictures as cultural artefacts is ever present. For example Ittelson (1996) has been concerned with the fact that marks on a surface at all can be meaningful to humans. Without invoking semiotic theory, he attributes this state of affairs exclusively to appreciation of the communicative intentions of the people who place marks. Reference is in his view the specification of intention, of which there can be several for any given picture, or collection of markings. The problem with this view is that inferring intention implies inferring a sender of the message. However, the private aspects of picture interpretation is incompatible with such a definite stance. While the picture as a cultural artefact might be intimately linked with communication, the picture as a vehicle of iconic meaning is not necessarily that. "Picture-attention" can be grabbed by the striking resemblance to an external entity in the markings on a surface, or by one's expectation to see an arrangement in the marks on a particular surface, but also by a more purposeful internal command to invoke a picture in less pictorial mediums, like in looking for figures in the clouds. In this sense the picture is not a cultural artefact but one of imagination. There must be a way to describe pictures without invoking human socio-cognitive factors as the crucial ingredient. Pictorial semiotics is one such way.

Semiotics is often described as the science and study of *meaning*, and more specifically the study of *signs*. Sonesson (e.g. in press a) describes the very point of semiotics to be to "continuously relate the kind of signs we are investigating to all other kinds of signs." Its purpose is thus to say something general and law-bound about meaning creation and mediation. To fulfil this aim semioticians recruit methods and findings from other disciplines as well as developing their own ways of analysing cultural and biological phenomena. Historically the focus has been on texts, but since the 1960s, starting with the analysis of visual rhetoric in advertisements, also pictures have been studied within semiotic frameworks.

In the field of cognitive science, Deacon's *The Symbolic Species* (1997) is perhaps the most well-known explicitly semiotic work. In this he relates classic semiotic concepts to neuroscientific and primatological research. Though overlapping in terminology, Deacon's semiotics differs markedly from that of Sonesson (see 2003a). I will subscribe more to the uses of the latter in this text since Sonesson makes several important distinctions. First of all he clearly separates sign function from symbolicity. This is an overlooked difference in many contexts, not the least in human and animal psychology. The word “symbol,” or “symbolic,” is used as soon something stands for something else. This is clearly different from the use in semiotics, as fuelled by the works of especially Charles Sanders Peirce, and manifested in the pictorial semiotics of for example Sonesson.

In a Peircian framework[28] a symbol is only a special case of sign. There are others, which are just as “representational,” such as icons and indices. These differ from symbols in important regards, but are still signs. A further useful discrimination, in especially Sonesson’s work, is the separation of principles from the signs that depend on those principles. An icon is for example a sign that predominantly owes its meaning to the principle of iconicity, or similarity. An index is evoked by the principle of indexicality, i.e. nearness. Lastly, a symbol is based on the principle of symbolicity, which is really conventionality. Often the defining character of a symbol is attributed to arbitrariness, but this is only a common effect of a conventionally induced meaning.[29]

The separation of signs from their principles is necessary because all three principles, i.e. iconicity, indexicality, and conventionality, can combine in meaning creation. A relevant example for this text is that there can be a fair amount of conventionality in an iconic sign, i.e. in many pictures.

A separation of the principles and the sign relation is necessary for a second reason. A sign is only one kind of meaning. Iconicity, indexicality, and conventionality contribute to other meanings that are not necessarily signs. Stimulus generalisation can for example be described as an iconic process: A second entity inherits properties from the first one because they are alike. Indexical processes are often involved in reinforcement learning. Perceived temporal or spatial connectedness between a reward and its contingency strengthens the bond between these two, as opposed to something more removed in time and space. Conventionality, on some level, is involved for example when animals agree on a joint activity. Play behaviour is for example imbued with agreements. I say “some level,” because attempts have been made to specify types, or degrees, of conventionality. If the animals for example are aware of the fact that they are involved in an agreed upon practice, it would have been a case of “full conventionality” (e.g. Zlatev et al., 2005), characterised by normativity (Zlatev, 2007). Full conventionality is required for systems of symbol use, i.e. language.

The three principles can interact in complex ways and can be described in terms of relative impact. That is, sometimes an e.g. iconic impact is low; sometimes it is very strong, and so forth. It is also possible to create complex taxonomies of kinds of iconicities, indexicalities, and conventions, but that is not necessary for my purposes here.

---

[28] Filtered through my understanding of Sonesson (e.g. 1989).

[29] If following a Peircian terminology “principle” is really reserved for iconicity (e.g. Sonesson, in press b), but that distinction is not necessary here.

4.1 The sign

When the principles of iconicity, indexicality, and conventionality are "encapsulated" in an expression, be it a gesture, a word, marks on a surface, etc., we can start to look for a sign relation. A sign is traditionally made up of two parts, usually known as expression and content, derived from Ferdinand de Saussure's "signifiant" and "signifié," the signifier and the signified (see e.g. Sonesson, 1989). In verbal language the sounds constituting the word is the expression, and what the word "means" can be said to correspond to the content. For it to be considered a sign these two parts need to be connected through at least one of the principles, but must also be understood as separated from each other. A word is typically "connected" to its meaning through convention and experience, i.e. a repeated co-occurrence. (Note the role of indexicality in this process.)

Within the sign, the expression is more directly connected to the senses, in virtue of being physical, but the focus of the perceiving system is on the content (Sonesson, 1989). This definition stems from phenomenology and contrasts signs to two other types of pairings, that of two co-present items which are grouped for one reason or the other, and that of two separate items but where the second one is only indirectly given through the focus on the first. In the sign relation the focus, or "theme," is rather the second item, the content. If one instead focuses on the expression, the sign function fails.

In accordance with the observations of Jean Piaget the crucial differentiation between expression and content necessarily takes place from the point of view of the subject and typically involves differentiation in time and/or space, as well as between the respective natures of the signifier and the signified (e.g. Sonesson, 1992).

Differentiation is an integral part of Jean Piaget's semiotic function.[30] This function is typically seen as a general capacity that develops in children between 1.5 and 2 years of age. It enables the child to draw representationally, to pretend, to use language, use mental imagery, to imitate from memory (deferred imitation), etc. The commonality between these abilities is to represent reality by means of a signifier that is separated, from the point of view of the subject, from that which is signified (e.g. Sonesson, 1992). However, that there is a commonality between these competencies does not imply that they necessarily have similar developmental trajectories or indeed sources (see e.g. Gardner & Wolf, 1983, in Sonesson, 1992). There does not seem to be a semiotic seed that bides its time and then bursts in all directions, looking for ways to be expressed. The notion of a semiotic function rather serves a descriptive role than explains competencies. However, subscribers to a central semiotic function are naturally also to be found. For my purposes here I will only be concerned with Piaget's contribution with the notion of differentiation.

Sonesson (1992) distils Piaget's ideas about differentiation into two categories. First is the recognition that the signifier and the signified pertain to two very different areas of the world of experiences, and secondly that these two areas cannot go over into each other in time or space. Differentiation is thus of a double type in that

---

[30] It was called the symbolic function in his earlier writings.

expression and content does not cross each other in time and/or space, and they are perceived to be of different natures (Sonesson, in press a). The nature of a picture, as a physical object, and that which it depicts is just very different, especially if you have a developed concept of "picture."

In terms of knowledge, or expectations, differentiation in the iconic sign relation can thus be many things, which is the reason I have used the term "sufficiently differentiated" in the definition of pictorial mode in section 1.4. Differentiation can not only be several things, but different for different media. For example in photographs, knowledge of displacement is a central cue for differentiation: the point of view that there is a space and/or time difference between the occurrence of the referent and its photographic counterpart. For constructed pictures the plain knowledge that they are constructed can be something that helps with differentiation. Their visual deviation from reality is another cue, as is the flatness of most pictorial displays, etc. All that one learns about pictures potentially helps with differentiation. However, one can only learn about pictures, as a category, due to differentiation in the first place. But this initial differentiation might not be enough. Sufficient, or proper, differentiation is a necessity for reference to be possible.

Because differentiation is not an all or none thing, pictures, even when seen as pictures, can contain properties of the real. This is for example evident in children who perform "realist errors" (e.g. Thomas et al., 1999; "false photographs": Zaitchik, 1990), which is when children believe that a referent can continue to affect a picture after creation of the picture. If the referent updates, so does the picture, they reason. There is enough differentiation to allow a sign function, but not enough to preclude misconceptions about the relation between pictures and their referents.

As adults we are not immune to degrees of differentiation. Attempts have been made to pinpoint what it is that determine, for any given image, whether adults reason about the referent of the picture, or the picture *as* the referent (Schwartz, 1995). Realism versus schematics has proven to be a promising candidate. This is likely due to retained properties of the real in the latter case.

To experience a retained reality perspective on pictures, despite fully fledged referential competence, try to tear up a photograph of a loved one and note your feelings. Or if you are afraid of snakes you probably know already in which books not to browse. A picture can very well have properties of the real when differentiation falters. This suggests that differentiation is partly connected to inhibitory functions. For some pictures, or content, we are good at inhibiting our reality responses, most likely not even noticing that they are there, while others break the dams, so to speak. The opposite case is of course also possible. People can turn the very real into non-reality, as when watching a gruesome splatter movie and react with amusement instead of disgust. So while parts of our systems treat pictorial stimuli as real, other parts "know" what pictures are and are not. In e.g. Slater et al. (2006) the reality part got the upper hand when subjects reacted with subjective, behavioural and physiological stress when having to administer electric shocks to an animated figure in a virtual reality setting. Subjects that did not have to see the animation but interacted with it through a text based interface did not experience such stress. The former group reacted *despite* full knowledge of the artificiality of the situation.[31]

In order to explain why some animals and children act towards pictures as if they were real, even though they can separate expression and content per se, i.e. discriminate pictures from reality, we must invoke a third entity. There seems to be a second necessary differentiation, that between the sign (i.e. expression and content), and that entity in the real or imagined world that appears as content within the sign; i.e. the referent. The referent is present in the sign, as content, but is not really there. It is the content which is perceived as actually "being there." If this experience is confused with the referent, the sign appreciation has also failed.

At first glance it might seem odd to distinguish between content and referent. After all, most of the time we do not seem to bother to connect a picture to a specific object in the outside world. However, that there is a need to separate content from referent is perhaps more apparent when we look at words, signs that are grounded on conventionality. "Fox" and "räv" (Swedish) are different expressions that result in the same content. But let us say that "fox" is exchanged for "vixen," where Swedish does not have a separate word for a female fox. Then, for the English speaker, content would change with the expression used. However, the English speaker and the Swedish speaker can still talk about the very same vixen/räv (fox), i.e. referent. Gender is a property of the referent, but is only a part of the content, given the expression, in one of the two languages. Similarly for the iconic sign, all drawings of foxes, although very different, captures the content "fox" (otherwise they would for example be dog drawings), but not all drawings of foxes can necessarily refer to the same referent fox.[32]

But there are variations in expression that do indeed affect the content and consequently the sign's possible referent. A fox can be drawn in a way that someone interprets as a specific fox, perhaps a dog, or cannot recognise as a depiction at all. The properties of the expression which is necessary for designation of a content is called *form* in the semiotic tradition of Ferdinand de Sausssure, and that which is redundant is named *substance*. Form and substance exist on both the expression and the content side of the sign. They are separated by what is called the principle of relevance, which defines which properties of the expression that are crucial for a certain content to be expressed, and also which properties of the content that can vary with expression. Putting a horn in the forehead of a horse changes the content to a unicorn. A horn in the forehead is thus form. For example the length, within certain boundaries, of this horn is substance. A unicorn with its head in the bushes can turn back into a horse, given that we did not know anything about the context that hinted that this was in fact a unicorn with its head in the bushes. Such knowledge is on the level of the referent. The principle of relevance is thus dynamic and sensitive to the dictations of reference. The same feature can be form or substance in different contexts.

---

[31] The social pressure stemming from "being bad" in a scientific experiment was arguably equivalent between the two conditions.

[32] It is clear from this example that content is closely connected to what is sometimes called categories in cognitive science.

Reference is indeed context. The very same circle (content) made out of lines (expression) on a surface, can be many things (referent), depending on context. It can be a table from above, a ball, a hole, the letter O, etc. Context can take the form of additional elements, such as a plate and fork on the table, or a shadow under the ball. Context can furthermore be the location where a marked surface appears, such as being in a written book, in a picture book, on a road sign, and so forth. Context can also be what we knew about the specific circle from before, for example that it was a table in the picture we saw on the previous page in a picture book, or what people tell us it is through its appearance in communication. It is no wonder that Peirce called the establishing of reference "interpretant," and that there are several kinds, which can be internal or external to the mind (Deacon, 1997). To conclude, reference can perhaps be seen as the specification of content due to context.

In an experimental setting, which we will see many examples of later in this text, the task at hand potentially set up context differently. This can have crucial impact on the perceived content of a picture. While a picture might be recognisable in e.g. a matching task, it may remain non-identifiable in e.g. a free response task.

Different types of signs designate their referents in different ways. A word, which has an arbitrary relationship between expression and content, relies heavily on context that is external to the sign as such. A picture, on the other hand, in virtue of being an iconic sign, can often specify referents with the context in the sign itself. There is a type of overlap between referent and content in an iconic sign that cannot be found in indexical or conventional signs. Consequently there are interpretive mistakes that are specific for pictures, such as mistaking picture for referent and act out on it. One seldom acts out on words, as sound waves or ink on paper, and when getting angry at the squirrel that steals food from the bird table one does not throw stones at the paw prints it has left in the snow.

The trinity of expression, content and referent is closely connected to the notions of surface, reality, and pictorial mode of picture processing. One could say that the focus of attention in surface mode is on the expression side of the potential sign, which is the reason recognition fails. In reality mode attention is caught in the content of the picture, but never moves beyond this. In a pictorial mode, attention is not only on the content, but also on the referent.[33]

4.2 Iconicity

According to Peirce, as described in the works of Sonesson (e.g. 1989), the principle connecting two things in an icon, i.e. similarity, shall exist independently of the sign relation, and the properties that are similar must be extant independently in the two things. That is, e.g. a circle in my picture shall be there independently of me viewing it as the shape of an apple.

The iconic relation between a picture and that which it depicts is seldom one of only isolated features, such as a shared colour. As perception at large, perception of

---

[33] Although the mapping is appealing, note that it is not just a question of distributing attention, as if it were a spatial issue. Attention itself is likely to work somewhat differently in the three modes.

pictures deals with parts and wholes and therefore relates to the combination of features. The combination of features, i.e. their presence together with other features, their spatial relation to these other features, etc., are also basis for similarity between iconic pictures and their real or imagined referents. Iconicity is furthermore not limited to concrete visual similarities, but more abstract ones can enter the relationship. An arrow can for example be iconic with direction. However, such relations become apparent only after a sign relation has been established.

In the photograph, the relationship between content and referent is very direct. The photographic picture is a real view, indexically acquired, of the referent. However, the expression of a real scene and a photographic scene can never be identical. Light does not bounce off flat, pigmented surfaces the same way that it bounces of real objects. The image that reaches the viewer's vision will be different, at least objectively. Colour, resolution, sharpness, etc., differs from reality in most photographs. However, a difference from hand-made pictures is that such deviations are usually global, meaning that it affects the whole image equally. Such constant deviation should not confuse the system more than looking at the world through a somewhat distorting window. Potentially more problematic, then, is the flatness of photographs, their unnatural motion parallax, reduction in size, and static display. In order to understand the photograph as a snap shot it is necessary to extend its content to time and space outside of the picture. Without such abduction making sense of anything but static objects may be limited. The realism of photographs shall thus not be taken for granted.

Although I sometimes use the term "degree" of iconicity in this text, I do not mean to suggest that iconicity is a physical thing that there can be little or a lot of. The notion of degree rather alludes to the subjective appreciation of something being very clearly pointed out by the picture. Since iconicity is subjective and dependent on the interaction of the interpretative system of the viewer with the "stuff on a surface" one cannot objectively count iconic relations across pictures. In one picture a shadow can be necessary for recognition, i.e. *form* in the Saussurean sense, and in the next the very same type of shadow can be completely redundant, i.e. a case of *substance*. With pictorial experience a former necessary property can become redundant and excluded from the picture, but if recognition is the same one cannot say that the picture has turned less iconic. Reduction does not equal abstraction, and abstract is not the opposite of iconic. The term used by Köhler (1925/1957), "nearness to life," or "realism," is perhaps a more suitable term when one shall describe whether a picture is a truthful rendering or an abstraction.

Another possible dimension for judging "degree" of iconicity is to look at the "exclusiveness" of iconicity, i.e. the relative importance of iconicity among indexicalities and conventionalities for construing a sign. Interestingly, also very realistic icons demand conventional components to function as a sign. One example is that we are dependent on knowing that objects in a shop window stand for the objects that one can buy in that shop. The similarity between display objects and the objects that one can buy is very high, to the point of identity (e.g. Sonesson, in press a; b), but this similarity has very little impact on the sign function of the objects. Drawings in the shop window might have served the purpose equally well. In this perspective the objects in the shop window are iconic conventional signs. The question, then, is when iconicity is the prime principle in a sign relation.

4.3 The picture sign

Sonesson (1989) has convincingly argued for the picture being a sign, en par with language symbols in terms of referential potential. Often pictures are predominantly iconic signs, but this is far from true for all objects that we call pictures. Sonesson reserves the term pictures to icons based on primary iconicity. He discriminates between two types of iconic signs, depending on whether likeness precedes the sign relation, and even is the reason for appreciating it, or whether likeness is only discovered as a result of an appreciated sign relation. The first type of likeness has been called primary iconicity, and the second type has been dubbed secondary iconicity.

The main argument for the existence of primary iconic signs is that a naïve viewer is able to decode such signs without instruction, and that cross-cultural and child data supports this case (e.g. Sonesson, in press b). However, the data for this argument is inconclusive. The Hochberg and Brooks' (1962) child study is just not enough, and cross-cultural data paints a mixed picture. Recognition of non-photographic material in picture naïve subjects has only been demonstrated after instruction, or in a setting that scaffolds the subject's view of the stimuli as informative. There are several instances in the literature of complete failure to recognise depictions, even photographs. If instruction or scaffolding is necessary the presence of secondary iconicity cannot be precluded. The extant data rather reserve primary iconicity to photographs, if even that. However, it should be said that the subspecies of primary and secondary iconicities in turn can be numerous (Sonesson, in press a), and if this is indeed the case one might expect that some of them are not easily ascribed to a pre or post sign existence.

However, rather than focusing on the possible interactions between primary and secondary iconicities in pictures, Sonesson (e.g. 1989) argues that there is a hierarchy in the real world of suitable and unsuitable mediums for signs, which accounts for the problems of recognising pictures as primary icons. That which is more prominent always serves as comparison to that which is less prominent, not the other way around. The reason picture-naïve people need guidance is because they struggle with this hierarchy. If they only get to grips with the surface of a picture, iconic interpretation will follow naturally. Beyond this point iconicity will more easily precede the sign relation and more types of pictures than before can qualify as primary iconic signs.

The process that is here proposed to account for the fact that iconicity can be appreciated before the sign function, even in non-realistic pictures, is successive approximation (section 3.3), or resemanticisation (Sonesson, 1989) (see *fig. 6*).[34] This implies that parts and wholes define each other in continuous feedback. Correspondence, and thereby recognition, occurs on the level of relationships rather than discrete features. An illustrative case in point is the caricature face drawing where indi-

---

[34] Sonesson himself would perhaps rather invoke the ecological optics of Gibson (e.g. 1979).

vidual features can be grossly distorted, but their interrelations still allow effortless and swift recognition (e.g. Perkins & Hagen, 1980). Following recognition, wholes can continue to induce expectations on the parts. This allows us to imagine with precision where left out parts should have appeared in the picture. The potential of conveying more information than goes into their production is a unique feature of icons (Sonesson, in press a).

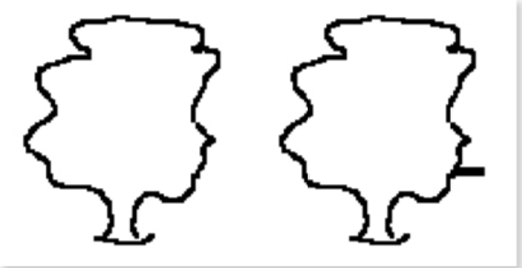


Figure 6. Gregory's tree/smoking lady. A case of resemanticisation where nose, chin, hair etc. is pointed out by the simple addition of a line in a critical position. The facial features in turn define the line as a cigarette. From Sonesson (1989).

Resemanticisation is of course even more powerful in a post-sign situation, where attention is constrained by other expectations. As mentioned above, some similarities can only be discovered *after* the sign relation has been established. Such a picture is more of a conventional sign, where iconicity predominantly exists in the mind of the observer, rather than extant in the expression. The source for similarity is in the expression, but the information is just too scant, or too ambivalent, and no help comes from context within the picture itself. However, many such pictures, perhaps especially ambivalent ones, are not approached first as a conventional sign, but as an ordinary iconic sign. Then, after a clue has been found that hints at another way of seeing it, the picture can turn into a secondary iconic sign. (This in turn could perhaps be the new primary iconic interpretation next time one sees it.) Since iconicity is not an objective state of affairs but relies on interpretation, a particular picture can have more than one nature over time. The content of the picture, however, can only be one at every given moment. The viewer must therefore settle for one view or swap between them. Sonesson (in press a) agrees that a sign can mix primary and secondary iconicity.

The discrimination between primary and secondary iconicity is important because it entails two very different types of causal relationship between perceived similarity and the sign relation. However, as suggested, this distinction does not translate well to two types of *pictures*. It would mean that the same physical object is a picture only in one of two possible views by the same pictorially competent subject. One could perhaps say that the picture is an objective picture in both cases, but a subjective picture in only one of them. However, since pictorial *competence* entails perception of both primary and secondary iconicity, we do not need to be quite as categorical. Perceiving secondary iconicity is in fact the most telling evidence for a perception of (objective) pictures in a pictorial mode, since it depends on the sign function being present. This is perhaps the best way to preclude processing in reality mode and is the reason for why I have used drawings as the critical stimuli in my empirical work in Chapter 13.

Now that we know what pictures and iconicity are, and are not, how children and picture-naïve cultures interact with them, and have a framework of modes in which pictures seem to be processed, let us turn to the nonhuman primates.

II

The program we have described avoids the question of whether an animal other than man can acquire language. As comparative psychologists we must reject this question. It is like the question of whether an animal other than man can have thoughts. It depends on the definition of language rather than on the observations of what animals do.

Allen Gardner and Beatrix Gardner (1971, p. 181)

Chapter 5

Primates in picture tests

Given what we have learnt from the empirical work with humans presented in Part I, we can assume with some certainty that pictorial competence in humans is a developmental matter, including both cognitive and cultural growth. It is then in its place to reflect upon what one can expect to find from scrutinising the use of pictures in animal research. We assume that animals do not share our pictorial culture, so what use is it to study this in animals? There are at least three areas, all connected to humans, that makes the endeavour worthwhile.

One obvious reason is to be able to say something about the onset of iconic competence in human development as a species. What are its requirements and early expressions? The comparative approach has always been a popular way to recreate human prehistory. This thesis will not be that ambitious, however, but I am sure that interested readers will be able to find useful information in this text for those kinds of inferences.[35]

An area where human and animal lives intercept, which involves pictures, is in experimental settings. Pictures are used as convenient stimuli in place of real objects, and far-reaching conclusions are drawn from such research. All three modes of viewing pictures have different implications for how the results are best interpreted.

The claim made further above, that pictures are exclusive to human culture, is not really true. Pictures are also integrated as part of some animals' everyday life, for example in ape language research. It is worth studying if these individuals perform differently with pictures than those where pictures are less integrated, say at an anthropoid station in Tenerife. We can in this manner investigate the effects of language and culture on pictorial competence as such. But we can also investigate concepts and imagination in new ways, by looking at categorisation and interpretation of pictorial material. This is applicable also to language-naïve subjects, if the step to a pictorial mode can be made. In this way, working with pictures can replace linguistic symbols as a window to the ape mind.

To what extent nonhumans are able to replicate human development depends on what similarities are there from the start, what is open to learning, and what is constrained by species differences. However, the contrastive approach is not the only one. A second one, that I will tend to favour in the remainder of this thesis, is to

---

[35] A recent attempt to relate iconicity to cognitive evolution by the present author and colleagues can be found in Zlatev et al. (2006).

investigate what nonhuman primates are actually able to do with pictures, irrespectively of the humanness of those behaviours. The three modes of picture competence, i.e. surface, reality, and pictorial modes, can stand on their own as a framework, regardless of typical human performance.

Pictures have indeed been used extensively as stimuli since the early days of animal cognition research, both as substitute for real-life objects, individuals, and events, and as abstract stimuli. They have for example been used to assess questions of spatial representation, social cognition, viewpoint consistency, and serial list learning (Fagot et al., 2000), as well as all sorts of discrimination and categorisation (Bovet & Vauclair, 2000). Pictorial stimuli seem to have been used with such success that few scientists have paused and asked the questions of why it works, and what it means. Since animals of all kinds readily accept pictures, in the right circumstances, as examples of real-world objects without fuss, it must mean that pictures are simple and intuitive phenomena.

This observation easily leads to the idea that there is only one way of viewing pictures: you either see what is in them, or you do not. The most common mistake in the literature is thus to lump performance in reality mode as performance in pictorial mode, or neglect to control for surface mode processing. However, it should be noted that we do not know about all those unsuccessful and unpublished attempts at training animals in discriminating pictures. Bovet and Vauclair (2000) could only find ten published papers between 1953 and 1998 that demonstrate difficulties for animals in recognising pictures. Most certainly a vast number of pictures have been dropped from experiments because subjects have had difficulties decoding them. And many subjects might have been dropped too.

Bovet and Vauclair (2000) cite more negative examples with birds than with primates. Photographs and video images are made for human (i.e. primate) vision, while birds have a very different visual system (e.g. Delius et al., 2000). It might also be the case that primates can conceptualise pictorial stimuli in a way that birds cannot. Bovet & Vauclair (2000) concluded that "[...] picture recognition in animals is not obvious and is dependent on experimental factors" (p. 158). What those factors might be will hopefully be made clear in the chapters to follow.

There is a serious problem related to the ease by which we grant animals a complete pictorial competence by looking at their behavioural performance only, and disregard the underlying processes. If you do not ascribe referential abilities to animals, the option is to explain their competence with pictures based on a complete correspondence between pictorial stimuli and real stimuli. Experiments that substitute real objects or individuals for photographs, and then draw conclusions about the everyday functioning of the subjects' perception or thinking, do exactly that. Conclusions based on such experiments can be misguided if it turns out that animals perceive real objects and depicted ones differently. But not only should we avoid presuming that subjects lack a picture concept of some sort, we must also avoid the opposite. When we use pictures and presume that animals *must* appreciate that there is a difference, maybe none is perceived, at least not of the kind that we expect. As I will continue to suggest in the present part of the thesis, reality mode can be quite broad and can encompass certain degrees of "magical thinking"[36] and unusual instances of "reality." After all, an object in a picture can differ markedly from ordinary reality in appearance and behaviour. But often animals still seem able to recognise such content, while at the same time not being able to use pictures in referential tasks. They even act out on pictorial displays that should, from an objective point of view, facilitate differentiation.

In the coming chapters I will review primate (and some pigeon) experiments that use iconic stimuli (pictures, replicas, scale-models, video, and mirrors) for one purpose or the other. I will also review some observations and anecdotal evidence from the literature. Such data is useful for painting a picture of the potential and variety of behaviours with pictures that can be expected from apes. But let me start, in this chapter, with a closer look at those few experiments that have been aimed directly at picture comprehension, regardless of which views one have had on the ability at the time.

5.1 Viki

In the first part of this text Köhler's (1925) experiments with chimpanzees and photographs were presented. What we found was that the chimpanzees could perform all right in the tests but that they showed several behaviours that hinted to the fact that they seemed to have processed the stimuli in reality mode, and not in a pictorial mode. The overall conclusion of Köhler, that the "nearer to life" a photograph is, the better a chimpanzee performs with it, is predicted by reality mode processing.

Besides Köhler, Kellogg and Kellogg (1933/1967) are mentioned in Hayes and Hayes (1953) as the first to have published information on the use of pictorial stimuli with chimpanzees, in their case by 15 months old Gua, who are said to have been able to point out drawings of a dog and a shoe (see section 8.1).

Thus, almost half a century had passed between Köhler's experiments and the second direct test of picture comprehension in apes, which is reported in Hayes and Hayes (1953). Their chimpanzee Viki was 5 years old at the time of the main study. She was raised in the home of psychologist Keith Hayes and wife Catherine Hayes with the purpose of seeing what happens when an infant chimpanzee grows up in a human social and material environment. It was an extension of the Kellogg project, which had lasted only 9 months. Of special interest was if Viki would learn to produce speech by mere exposure. Viki was only a few days old when C. Hayes started to take care of her at Yerkes Laboratories of Primate Biology, and by the age of 6 weeks she was incorporated in the Hayes' household.

Viki had an early experience with pictures, since she was raised as a human child, but she was never specifically trained to perceive pictures. Viki enjoyed browsing in picture books on her own from the age of 6 months (Hayes, 1951). Later, at 9 months, she started to spend more time in the lap of C. Hayes, including looking in picture books. However, picture books did not become a favourite toy until the age

---

[36] In this text "magical thinking" is not meant to be taken as naïve or contrived thinking. It is just the most parsimonious explanation from a different perspective.

of 1.5 years, when she also started to spontaneously respond to pictures differentially. Details on these behaviours are not given in Hayes (1951). Surprisingly, her favourite book was not to become a picture book but *This Simian World* by Clarence Day, which contains only ten black and white cartoon drawings by the author (Day, 1920/1941). Thus, contrary to one's assumptions, Viki's interest in browsing through books was necessarily not a pictorial one, and her interest in a specific book might be one of smell, paper feel, format, association to a particularly good event, or something completely different.

When she at age 3 tried to listen to a depicted wrist watch the Hayes judged that she did not mistake it for a real watch since she had not tried to pick it off the page and put it to her ear, as she usually would with a watch. Instead she had bent down to it (Hayes, 1951; Hayes & Hayes, 1953). However, it is not implausible that someone would suspect a picture to give off sounds, especially at a young age.[37]

When 4 years old Viki readily pointed to pictures of e.g. beverages, followed by leading the addressed person to a refrigerator. This suggests that Viki could relate pictures at least in some ways to the real world. Premack (1976) sees communicative behaviours with pictures as the strongest evidence for a true pictorial competence. After all, the ape uses the picture to achieve a goal that pertains to an object external to the picture while at the same time *not* performing the same act on the pictorial version. In principle this criteria would be correct, but we need to know more about the context and generalisability of the behaviour before we can say that Viki knows that pictures are pictures and objects are objects. In terms of learning, Viki could have discovered that she could barter those flat and flimsy *special cases* of drinks for a more drinkable version. If her performance was limited to certain categories of objects there is reason to believe that she had learned specific links between pictures and their objects within those categories, rather than discovering the general nature of pictures and their communicative potential. If this was indeed the case one could say, in terms of picture processing modes, that she used reality mode in a communicative context. In addition, this context could be limited to bartering situations.

This interpretation is supported by the fact that Viki was apt at developing conventions, or habits, once the co-occurrence between two entities had been apparent to her. For example she used to bring diapers from a special storing place in the bathroom to signal that she wanted to go for a car ride, since extra diapers were always brought on such occasions (Hayes & Hayes, 1954). When diapers were not available Viki generalised to bringing tissues from the bathroom instead. In this episode the co-occurrence, an indexicality, developed into a convention since Viki's actions and her caretakers' recognition of their meaning resulted in the fulfilment of Viki's wants. A similar development could have occurred with specific types of pictures, bypassing a general and flexible understanding of pictures as potentially communicative.

Despite Viki's limited use of pictures the Hayes initially invested much hope in it. They believed that pictures could help where words failed, in the same way that they had discovered that relying on Viki's imitative abilities of certain tasks was fruitful in place of verbal instruction. Imitation from pictures would be an especially

---

[37] Three year old humans likewise occasionally attribute depicted content with properties of the real (see e.g. Beilin & Pearlman, 1991).

powerful tool in the future they reckoned. With these musings the book about Viki's first three years, *The Ape in Our House* (Hayes, 1951), ends.

Unfortunately, pictures as language would disappoint after all. Hayes and Nissen[38] (1971) came to report major limitations in Viki's use of pictures for intentional communication. The Hayes had tried to encourage Viki to use pictures for communication by introducing her to a deck of picture cards made from colour magazine illustrations (further nature not specified). When she verbally asked for a drink or a cigarette Viki was asked to clarify her request by indicating a picture from her picture cards. This she could do from trial one with an initial set of three cards, and continued to do without fail. Then a novel picture of a comb was introduced. This threw her completely. After much coaxing she used the picture randomly. In the end only cup and car pictures were used dependently and the attempts to make her state her requests using pictures were dropped after seven months.[39] A promising project start led to a disappointing end. Nevertheless, Viki gladly continued to tear pictures of cars from magazines and trade them for car rides long after the deck of cards had been discarded.

At the time of formal testing of picture perception in Hayes and Hayes (1953) Viki was already familiar to matching procedures. She was shown a picture and was made to choose one of two objects that matched the category in the picture. There is no information regarding the degree of iconicity in these pictures. She was correct on 78% of these choices. It is not said whether individual pictures were shown more than once and no controls were made for matching based on surface features, like colour and form.

At 3.5 years of age Viki was tested on her ability to imitate actions from pictures. The actions to perform were clapping hands, patting one's head, and sticking out the tongue. Hayes and Hayes (1953) report that she did fairly well on stimuli ranging from movies, via black-and-white photographs, to "simple line drawings." However, she had a preference for performing actions that she particularly liked, regardless what was modelled. The same was true for her imitation of real people. Her successful interpretation of line drawings speaks for a pictorial competence. But there is no data on the novelty of the pictures, thus rote learning cannot be ruled out in the present analysis of this particular test.

Besides the performance with line drawings it is noteworthy that imitation of dynamic actions depicted in static pictures requires imaginative interpretation on behalf of the viewer. One must infer what happened just before the static view, and what will happen just after it, in order to read *clapping* and *patting* into the relations of body parts in a picture. This might not be possible when viewing a picture in reality mode. Unfortunately, without a detailed report on the action response in Viki, we cannot know if she read *clapping* or *patting* into the pictures, or just *hands together* and *hand on head.* Viki, aged 4, did for example not learn how to solve problems when the solution was presented in pictures, but she did learn when human models demonstrated the solution in real life (Hayes & Nissen, 1971). Two com-

---

[38] Formerly Mrs. Hayes. Not to be confused with H.W. Nissen.

[39] Car cards had to be taken out of the deck because as soon as she saw them her requests were all about car rides and all other cards were ignored.

parison children of Viki's age succeeded by watching only the pictures. While purely visual, static, information was enough for the human children, Viki apparently needed to see the solution acted out. The ability to read dynamic content into pictures is a possible venue for exploring pictorial competence in future empirical work.

At 5 years of age Viki was given a discrimination task with generalisation trials using pictures of familiar objects like chairs, cars, dogs, flowers etc. Cups in a tray were differentially baited with food, and pictures drawn from two categories of objects were placed on the cups. The simple setup of using food wells covered by choice stimuli is called a Klüver form board, or just Klüver board (Meyer et al., 1965). Viki was rewarded for choosing from the same category throughout a session of about 12 trials. Importantly, a given picture was only used once, precluding rote learning. Reliance on a surface mode could otherwise have been possible. In total she was given 28 sessions, and thus tested on 28 pairs of objects, divided into four successive groups of picture types. In the first 7 sessions "naturalistic" colour pictures were used, and in the following 6 a mix of realistic photographs and stylized drawings. The third group consisted of 7 problems that utilized realistic colour as well as black-and-white photographs, and the last 8 sessions involved only black-and-white line drawings (*fig. 7*). The pictures within an object category were varied not only when it comes to expression, i.e. picture type, but also in content, e.g. chairs of all styles depicted from various angles. The pictures were a mix of illustrations, cut out from magazines, photographs, and handmade line drawings. Many of the drawings were freehand copies of photographs from earlier sessions, and can thus be criticised for lacking in novelty. Viki is reported to have been able to generalise discriminations to novel pictures of all the types mentioned above. Success rate on the four groups of problems were 85% (colour pictures), 75% (mixed), 82% (photographs) and 73% (line drawings) correct respectively. No data is given for the distribution within the mixed group, but the lower figures in this and the line drawing group might be due to the abstractness of the stimuli. In the third group Viki was 95% correct on the colour photographs, but only 68% on the black-and-white ones.

Figure 7. A subset of the line-drawings used in discrimination and matching tasks with chimpanzee Viki, age 5. From Hayes and Hayes (1953).

Following the discrimination work Viki was set to work on a reversed form of matching from the first one she did. Now she was shown an object and was made to choose between two pictures the one that depicted an object from the same category as the shown sample object. Pictures were controlled for size and colour matching. Mostly objects that could be depicted in pictures without too much size transformation were used. Pictures of real cars and animals were matched to stuffed animals and toys. This experiment was also divided into four conditions. In the first one Viki received 37 trials using realistic photographs reused from the discrimination tasks. She was correct on 95% of the trials. In the second condition she received 13 picture pairs from the former group but now the matching pictures had become the non-matches, and vice versa, and new sample objects were used. A 10 second delay between a removal of the sample and the presentation of the matches was also introduced, to hinder Viki from running off and play with the sample objects. She was correct on 85% of these trials. The third group of trials, 25 in total, utilized line drawings, some of which were reused from the discrimination tasks and was thus not novel (see *fig.7,* p. 62). Her performance dropped somewhat to a still good 80%. The last group of 22 trials consisted of rearrangements of the line-drawing pairs from the previous group, so that previous matches and previous non-matches were pitted against each other in the presented choice. Viki was 91% correct.

The last study reported in Hayes and Hayes (1953) is a comparison between how Viki performed on discrimination tasks with pictures versus with nonsense designs. The reason for this comparison is interesting. Discrimination problems are, according to Hayes and Hayes, learned more quickly with three-dimensional objects than with designs on flat surfaces. If Viki saw pictures as motifs she would perform better with them than with totally abstract stimuli, arguably because the latter are purely dependent on visual matching and memory of arbitrary patterns while the former bears also on conceptual resources. It is a nice test of the constraints of a strategy based on a surface mode rather than a reality or pictorial mode. As pictures, "realistic, coloured pictures" [40] were used and as nonsense designs single coloured shapes made with crayon.

There was a significant difference in performance between the two categories, with pictures having a mean error rate of 0.8 and nonsense designs 2.7. No development of the performance with nonsense designs is given so one cannot exclude that some of the difference might be explained by the fact that nonsense designs were novel stimuli, while pictures by now was commonplace. Another competing hypothesis for the difference is that Viki might, if she analyzed abstract pictures in a pictorial mode, have inferred meaning in the abstract pictures, but happened to construct the wrong theories about what she was seeing and thus fail on the trials. However, the most parsimonious explanation is that the superior performance with pictures was due to Viki being able to categorize the objects depicted and not the nonsense designs.

With another type of abstract, but still depicting, stimuli Viki did not show a strong tendency to infer meaning in their lines and shapes. Viki was fond of drawing but never seemed to make depicting pictures (see Chapter 10 for more on drawing), but she learned to connect multiple dots that was put out on a paper in order to create shapes when filled in, just as the popular child's game found in colouring books and magazines. (If the dots were too far away she started to draw on and around them instead.) She never showed any recognition of the shapes she made until, in response to the familiar words "get me one of these," she fetched a stuffed dog after having connected a "rough approximation of a terrier" (Hayes & Nissen, 1971). The only other self-made drawing that elicited an equally successful response was one of a

---

[40] We can thus not separate reality from reference in this case.

cup, which she named (one of the few words she could voice) and fetched. To other self-made drawings she was indifferent and, importantly, did never try to fetch an object that was wrong, even when asked. The dog and the cup were thus not chance events. However, without replication with control for contextual cueing, we cannot move beyond "interesting anecdote" on this one. In theory, self-made pictures, especially of a low-iconic nature, makes performance through reality mode unlikely. Unfortunately, the numerous occasions where Viki did not fetch objects in response to the dot connecting exercise argues against her understanding the depicting potential of self-made drawings.

To conclude, Viki showed clear evidence of recognizing the objects in both realistic and more abstract depictions. In an analysis of Viki's mistakes with photographs and realistic pictures, Hayes and Hayes (1953) could not find a reliable trend other than lack of attention. They make no similar error analysis for line drawings. Furthermore, they do not use the line-drawing data to argue for a representational ability in Viki but instead use the nonsense design discrimination data, which to me sounds like they believe that performance with realistic pictures in matching and discrimination tasks proves their point. Without doubt this at least allows for a reality mode processing, but I think that Viki shows something more when she succeeds with abstracted drawings. If successful categorical performance with novel line drawings is dependent on processes that cannot be contained in reality mode, Viki must be granted a referential understanding of pictures, although with some caution since the novelty requirement is sometimes violated in the Hayes study.

The role of growing up with humans, i.e. human enculturation (see Chapter 8), is likely a factor in Viki's pictorial development, but that is like saying that experience is important. We need to figure out exactly what it is that makes this development possible, and we need to compare experienced and naïve subjects on comparable tasks. A start in that direction is presented in Chapter 13.

5.2 Cross-modal matching

Davenport and Rogers (1971) notes in connection to the Köhler (1925/1957) and Hayes and Hayes (1953) studies that they "are unaware of any study which unequivocally demonstrates the ability of chimpanzees or any other organisms to perceive the representational character of photographs without specific training" and Winner and Ettlinger (1979) criticise Köhler, and Hayes and Hayes, for not producing a control for associative learning. So by the 1970s the question of recognition of pictures by apes seemed to still have been unresolved.

Davenport and Rogers themselves seemed to have hit upon the neglected importance of pictorial processing as a secondary effect of their research into cross-modal matching in apes. In cross-modal matching a visual sample has to be matched to a haptic comparison stimulus (e.g. Davenport & Rogers, 1970) or vice versa: a haptic sample matched to visual comparisons (e.g. Davenport et al., 1973). The subject typically puts its hand into a box or behind a screen and feels an object that cannot be seen. The subject then has to match what it feels to visually accessible objects or pictures of objects. The matching can be simultaneous or delayed.

The initial interest in cross-modal matching stemmed from the idea that such transfer between modalities seemed to be uniquely human.[41] It was reasoned that symbolic mediation was the key ingredient. Since it was shown that apes after all could pass these tests, although it took about 500 trials to learn the matching procedure, the conclusion had to be that apes possessed a "metamodal concept of stimulus equivalence […] independent of verbal language" (Davenport & Rogers, 1970). The apes that participated in the complete testing were ultimately two chimpanzees and one orangutan of unreported background. The work had started with 11 subjects, which illustrates the long process involved in using tests that are not intuitive to the subjects and that require much drilling. However, the benefits of teaching the cross-modal matching apparatus soon opened up an easy way to test also picture comprehension.

The three apes from the above study performed a cross-modal matching procedure with photographs instead of objects, reported in Davenport and Rogers (1971). Life-sized colour and black-and-white photographs of mostly unfamiliar objects were used as target stimuli, and real objects as haptic matches and non-matches. There was good control for learning effects since each photograph was used only once. The subjects performed above chance and there was no difference between the two photograph categories. Since the subjects were naïve to pictures, the pictures were highly realistic, and furthermore placed behind glass, a reality mode of picture processing is the given candidate for the apes' performance. If this was indeed the case, we can also be confident in assuming that colour hues are not always a necessary element for differentiating photographs from reality. This makes sense since colour hues are a variable property and under certain conditions, i.e. in dim lightning, most real-life objects approach greyscale.[42] Davenport and Rogers (1971) concluded that apes can perceive the objects of photographs at first sight, but their own study does not capture what they sought in their introduction: A study that unequivocally demonstrates the ability to perceive the representational character of photographs.

Davenport et al. (1975) introduced delayed matching in the paradigm, and also the use of pictures that would strengthen the representational character of the task, i.e. non-photographs. They wanted to further demonstrate the ability of apes to keep, and act on, a representation of an object that was only present in their minds and nowhere else. The subjects in this study, five nursery-reared chimpanzees, were different from the ones in the study reported above but they had all participated in an inverted version of the original 1970 study, i.e. they were familiar with matching haptic samples to visual comparison objects. Four of the five had reached the criterion of 70% correct matching (Davenport et al., 1973). The apparatus in Davenport et al. (1975) was basically the same as in earlier studies with the haptic sample occluded from sight but reachable, and the matching and non-matching pictures be-

---

[41] Which ability does not in its scientific infancy?

[42] Beilin and Pearlman (1991) could for example not find a difference due to colour in 3 and 5-year-old humans' nondifferentiated reasoning about photographic stimuli and their referents.

hind touch-sensitive glass.[43] Five classes of pictures were used in the simultaneous matching condition. (Only colour photographs were used when testing delayed matching, of up to 20 seconds.) The picture conditions were the following: Full sized colour photographs, full sized black-and-white photographs, half-sized black-and-white photographs, full-sized silhouette pictures and, lastly, full-sized line drawings. The silhouette pictures were created by increasing the contrast in black-and-white photographs until only the black mass of the depicted object against a white background was discernable. To control for learning effects, and thus make sure that choices were based on similarity judgements, 40 critical one-trial problems using completely novel stimuli, or novel combinations of stimuli, were given for each picture condition. However, they seem to have gone through each picture condition before moving on to the next, thus neglected to control for order effects. At the time the subjects received the line drawing trials they had thus already had extensive experience with the previous conditions.

For the simultaneous matching problems four of the five chimpanzees performed above chance in the full-sized colour and black-and-white photograph conditions. All five were significantly above chance on the half-sized black-and-white photographs. Three of five were correct on the silhouette pictures and four out of five passed the line drawing condition. Only one of the subjects performed below chance in more than one condition, and that was for the colour and high-contrast condition. Taken together, the subjects performed a bit worse than they had made in the 1973 study that had utilised objects instead of pictures. In the delayed matching condition with colour photographs, four out of five subjects performed above chance, but in the critical tests with novel stimuli only two performed above chance, but they did perform better than on the earlier simultaneous matching with colour photographs.[44]

It should be remembered that the above testing was all in the context of cross-modal matching, which might very well have been a significant factor for the picture performance shown. As long as the pictures are not weighted against each other to counteract critical visual properties, one cannot rule out that shape matching rather than identity matching took place. One should be able to perform quite all right by comparing the remembered sample shapes to the pictured shapes, which remains intact in all picture conditions. If the animals had hit upon this strategy in the silhouette condition, which was a link between the photorealistic conditions and the presumably more abstract line drawing condition, the transfer to line drawings might be a simple task. It is a pity that the report does not include examples of the line drawings used. The fact that most sample objects in the study were fairly unknown to the subjects strengthens the advantage of matching based on shape similarity rather than object identity. (The fact that the pictures were behind glass highlights the aspect of the pictures as shapes in their turn, as opposed to pigment on a two-dimensional surface.) I think the data on the delayed matching task, i.e. the poor performance with novel stimuli, supports my concerns that the subjects could

---

[43] In those days touch-sensitive meant that the glass moved with pressure and activated some switch.

[44] The correct identification of line-drawn silhouettes of simple shapes in a cross-modal matching task (haptic sample, visual comparison) was replicated with one year old human children by Rose et al. (1983).

not reliably identify the objects and therefore had problems remembering them. If the task instead had been to match identifiable objects, the delayed matching would improve but pictorial performance might have declined, because then the pictures would also have to be decoded with their identity taken into account. Line drawings might then fall short. I am afraid that the cross-modal matching experiments by Davenport and colleagues will have to pass as combinations of picture processing in the surface and reality modes.

Winner and Ettlinger (1979) might agree with the above conclusions since they failed to reproduce the results of Davenport et al. (1971; 1975) in a study on both regular matching-to-sample (MTS) and cross-modal recognition using photographs. They criticised Davenport et al. on the grounds that they did not create controlled pairs of comparison stimuli that were matched on dimensions such as size. However, they do grant rhesus macaques in a study by Zimmerman and Hochberg (1970) the ability to transfer discrimination of "simple object shapes" to photographs and drawings. But see my objections above regarding shape matching that might apply for the results with drawings in this case as well.

In their 1979 paper Winner and Ettlinger tries to address the shortcomings of Davenport et al. using both regular, unimodal, MTS, and a cross-modal paradigms. Beside size cues they also wanted to test if familiarity with the depicted objects or reliance on colour affected performance.

For the unimodal testing subjects were two juvenile chimpanzees with extensive object-to-object matching experience. They used a procedure quite different from Davenport's. They put rewards inside holes in a Klüver board which they then plugged with cork. On top of the cork the stimuli were fastened. In a successive version of the test the subject first removed a single cork with the sample on it and found a reward underneath. It then got to choose between two corks with the match and non-match on them. Objects, life-sized colour and black-and-white photographs were used as stimuli. After object-to-object matching the subjects received trials on object-to-picture and picture-to-object matching. They then received a simultaneous version of the mentioned conditions before they were tested on successive picture-to-picture matching. In order to make sure that the chimpanzees were paying attention to the stimuli, presentation was varied systematically: flat presentation, or at a 45 degree angle behind, or in front, of Plexiglas. However, the stimuli were still placed on the corks in the board and the subjects themselves manipulated the pictures when choosing. Throughout the testing period, following the sessions, the chimpanzees were tested for object-to-object matching to make sure that they had not developed a lapse in memory for matching as such.

In total the chimpanzees were given 40 trials per day for 16 days. While they performed at 90% success rate with object-to-object matching they mostly performed at chance with pictures. They also failed to match two identical photographs on four consecutive days of training (number of trials unknown). To me this implies that the chimpanzees did not attend to the motifs of the pictures.

Two different juvenile chimpanzees from the above two were tested in the cross-modal recognition tests. They had previous experience with three-dimensional stimuli using the same setup. Objects and pictures from the earlier experiment were reused for this experiment. The subjects were rewarded for displacing one of two corks for six trials, and was then given a single generalisation trial in the opposite modality of the training one (e.g. six visual trials followed by a haptic one). Sometimes the objects in the visual mode were substituted for their photographs. Significant performance was only obtained with objects and not their photographs, and no difference between colour and black-and-white photographs was found. The subjects had been given more than 100 trials.

Although Winner and Ettlinger (1979) call their study a replication, the setup used is very different from the one of Davenport and colleagues, where the reward was given separately from the manipulation of the stimuli. The chimpanzee pulled the stimuli (when haptic), or pressed the window (when visual), and then the reward was administered from a separate part of the machinery. Here the rewards were baited underneath the correct stimuli. This ought to affect the attention of the subjects.[45] Furthermore, the subjects were allowed to interact with the pictures, disturbing the illusion of being somewhat real objects behind glass, and allowing for action guided by reality mode processing of pictures.

Winner and Ettlinger concluded that for their four chimpanzees in the two studies photographs were treated as meaningless two-dimensional stimuli and not pictorial stimuli that had to be interpreted. A problem with the study is that they could not come up with *any* situation where the subjects showed that they recognised something in the photographs, such as mouthing a picture or the like. (Maybe they would if they had been using food pictures.) Human judges had reported the photographs to be very clear, which probably entails them being near to life. We can thus not exclude that these animals suffered from a prominence problem, where the situation of the presentation and use of the material digressed attention from the motifs of the pictures. I believe that the finding nature of the task might have been such an obstacle, as well as the appearance of the pictures as flat surfaces.

Malone et al. (1980) also tried to reproduce the cross-modal work with chimpanzees by Davenport and colleagues. They found that macaques matched objects to photographs, and photographs to objects equally well as did the chimpanzees. However, they seemed to have needed more training on the matching per se. (Only two subjects were used since three failed to learn matching altogether.) Full-sized colour photographs were used as visual stimuli and an assortment of small, mainly unknown, objects were used as tactile stimuli. They raise, but do not test, the issue of whether familiarity with the objects is a relevant factor for matching performance. They cite Rumbaugh and Gill (1976, in Malone et al., 1980) who found that Lana, a chimpanzee trained in using visual symbols (lexigrams), performed radically different with familiar and unfamiliar objects in a cross-modal matching task, and also with familiar foods with and without lexigram associations (Rumbaugh & Gill 1976, in Tolan et al., 1981).

The apparatus used by Malone et al. was the same used in the studies by Davenport and colleagues, with the photographs enclosed behind glass. It seems to be the very first monkey data on object - photograph equivalence in a cross-modal task ob-

---

[45] In the positive direction one would presume, but see section 12.7. Attention was probably fixed on the corks, not the pictures.

tained. However, the authors admit that because of the problems of teaching matching to the subjects it is not clear from this experiment if the monkeys could recognise photographs at first sight or not. Either they could, but could not show this since they did not grasp that they were supposed to match to sample, or they could not, and could therefore not learn matching until they had learned to perceive the content of the photographs.

In a follow up study Tolan et al. (1981) exposed the same two macaque subjects as above to black-and-white photographs, silhouette pictures, and line drawings in simultaneous cross-modal matching in an attempt to extend their data to match also those aspects of the chimpanzee findings of Davenport et al. (1975). They also tested colour photographs in simultaneous and, furthermore, delayed (10 seconds) matching. With the colour photographs the monkeys could perform above chance in both the simultaneous and the delayed condition, but in the latter familiarity with the depicted stimuli seemed to have been crucial. The monkeys also performed above chance with all the other types of pictures except the line drawings. Further training was needed for one of the subjects in order to transfer from colour to black-and-white photographs. Generalisation to novel silhouette photographs does not seem to have been a problem, although initial transfer from black-and-white photographs was shaky. The ability to match silhouette photographs remained when they failed at matching above chance on line drawing. Even when allowed to both see and handle the objects, thus no longer a cross-modal problem, and match them to line drawings, did they fail to perform above chance. The line drawings used are not shown in the report but are said to be of an outline nature "with no internal details drawn in," and thus different from the drawings previously used with chimpanzees which had more features than the outline drawn in.

The fact that photographs but not line drawings could be matched to objects suggests that a reality mode, and not a pictorial mode, of picture processing were employed by the macaques. The authors also acknowledge that the "[...] photographs were probably perceived in much the same way as visible objects, especially since the animals were prevented from having any tactile experience with the photographs" (Tolan et al., 1981, p. 298). They suggest that the reason that Winner and Ettlinger (1979) got different results from Davenport et al. (1975) was exactly because the subjects had different opportunities to handle the pictures and thus focus their attention on the differences rather than similarities between photographs and real-life objects. Discovering the pictures' flatness, lack of appropriate texture, and so forth, could be such spoilers. The subjects could therefore never learn to reliably match with photographs.

The monkeys' successful performance with silhouette photographs is rightfully not seen as an intermediate stage between photographs and line drawings by Tolan and colleagues (1981), although the pictures differed markedly from the three-dimensional objects. They propose that the macaques, and previously the chimpanzees, might have learned to match visual profiles to haptic profiles. This might be the case, but it is not surprising if the silhouettes after all could be identified on an object level rather than as an arbitrary shape. The silhouette of a bird of prey is successfully (one would presumed) used to discourage other birds from crashing into windows. Petit and Thierry (1993) report that Guinea baboons (*Papio papio*) react aggressively towards baboon-like silhouettes cast on their cage wall. In dim lightning most objects are recognisable through their silhouette, unless the view is too atypical. A shared silhouette between objects in real life and on a picture is a small commonality if one considers all features being equal, but shape is a key feature for recognising objects for most visual species, from bird (e.g. Looney & Cohen, 1974) to human (e.g. Quinn et al., 2001). Pigeons, like the primates above, find silhouettes easier to discriminate than outline drawings (see Cabe, 1980).

However, drawings can capitalise on the ability of the viewer to identify an object through its shape. When the conditions are right some drawings can therefore be recognised in reality mode. To support this interpretation one should find that line drawings that enhance the figure – ground appearance should give higher success rates in recognition than line drawings that do not. Colouration, shadowing, and variation in density might be such factors. From this perspective colour does not help with recognition of the features of a drawn figure, but rather points out its status as a figure against a background as such. However, with subjects that are on the verge of pictorial perception the identification of a recognisable shape in a non-photographic picture potentially feeds back to the recognition of local features as well, and in that case colouration, shadows etc. enters iconic significance.

That shape would be enough of a feature for identifying objects from a reality perspective is thus not surprising. Shape alone seems to be sufficient for matching pictures to objects, regardless if this is done in a reality or pictorial mode, but would shape suffice for matching on the basis of a surface analysis, as Tolan et al. (1981) suggest? We know from monkey data that some discriminations are indeed based solely on local features, such as colour, even though the experimenters intend more holistic solutions (e.g. D'Amato and van Sant, 1988).

The context, a silhouette viewed in broad daylight, implies that reality mode can be quite flexible and allow for atypical, but not impossible, views of objects. However, the drawing results show us that there are limits. Shape in the form of only an outline, with less of a figure - ground appearance, does not seem to be sufficient. This is supported by results of Zimmerman and Hochberg (1970) that show that monkeys discriminate drawn shapes better when the figure-ground relationship is enhanced by contrasting colours or shadows. Black lines on a white background did not work well at all.

5.3 Ai

Itakura (1994) tested black and white drawings on the lexigram-competent female chimpanzee Ai at the Primate Research Institute in Kyoto, Japan. Ai was 12 years old at the time. Matsuzawa (2003) gives some background on Ai. Although a project involving symbol learning and "language-like" competencies, like counting and ascribing numbers, the main goal of the Ai project has never been one of interspecies communication, as in the American language studies of the 1970s and onwards. Rather, the ambition has been to map how chimpanzees perceive their world. The Kyoto researchers favoured a Japanese version of the computerized lexigram system used with the chimpanzee Lana by Rumbaugh and colleagues (Rumbaugh, 1977), instead of a gestural system (Gardner & Gardner, 1969), or one based on physical tokens (Premack, 1971). This choice was made in order to integrate symbol use with other cognitive tasks on a computer screen.

Itakura's (1994) paper does not report what previous experience Ai might have had with non-photographic pictures before she was tested on line-drawings, but at least she had extensive training on discriminating figure from ground in drawing-like stimuli since her lexigrams are all black-and-white patterns (Matsuzawa, 2003), as is the case for numerous other discrimination tasks that she has been exposed to during her lifetime.

At the onset of the study Ai could "name" six chimpanzees, six humans and two orangutans with individual letters from the alphabet and she could ascribe these to photographs or movies of the individuals. Of these, three humans, three chimpanzees, and one orangutan were included as stimuli in the testing. (Ai herself was one of the chimpanzees.) Colour photographs and black-and-white line drawings were used as representations of the individuals and were to be named with the respective letter. Two extra letters were included, that designated one chimpanzee and one orangutan that Ai only knew from photographs. The line drawings had been constructed from photographic templates and were quite detailed (*fig. 8*).

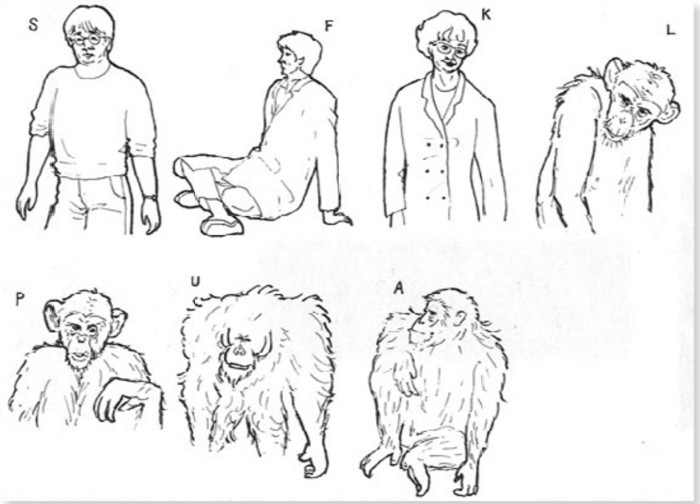


Figure 8. Some of the drawings used in Itakura (1994). The letters designate the "names" that Ai was required to match to the pictures. From Itakura (1994).

The presentation and apparatus used was computerised. One of the seven individuals was the target in each trial and after the sample had been presented on the screen two seconds elapsed before the letters were presented. A food item was automatically dispensed each time Ai made a correct choice among the nine letters. The description in the report is somewhat unclear, but it seems that the matching was simultaneous, i.e. the sample picture stayed onscreen during the choice. If Ai had been incorrect a signal was sounded and the trial restarted after a 3-second timeout. Ai seems to have been allowed as many correction trials as she needed. Twenty-one pictures (three for each of the seven individuals) were used for seven sessions, each session consisting of 84 trials. This means that each picture was used four times per session. One new drawing replaced a photograph in each session, starting with one individual in session one, and another one added in session two. Thus, in session seven all seven stimuli individuals were represented by two photographs and one drawing. Photographs were changed for each session, but the paper does not say if the drawings were also changed or remained the same throughout testing. This would mean that in the seventh session the first individual's drawing had been presented on four times seven trials, the second individual's drawing four times six trials, and so on, meaning that chances for rote learning of the drawings were high for most of them.

Ai was correct on 100% of the photographs. Performance was not above chance on the first trial on each of the line drawings. Ai tended to choose the letter that had been reinforced in the previous trial, and she responded randomly during correction trials. Some of the drawings only needed to be rewarded once for Ai to perform correctly on the next trial that involved that picture, meaning that her memory for correct pairings was very good.

In a second phase of the study all stimuli individuals were represented by two photographs and one line drawing each from the onset. Again 84 trials (not counting the correction trials) were given per session, and new line drawings and photographs were introduced in each session. But still, each picture was used as target four times within a session. The report does not say how many sessions were run in this second phase. For some reason Ai was much better in phase two and responded accurately to line drawings at first presentation for four of the seven individuals at a level of 60% correct, which is significantly above chance. As in phase one, her success rate improved during correction trials. The four individuals that she seemed to be able to pinpoint were the three humans and the one orangutan. The three chimpanzees could thus not be identified in line drawings by Ai. Nor could they be readily identified by a control group of human subjects. The control group had no problems with the drawings of the humans or the orangutan though, similarly to Ai.

That the orangutan sticks out as stimuli might not be surprising at first glance. It is not implausible that the orangutan drawings show some invariant features that the other drawings lack. However, that judgement will have to be based on an analysis of the complete drawing material, which is not included in the published data. The individual humans could possibly also be discriminated on surface features since, for example, only one of the two males wears glasses, and the third person also wears glasses, but is a woman. Combining the glasses feature with, for example, the presence of much hair on the head, will suffice for mutually exclusive identification of the humans. It is in theory possible to appreciate these distinctive features and combinations without attributing "glasses" and "hair" to them, and thus solve the problem completely in a surface mode. However, this is also an alternative explanation that has to be checked against the full stimulus set. Pattern recognition rather than categorical identification is therefore a possible explanation for Ai's success rate, and might also account for her uneven performance.

When looking at her errors one finds that it was more common for Ai to confuse individuals within the chimpanzee category than between species, although one of the chimpanzees was paired with the name of the orangutan as many times as with those of incorrect chimpanzees. The orangutan in turn was named as a human several times. The humans, however, were only named as other humans. Does this mean that chimpanzees look like chimpanzees and humans look like humans in these drawings, and that they are just a bit difficult to tell apart as individuals? Or, does it rather mean that the visual pattern is recognisable as a class, i.e. "this is one of the human-patterns"? This latter alternative might explain the hardships of keeping individual patterns apart while keeping the classes apart. The invariant features between the humans as a group as opposed to the chimpanzees are easier to keep track of than the invariant features of the individuals, although that was apparently possible to a degree for four of the individuals. This can all be done without actually seeing "humanness" or "chimpanzeeness" in the pictures.

Given Ai's extensive experience with pictorial stimuli in the past in the form of photographs, and her ability in the present task to accurately "name" photographs, we must consider the possibility that she might have been very aware of the depicting function of the drawings, but could just not decode the transformations in most of them. To ease her task I would propose to do the matching the other way around, so that she is given a name and then has to pick out the correct picture in an array to go with it. In that format she can compare pictures with each other and pinpoint the drawing that is most like the referent. This way there is no need to invoke an absolute resemblance, only a relative one, and the matching ought to become easier, even when the stimuli is imperfect from the point of view of the subject.

Another hint that Ai knows what is actually required of her in this study is the use of "names" as matches. Ai has shown an understanding that the letters pair with specific individuals in her earlier performance and on the trials with photographs. It would be strange if she somehow did not make the connection that the line drawings also must pair with these "names" and thereby the individuals behind them. Analogical reasoning is not beyond the ape mind (e.g. Oden et al., 2001). We could then presume that Ai might be reasoning that not only do the drawings pair with letters, but just as letters pair with the content of photographs, i.e. people, perhaps drawings also depict people, although in a very poor format. However, we cannot be entirely sure of what the names actually decode for Ai, especially after having subjected her to stimuli that unreliably match to the names, i.e. the drawings. Successful matching between drawings and photographs should therefore be the next necessary step in evaluating her competence with line drawings. For now we must conclude that Ai's performance in this study unfortunately leaves us with inconclusive evidence.

Another study with Ai, which directly targets the questions raised in this thesis, was recently published by Tanaka (2007b) of the Kyoto Primate Research Institute. Tanaka tested Ai and other residing chimpanzees on a generalisation task involving photographs and drawings of various degrees of (experimenter defined) realism. Besides Ai, three other adult females from the institute, as well as three 4 to 5-year old juveniles served as subjects. On a computer touchscreen the chimpanzees were required to choose 3 pictures in an array of 12. Correction trials were allowed and the subjects were food reinforced for every correct indication of a flower picture. Training went effortlessly for all but one of the adult subjects and generalisation to novel colour photographs was above chance for all but the said subject. At the second exposure to the new photographs they were all well above chance. This does not have to indicate a learning effect due to memorising correct choices per se, but can rather indicate that the procedure itself was somewhat upset due to new stimuli, or just reflect perseveration errors from choosing old stimuli, regardless of whether they were correct or not on previous trials. Tanaka (2007b) admits that the subjects sometimes chose non-flower pictures that contained colourful patches. Whether they mistook these for flowers or did not chose in accordance with a flower concept at all we do not know.

Next step was generalisation to "realistic" colour sketches and "less realistic" colour cartoons, or computer clip-art. The criteria for these two groups are not given so we have to assume that it was the subjective judgement by the experimenter. 480 novel pictures of each type were tested, of which 120 in each group were flowers. (Non-targets were trees, leaves, branches, grasses, ground surfaces and various everyday objects.) Ai was the only adult subject who could chose correctly in both categories at first exposure readily above chance. Chloe, one of the other adults, managed for both categories as well, but sketches outdid cartoons/clip-art. The third adult performed equally low for both categories but still above chance. The fourth adult subject performed at chance level for all non-photographic stimuli. Only Ai benefited from repeated exposure and increased her scores as the test went along. In contrast to the variable performance of the adults all three juveniles performed above chance from trial one on both types of pictures. They also showed a strong increase in performance with repeated exposure. But from the presented data one cannot draw any conclusions regarding which category was the easier one for neither the juveniles nor the adults as a group. Only Chloe showed a clear decline in performance with decrease in "realism."

To further abstract the pictures colour was removed from 48 novel line-drawn pictures, of which 12 were flowers. Now only Ai performed above chance in the adult group, but not until the third presentation of the new pictures. She then showed a steep learning curve. It is difficult to say if Ai suddenly grasped, by the third exposure, that also the line-drawings were meant to be flowers, if she had formed a new category perhaps not remotely connected to flowers, or if she had just memorised the correct pictures. After all, they were only 12. A fact that speaks against the latter is that when tested with interspersed novel Kanji (Chinese) characters in the set instead of novel pictures, she did not learn the correct choices with repeated exposure. In fact none of the subjects learned the rewarded Kanji character for flower. Again in contrast with the adults, two of the three juveniles were above chance from trial one on the black-and-white line-drawings, and the third juvenile on second exposure.

Tanaka (2007b) drew three conclusions from this experiment. 1. Line-drawings are not seen as equivalent to photographs for all subjects. 2. There is a critical period for learning "pictorial representation." 3. There is a relationship between symbolic and pictorial competence. This last conclusion is based on the fact that of all subjects only Ai was skilful in using a visual symbolic system, i.e. lexigrams. As mechanism Tanaka (2007b) evokes naming, supported by the research of Callaghan (e.g. 2000), rather than a general symbolic function. However, this does not explain why the juveniles, who had not acquired lexigram use, were proficient in the task. The conclusion that they were experiencing a "critical period" needs to be clarified. Pictorial competence is a complex ability and there are many constituting abilities for which there might be a period of heightened sensitivity. This explanation must also rule out a critical period, i.e. a heightened non-adult sensitivity, to some crucial part of the experimental procedure. Perhaps young learners might just be more proficient in generalising visual stimuli.

To control for visual generalisation from photographs to drawings, perhaps the abstracted pictures should have been tested before photographs. If transfer would occur in this direction the claim for recognition of drawings as objects would be stronger. On a further note, repeated response to flower-like stimuli, regardless of level of abstraction, will not do as an indication of pictorial competence. A minimum requirement would be a conditional discrimination, as in a MTS paradigm, where flowers are sometimes correct and sometimes wrong, and where several categories of drawn objects are tested at the same time. That said, remember the results of Hochberg and Brooks (1962), where the 19 months old boy was able to recognise line-drawings, albeit highly prototypical ones. Flowers are unfortunately also a type of stimuli that takes highly canonical forms in most cartoons and clipart. A detailed analysis of those particular flower pictures that proved problematic for the subjects in Tanaka (2007b) would be most interesting.

Almost no studies targeted at picture perception per se have been made with monkeys. Bovet and Vauclair (1998) studied the categorisation of objects and photographs into foods and non-foods by baboons (*Papio anubis*). Interestingly, Bovet and Vauclair would be happy to see confusion behaviours. Their goal was to show that primates can establish a correspondence between picture and object, and the very criterion for recognition of a photograph is treating it as the real object. This is shown by transfer from categorising real objects to categorising their depictions.

It was found that the baboons could do this. However, it was also found that cut-out pictures were easier to categorise than whole photographs. Cutting out a figure highlights its shape and similarity to a real object. This is not only indicative of performance in reality mode, but also of feature matching and performance in surface mode. However, the researchers matched the two groups of objects (foods and non-foods) as close as possible in terms of size and colour to counteract this latter possibility. In addition, it was noted that the baboons often held out their hand towards food items but not towards non-food during presentation. When presented with cut-out photographs the response was the same. Food pictures elicited a stretched out arm, while non-food did not. The food cut-outs that did not yield a response were exactly those that were often categorised as non-food in the previous experiment. The baboons never reached for any of the intact photographs, regardless of motif. When the realness of the cut-outs was decreased by rotating them slowly, and thereby enhancing their two-dimensionality, subjects stretched towards the pictures less and also performed worse on categorical judgement.

Another interesting finding was that the most difficult food pictures to recognise were large spherical fruits. The authors speculate that the lack of a third dimension in pictures is particularly problematic for such fruits. Cutting them out did not help. The above study demonstrates, according to Bovet and Vauclair (2000), an ability to "relate real objects to their pictorial representations" (p. 150). The actual nature of that relationship is not further defined, however.

That a cut-out effect facilitates recognition has been found also for humans. Fuglesang (1970, in Jones & Hagen, 1980) found that Zambians (picture experience unspecified) performed better when the background information was blocked out in photographs. They also performed significantly better with silhouette pictures than with line-drawings. Figure-ground cues are strengthened in both these modifications.

5.4 Referential use of pictorial information

If it is difficult to construct stimuli per se that can only be decoded within a pictorial mode while excluding the other modes, using pictures in a referential *task* is another way to investigate the ability. We do not eliminate the problem that also pictures seen in reality mode can be used in such a task, or even surface mode if training is allowed, but reference is an important piece of the puzzle, and if subjects can solve referential tasks using e.g. photographs while they cannot decode pictures that are low in realism or depict more dynamic content, we can deduce that the problem is not one predominantly of reference, but depends more on the properties of the medium.

Apes have been shown to have problems following indexical cues in a number of experiments (for a review, see e.g. Byrnit, 2005). An indexical cue is one where nearness to the target singles it out, of which perhaps close-range pointing is a paradigmatic example. In order to test whether iconic information would enhance the salience of indexical cues, as well as being informative as a cue on its own, Herrmann et al. (2005) conducted the so far largest, in terms of participants, pictorial study made with apes. 27 chimpanzees, bonobos, gorillas, and orangutans at the Wolfgang Köhler Primate Research Center in the Leipzig Zoo, Germany, were subjected to several versions of an object-choice task.

The first task was simply to choose one of two cups of which the one baited with food was indicated by a photograph or a replica placed on top of it. The photograph (in colour) and the replica (of rubber) depicted the content of the cup. Nothing was placed on the cup that was empty in this experiment. The placement of a cue on one of the cups was made either in view, or occluded from the subject's sight. Of the 27 subjects 11 managed to chose above chance in at least one of the conditions. Of the 12 chimpanzees only one succeeded. Four of the six orangutans, two of the five gorillas and all four bonobos likewise succeeded in the task. A clear species difference thus suggests itself in this study, where especially common chimpanzees fell short. While above chance in three conditions, as a group the subjects did not perform above chance when the replica was used in hidden placement. This implies that the photograph afforded different information than the replica. When the actual placement (irrespectively of medium) in full view of the subjects was enough to inform them of the correct choice, this may have been based on other information than the iconic one. In the hidden condition the replica lost its cue value while the photograph retained its value. (Or acquired new value.) This time the subjects probably used their recognition of food in the photograph and made their choices accordingly.

The authors suggest that the replica, due to its three-dimensionality, was too much an object in itself to serve as a representation, which is in accordance with the dual-representation hypothesis of DeLoache. Somewhat contrary to this explanation

I rather believe that the photograph was successful just because it was perceived as even more "real" than the rubber banana used in this experiment. The photograph used might have captured more of "banana-ness" than the rubber replica did. This effect would have been enhanced by some experience the subjects had of the replica, but not the photograph, before the experiment. To make sure that the subjects could tell the difference they had namely been given a choice between a real banana and the rubber banana. If they chose the replica they were given this to smell and touch. The subjects thus knew beforehand that the rubber object on top of one of the cups was not remotely close to an edible banana. (Or *any* banana.) The perceived iconicity was probably lost with this information. This might be the reason why so many subjects of the 27 failed altogether. They might of course have failed simply because they could not understand indexical cueing throughout the experiment, which is not uncommon for novel cues (Tomasello et al., 1997), but to complicate the matter they might also have failed to recognise the value of the iconic information of both the replica and the photograph. This information might otherwise have helped them as it did the subjects who performed better with the photograph than the replica. However, this experiment cannot arbitrate between performing on indexical versus purely iconic grounds.

The second experiment in Herrmann et al. (2005) takes care of this problem by placing photographs or replicas on both cups for the 11 successful subjects from the previous experiment. The distractor items could either depict colourful non-edible objects or other fruits than the target ones. Also in this test the subjects got to inspect the replicas but not the photographs prior to testing. It was found that the subjects as a group chose above chance in all conditions except when replicas were used and the distractor item was another fruit. The surprisingly good performance with object, as opposed to fruit, distractors in the replica condition Herrmann and colleagues attribute to the individual scores of three orangutans and one bonobo who seemed to develop a fondness for fruit replicas. When another object was the distractor item they just went for the replica that they found most interesting. Thus they might have appreciated the likeness to fruit, but they did not necessarily use this information in relation to the baiting of the cups. With photographs four subjects performed above chance with object distractors and two with fruit distractors. The task was in other words difficult.

Overall, the subjects in both experiments show great variation between conditions which suggests that other factors than a general "iconic" ability, as defined by the experimenters who chose the stimuli, confounded the responses. A detailed analysis of the individual stimulus items is probably necessary in order to conclude what was going on. In the second experiment only one subject, a gorilla, managed all four conditions. Had only one subject of an original 27 gotten the hang of iconicity, irrespectively of medium? It seems so. But Herrmann et al. (2005) were not satisfied. They also wanted to see if iconicity could work independently from indexicality. Instead of using pictures to label a content, they wanted to use iconicity to communicate the right choice.

In the third experiment very different looking cups and boxes were used to hide the food and the correct choice was cued by holding up a photograph between the containers depicting the correct hiding place. The fourth experiment was similar, but this time the containers were transparent and both contained food. However, only one of them could be opened and the photograph held up by the experimenter indicated which food that was possible to get hold of. In both these experiments none of the 11 subjects succeeded above chance. But perhaps with time they would, since there was a clear learning effect between the start and the end of testing.

Herrmann et al. (2005) conclude that when indexical information is removed great apes cannot easily substitute this with the communicative intent of the experimenter. That is, the "reason" for using iconic information in the first experiments was the pictures' or replicas' closeness to the containers, while the "reason" in the last two experiments was the helpfulness of the experimenter. How to discover these connections between pictures and baited cups is thus not merely a pictorial problem, but depends on how pictures manifest themselves in relation to other things, such as cups, experimenters and experimenter's minds. When no relations can be discovered pictures seem to lose their meaning, i.e. usefulness. Pictures in a non-pictorial mode, might be added. In a fully pictorial mode reference is part of the picture concept. Even though one might not necessary read others' communicative intentions into a situation, relationships that are not physically salient can still follow from the mere fact that one views a picture as being about something other than itself. In this case the referential act is a private act.

Herrmann et al. (2007) repeated one part of the above study, the combining of iconic and indexical information in an object-choice task, as part of a large battery of tests administered to over a hundred children (2.5 years), chimpanzees and orangutans. Unfortunately they lump the result of the iconic test with two other tests (pointing and looking cues) in their report. For this combined group of trials the chimpanzees and orangutans were 63% and 65% correct respectively. The human children were significantly better at 84% correct.

In a study by Tomasello et al. (1997) only chimpanzees and orangutans that had experience with human pointing or the placement of a marker could solve an object-choice task that involved these cues. Showing a replica of the container that harboured a reward was not informative for the subjects, mirroring the findings in the third and fourth experiments in Herrmann et al. (2005). Not understanding the communicative intent of the experimenter was evoked as an explanation in both studies. It should be noted that in Tomasello et al. (1997) only seven out of forty-eight children were above chance when a replica of the correct target item was used as a communicative cue. Human children had arguably extensive experience of toys and replicas compared to apes. The dual-representation hypothesis of DeLoache is mentioned as an explanation, i.e. that the replicas were too interesting as objects in their own right to serve as signs for something else, but also the lack of indexical information is blamed. The conclusion is thus that "[…] any problems children had did not concern the comprehension of communicative intentions, but rather concerned their understanding of how the particular sign functioned in the context of this particular game" (Tomasello et al., 1997, p. 1078). Applying a double standard, the data for the apes, although "not definite on the issue," was judged to be indicative of a lack of comprehension of communicative intention, while the data for the children was blamed on the stimuli.

Although framed as a test of the ability of language and non language-trained chimpanzees to delay gratification, Beran et al. (1999) is, like the above experiments, also a direct comparison between stimulus types, of which photographs is one, as well as a test of understanding the referential nature of these.

Subjects were two chimpanzees trained in the use of lexigrams (i.e. arbitrary graphic symbols), Lana (Rumbaugh, 1977) and Sherman (Savage-Rumbaugh, 1986), and a non language-trained control subject, Mercury, from the Language Research Center of Georgia State University, USA. All had experience in cognitive tasks of various sorts. The present task was to say no to an immediate reward in favour for one that was given three minutes later. Following training, three conditions were given: the immediate and deferred food visually present, the respective foods being designated by laminated photographs, and the foods being represented only by their lexigrams.

In a control session aimed at making sure that the subjects understood the stimuli, they were given a choice of a photograph of their preferred food or a photograph of a less preferred food. They all chose the picture of the preferred food. Likewise, the two lexigram competent subjects chose lexigrams that designated their preferred food before lexigrams designating their non-preferred food. Apprehending the motif of photographs was thus not a problem for any of the subjects. Likewise, when the preferred food, photograph or lexigram was put in the immediate reward position, which was a bowl by a bell-button that was to be pressed if one wanted the contents of the bowl, all subjects pressed the button to receive the immediate reward. When the preferred food instead was in the delay position, i.e. a bowl further away whose content was given to the subject only if it had refrained from ringing the bell for three minutes, the story was very different. All three subjects managed to inhibit the want for the direct food in order to receive the delayed food in about half of the trials or less. Likewise for the lexigrams, performance was low but significant, but only for the symbolically trained subjects Lana and Sherman. When it came to photographs, only Sherman reliably delayed gratification. In fact he was equally good as when the reward bowls were baited with actual food. Lana and Mercury could not delay gratification when photographs were used but pressed the button before the three minutes were up.

It seems that the two subjects that did not delay gratification in the photograph condition did not only fail to see the connection between the photographs and the foods that were given in reward, they also failed to see the photographs as real food. If they had performed in a total reality mode, they should as easily delay a response to a photograph as to a real food item. But they did not. They seem to have been stuck between two modes. On the one hand they differentiated the photographs from real food, simultaneously as apprehending their content, but on the other they could not attribute a referential function to the pictures in the task at hand. It did not occur to them that the photographs designated foods that were to be given later, in place of the pictures. That the food was placed directly in the bowls, while the photographs and lexigrams were placed against the front of the bowls, might possibly have contributed to this effect. The correspondence might have been clearer if also the photographs were placed in the bowls.

The authors explain the lack of pictorial performance with the subject's rearing histories. Both chimpanzees had extensive experience with pictures as enrichment items, but none were "trained or exposed to photographs as representational symbols, and [they have] not used them as such during [their] lifetime" (Beran et al., 1999, p. 125). Sherman on the other hand, the only one who delayed in all three conditions, learned during his early training that "not only lexigrams but also photographs and labels could represent other things in the world. Therefore, for Sherman, a photograph or lexigram representing a food produced results the same as having the food itself present" (p. 125). This makes sense. A more puzzling finding in Beran et al (1999) is why all three subjects pressed the bell when the *preferred* photograph or lexigram was in the immediate bowl, especially since Mercury did not know lexigrams. Perhaps it was simply due to the fact that they did not see a point in waiting for the delayed reward, because it was non-preferred, regardless of what was attributed to the immediate reward bowl. Thus, it does not need to mean that they understood the role of photographs or lexigrams in the immediate condition but not in the delayed condition.

Sherman's linguistic as well as pictorial training is described in Savage-Rumbaugh (1986), and a crucial part of it in Savage-Rumbaugh et al. (1980). In this seminal paper, *Reference: The Linguistic Essential*, Sherman and his companion Austin, as well as Lana, are required to sort objects, photographs, and lexigrams as "food" or "tool" into two bins.

Lana, at the time 8 years old, and Sherman and Austin, 5 and 4 years old respectively, had very different language training. It is probably this diverse background that Beran et al. (1999) refers to. Also pictures were likely used differently in the two projects. Lana's training with lexigrams focused on symbol sequencing and object naming (see Rumbaugh, 1977), while Sherman and Austin had been involved in the pragmatic use of symbols in communication (see Savage-Rumbaugh, 1986). Consequently the relationship between lexigrams and objects tended to look very different for the subjects in the two projects. Sherman and Austin had been required to ask for objects when they specifically needed them in a problem-solving situation. For this reason their vocabulary became heavily tied to the respective use of objects. If they were not allowed to manipulate objects they initially had problems naming them. Lana, on the other hand, could readily name objects that she had only visual access to, but it did not easily occur to her that she could use these names to request objects in other contexts. Her training had been to use specific sequences of lexigrams to request foods and favours of her trainers and her computer, but these interactions did usually not start with a problem to be solved.

The result of their diverse training manifested itself in the sorting study of Savage-Rumbaugh et al. (1980). Following extensive training, all three chimpanzees could sort six objects either as "food" or "tool," physically into two bins, and also by naming them with the lexigrams "tool" and "food." Given 10 novel objects, Sherman and Austin categorised them correctly on trial one. Lana, on the other hand, identified only three items, suggesting that she had not attributed the concepts "tool" and "food" (or something like "non-edible" and "edible"), but had learned to pair individual objects with lexigrams associatively. However, when required to sort the 10 novel objects into the food and tool bins, without labelling, she sorted them all correctly. Thus it seemed that the problem did not lie in conceptualising the two categories, but to encode this in terms of lexigrams.

Sherman and Austin were given a further 28 items which they could sort without difficulties as "tools" and "food." This ability generalised to photographs for Sherman but not for Austin.[46] Despite training both of them to criterion by taping photographs to the respective objects, and then require them to label the photographs on their own, only Sherman continued doing so when 10 novel photographs were presented. Austin had for some reason not treated the novel photographs, and probably neither the training photographs, as "representations of real objects" (Savage-Rumbaugh et al., 1980). Austin seemed to have learned photographs as Lana had learned lexigrams.

However, the problem was found to be attributable to the medium. The pictures were enclosed in thick plastic casings which produced artefacts, such as reflecting light. While Sherman reduced these by moving his head to get a clear view, Austin never did so, suggesting that it had never occurred to him that something informative was lurking in there.[47] When the experimenters encouraged Austin to look more carefully and slowly rotated the stimuli so as to give him opportunity to catch the content, Austin correctly identified novel photographs. This is a simple but important finding that must be kept in mind in all analysis of negative results with pictorial stimuli. The medium can obscure the message.

The last step was the most critical of the study. Sherman and Austin were to label lexigrams with lexigrams, the first experimentally controlled display of completely detached symbolic manipulation in a nonhuman species. But before the crucial test, again training was employed. Lexigrams were taped to photographs of objects and both were classified as either "tools" or "foods." Then lexigrams alone were labelled. When, in the crucial test, novel items were interspersed among the training items, Sherman and Austin correctly labelled these as "tools" and "foods" on their first attempts.

It is noteworthy that Savage-Rumbaugh and colleagues seem to have chosen photographs as a middle step, or a bridge, between objects and lexigrams. However, from a representational viewpoint, there is nothing middle about such a step. If the photographs were seen as icons they would be en par, in terms of being signs, with symbols, i.e. the lexigrams. If, on the other hand, they were seen as iconicities devoid of sign function, performance would be in the realm of actions with real objects and the photographs would only be additional training items. Photographs used in e.g. a sorting or naming task need not be seen as representations at all. In fact, Austin's confusion with the plastic medium strongly suggests that he did not intuitively ascribe useful iconic information to sheets of plastic. Not until he could recognise something familiar in them did they become useful, not the other way around.

---

[46] Lana was dropped from the experiment after her failure to apply the "tool" and "food" lexigrams.

[47] Three year old human children are quite poor at compensating inadequate viewing angles of pictures (Olson et al, 1980).

Chapter 6

Monkeys and pigeons in indirect picture tests

6.1 Earlier reviews

When it comes to indirect evidence a central resource on picture recognition in animals is the extensive review by Bovet and Vauclair (2000). It categorises experiments based on “convincing demonstrations,” experiments that indirectly could indicate picture recognition, and experiments that fail to show picture recognition. They look at two classes of performance, acquired and spontaneous responses, for still pictures and motion pictures, across several species of animals, notably monkeys and birds, but also apes, sheep, fish, humans etc.

Then Bovet and Vauclair structure their findings in a way that: “[...] at least three stages could be considered with respect to the level of precision and the nature of relationship between the objects and its picture” (Bovet & Vauclair, 2000, p. 161). They could be labelled “feature discrimination,” “object-picture correspondence” and “object-picture equivalence.” This sounds similar to the three modes of Fagot et al. (2000) and those of my own (see section 1.4). The difference between me and Fagot et al. might be slight. They put their emphasis on the relation (independence, confusion, equivalence) between objects and their pictures from an objective (i.e. the experimenter’s) viewpoint, while I put the emphasis on the *subject’s approach* to the picture, and consequently define and name my modes differently. But a crucial difference that sets both me and Fagot et al. apart from Bovet and Vauclair is the latters’ notion of “stages.” This implies that in Bovet and Vauclair’s model there is a possible transition in a hierarchy where abilities build on lower levels. What sets the modes apart from each other, on the other hand, is the interaction between specific pictures and whatever cognitive resources the perceiving subject can apply in a given context. It is thus a dynamics that could not easily, nor meaningfully, be forced into stages. A second point that sets us apart is Bovet and Vauclair’s focus on recognition in all three of their stages. For me and Fagot et al. there are only learned connections between picture and referent, if any, in the surface/independent mode. (Responding to invariant features, like colours, is not *recognition* of content.) Thus the model of Vauclair et al. does not map onto mine, but possibly pertains to three steps in a hypothetical transition between the reality and pictorial modes. This will be made clearer as we now take a closer look at the three stages of Bovet and Vauclair (2000).

As their first stage Bovet and Vauclair (2000) offer recognition due to the discrimination of salient visual features. A picture is recognised because of shared properties with something else, which can be few, but crucial properties. Salience of certain local features can sometimes be attributed to adaptive predispositions, but can also be learned. An example would be to react to the redness in an apple picture or the shape of a bird of prey on a window sticker. Bovet and Vauclair give as a typical example of their first level of recognition the case of confusion behaviours. An appreciation of certain salient features is enough to evoke recognition in the onlooker. From the perspective of my model I would call their example a special case of reality mode, since the reason that some animals react to single features in pictures is likely to be because such features are sufficient to invoke a response in real life. In the review of Bovet and Vauclair (2000) moving social or predatory cues from video stimuli typically invoke these types of responses, even from lizards, fish and spiders (see also Cabe, 1980; D'eath, 1998). But they also give several examples of monkeys (marmosets, hamadryas baboons and rhesus and pigtail macaques) that show reactions to photographic still pictures in a way pertinent to the physical counterpart, i.e. grabbing for food, giving social responses towards conspecifics, and attending to parts of depicted stimuli that are also focus of attention in real life. In some of the studies "confusion" reactions were only initial, then worn off with prolonged experience with the stimuli. Presumably they lost their confusing properties.

Bovet and Vauclair deal with confusion between pictures and objects as a separate problem and not directly a part of the definition of their stages. Their second stage cannot therefore be required to rule out confusion behaviours. What instead sets it apart from the first is that enough of the picture is perceived to make transfer between objects and pictures possible. They call this correspondence. Bovet and Vauclair do not explicitly invoke any need for differentiation at this stage although the terms "transfer" and "correspondence" hints at this. I would therefore propose that reality mode could account for all performance grouped at this second stage. As mentioned, that pictures are perceived to differ from reality in *some* regards does not exclude reality mode processing in one's attitude towards them. After all, children and animals act towards some pictures as if they were real even though, at a minimum, the third dimension is lacking. Reality mode thus cuts across the first and second levels of Bovet and Vauclair.

But let us assume that recognition on the correspondence level can involve some degree of differentiation. In that case the correspondence level would imply that differentiated similarity can precede a referential use of that similarity. This can still occur within the confines of reality mode since such generalisation can occur between objects as such. If the picture is seen as an object, albeit in an atypical form, it can relate to similar instances where the third dimension etc. apparently is retained without being about this second category. This is in accordance with the notion of iconicity as being independent of an iconic sign relation.

The third level in Bovet and Vauclair (2000) invokes the equivalence term again, just as the third mode of Fagot et al. (2000). Two criteria are proposed that would distinguish the equivalence of the third level from the correspondence of the second.

The first is the symmetry argument heard before, which means that transfer in performance should be equally good from picture to object, as from object to picture. As for Fagot et al. (2000), my objection would be that symmetry pertains best to pictures that happen to be suitable for reality mode processing. Many pictures that are more or less exclusive to a pictorial mode are not themselves very good referents.[48]

However, Bovet and Vauclair (2000) are aware of this problem and propose the same control conditions as Fagot et al. (2000), and ones that I subscribe to. This is their second criteria for the equivalence level. Subjects should be able to acknowledge a relation between an object and a picture that shares only some properties with that object, outline drawings being a good example, and especially in a cross-modal situation. One would presume, however, that those properties that are shared should not be the ones that are crucial on the feature matching level.

The Bovet and Vauclair (2000) paper is a good point of departure for the present review and deserves extra credit for being the most extensive one as yet on the subject, examining about ninety studies that involve pictures. The reader who seeks a survey of a lot of original material will find it very helpful.

The paper of Fagot et al. (2000) is also helpful, focusing on pigeons and primates. Their theoretical contributions were discussed in section 1.4. In terms of number of reviewed studies they add little to Bovet and Vauclair (2000) however.

Cabe (1980) is an in-depth review of the older material, especially about pigeons. Some of his theoretical contributions were likewise mentioned in section 1.4.

A brief review of picture research on pigeons is included in this thesis in order to illustrate variations on the three modes, notably surface mode. There are also some striking claims regarding the pictorial competence of birds (e.g. Spetch & Friedman, 2006) that need to be addressed.

6.2 Birds

Pigeons are able to learn many categories, both natural and artificial ones, from photographs (e.g. Bhatt et al., 1988), and to perform same/different judgements with pictorial stimuli (e.g. Cook et al., 2000), but the question for the present analysis is which mode of picture processing that is at work. In Cook et al. (2000) for example, judgements can be based purely on a level of pattern recognition.

Another classic example of local processing of images in animals is the “Charlie Brown” study by John Cerella (1980). Pigeons learned to discriminate images of Charlie Brown, the character from the cartoon Peanuts.[49] Discrimination generalised to novel pictures, and also to scrambled versions of Charlie Brown. This meant that local information, and not Charlie Brown as a complete figure, was the basis for discrimination. But it was also found that no single critical feature accounted for the

---

[48] Men in red jackets do not stand for the blobs of pigment in the cartoon at my desk equally well as those blobs of pigment stands for a man in red jacket.

[49] In Sweden more known for the dog character Snoopy.

performance. It seemed that Charlie Brown could be defined as a redundant collection of discrete features. The pigeons had learned to respond to several independent features. This is a clear example of picture processing in surface mode.

However, surface mode does not only work for isolated features, but can also take into consideration the relations between features. One can speak of a surface mode of a local and a global sort, where the global is very close to be a form of reality mode, but without any correspondence to previously experienced objects, i.e. a gestalt without recognition of a referent. But this gestalt can be recognised between exemplars and thus work in reality mode between pictures as such.

This is probably the reason why it looks like pigeons treat line-drawings in studies by e.g. Wasserman and colleagues as representations of objects. The pigeons' discrimination is sensitive to deletion of several features at once, but not single features, and scrambling of features (Kirkpatrick-Steger et al., 1998), just as if they had perceived motifs in the pictures. When single features proved critical, however, individual pigeons were found to have had idiosyncratic recognition strategies. Had the pigeons seen the drawings as categorisable objects from the real world, effects like this ought to be rare.[50] Pigeons can also recognise three dimensional rotations of line-drawn object shapes (Wasserman et al., 1996), which is difficult to account for with a theory based on independently perceived features.

Kirkpatrick-Steger et al. (2000) exposed pigeons to line-drawn object shapes, called "geons," and found that the shapes were difficult to learn to discriminate, but also that once they had been learned they were treated as compositional wholes, perhaps as three-dimensional objects. It is thus possible that the line-drawings also in the studies above were seen as three-dimensional objects simply by virtue of being perspective drawings. (But see e.g. Cerella, 1977.) The step to reality mode when seeing three-dimensional shapes in line drawings is very small, but to make the distinction clearer I would like to reserve the term reality mode for those occasions where recognition of pictorial content is due to spontaneously taking it for a kind of real world exemplar. When finding out with time that a stimulus behaves comparable to objects in the real world one is constructing the relationship, or rather the confusion, from a different angle altogether. It can perhaps be described as a special case of surface mode, approximating reality mode. However, these variations likely stem from similar processes. Reid and Spetch (1998) for exampled showed that pigeons could discriminate three-dimensional from two-dimensional abstract objects in pictures by using depth cues such as shading, as well as perspective transformations.[51] The latter were present in Kirkpatrick-Steger et al. (2000).

In a study on object rotation Peissig et al. (2000a; 2000b) used geons similar to the ones in Kirkpatrick-Steger et al. (2000), with the difference that they were computer 3D rendered, complete with shadows and reflections. They found similar results for rotation as Wasserman et al. (1996). With this type of stimuli it is much easier to ascribe performance to reality mode, but in order to avoid double standards differences between the performance with rendered geons and drawn geons must be

---

[50] Unless pigeons form radically individual recognition strategies of real-world items as well. (If this is indeed the case there is little hope of learning anything about pictorality from pigeons.)

[51] For human children (3 years), shape perspective is a superior pictorial depth cue in relation to shading, relative sizes, linear perspective, interposition etc. (Olson et al, 1980).

found in a comparable test. Such a test was Peissig et al. (2006), where the line drawing stimuli and the 3D shape stimuli were used in a test of generalisation to novel sizes. Generalisation was equally successful for both types of stimuli. However, generalisation performance got worse the bigger the size difference was from training stimuli. The rotation studies also found that the larger the rotation the less response from the pigeons. These two findings imply that the object status of the depictions was not completely detached from experience. Meanwhile, it was found that the transfer to novel rotations in pictures became better the higher the number of different views had been during training sessions (Wasserman et al.,1996; Peissig et al., 2000b). This seems to suggest that the birds did indeed learn to see the shapes in the pictures as three-dimensional as a result of training.[52]

To really arbitrate between reality mode and a global, relational, version of surface mode, a comparison between e.g. rotation in pictures and rotation of real objects is needed. Friedman et al. (2005) did exactly this and found a notable difference in performance on novel rotational views between photographs and objects. The latter was significantly easier for the pigeons.[53] It was concluded that pigeons perceive objects and their pictures differently.

That discrimination of scrambled pictures is more difficult when viewed in a global fashion, or reality mode, than in a local surface mode (or pictorial mode) is shown in Watanabe (2001) for pigeons. Watanabe compared discriminations of a specific human individual among others in photographs, of pigeons among other bird species in photographs, of a specific human cartoon character among other cartoons, and of a specific pigeon cartoon character among other bird cartoons. In the case of human and bird cartoons the target and comparison stimuli were made by the same artist in each group, but not across groups. Human and bird cartoons were thus very different in style and composition. One notable difference was that bird cartoons were monochromatic while the human stimuli were in colour. All stimuli were cut out and pasted on green backgrounds (*fig. 9,* p. 88).

When scrambled, photographs, and especially of pigeons, stopped being recognised while cartoons still elicited discriminative responses, but only people cartoons. The reason for the unsuccessful discrimination of scrambled photographs is probably due to them being analysed in reality mode, at least pigeon photographs. A reason to suspect that bird photographs were processed in another way than the other categories is that discrimination of bird cartoons, human photographs, and human cartoons took roughly the same number of sessions for reaching criteria in learning trials, while bird photographs were learned in less than a third of that time (Watanabe, 2001).

---

[52] Macaques (*Macaca fuscata*) in Sugihara et al. (1998) seem to have responded in a matching task with rotated stimuli as if computer rendered 3D objects were three-dimensional. However, extensive training including 360° rotations of the stimuli preceded testing.

[53] The same effect could not be found in humans, probably because they were pictorially competent.

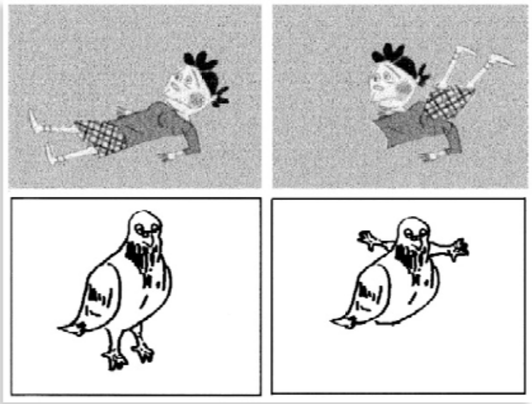


Figure 9. The cartoon human and one of the pigeon drawings used in Watanabe (2001). Whole and scrambled. From Watanbe (2001).

Photographic displays are far from certain to evoke global recognition. Even photographs of pigeons are sometimes difficult to perceive (e.g. Ryan & Lea, 1994) and many exemplars are needed to learn to discriminate individual pigeons in photographs (Nakamura et al., 2003). This suggests that even for pigeon photographs processes more close to surface mode than reality mode might often be at work. Similarly, Trillmich (1976, in Watanbe, 1997) failed to show transfer for discrimination of living individuals to their pictures in budgerigars.

That conspecific bird stimuli might be extra sensitive *in both directions* is not surprising. On the one hand, signals that single out birds from other objects might be readily reproduced in pictures and reacted upon with a predisposed heightened sensitivity. On the other hand, additional signals that are potentially relevant for identifying birds, such as vocalisation, olfactory cues and ultraviolet markings, are not reproduced in photographs. Perhaps information for discriminating *individual* birds are found in this latter category.

That people cartoons in Watanabe (2001) could be scrambled and still remain discriminable was probably due to the fact that the cartoons were processed with respect to local features. In addition, successful generalisation of discrimination to novel items occurred for all categories except people cartoons, which again suggests a local strategy for this stimulus type.

So why were pigeon cartoons sensible to scrambling? Remember that the two categories contained very different types of pictures (see *fig. 9*). The difference between people and bird cartoons can thus be attributed to different configurations of stimulus features rather than the people – pigeon dimension as such. Gibson and Wasserman (2003) showed that pigeons are able to simultaneously learn about cues as well as their spatial relations when both sources of information are available. It is plausible that the relative weight of each strategy is influenced by the presented stimuli. Matsukawa et al. (2004) confirm that this might be the case. Revisiting the cartoon and line-drawing studies already mentioned with one of their own they concluded that "pigeons use both global and local aspects, with different mixtures of these types of information depending on the particular perceptual context." In Watanabe (2001), for example the coloured elements of the people cartoons could have made recognition of local elements easier, while the black-and-white bird cartoons forced the subjects to rely more on relational properties, hence a more global processing and sensitivity for scrambling.

Watanabe's (2001) results for the pigeon cartoons closely resembles those of Kirkpatrick-Steger et al. (1998) who also used black-and-white outline drawings, but of watering cans, irons, sailboats and desk lamps. Another explanation for the sensitivity to scrambling in bird cartoons is possibly that there was some bird typical silhouette in the pictures which motivated a reality mode processing. But given the just mentioned watering cans, irons, sailboats and desk lamps, this can be easily argued against.

Whether human photographs, on the other hand, were processed in reality mode, like pigeon photographs, or was a case of processing in surface mode is less clear. If the latter it is reasonable to suspect a more relational parsing (global strategy) since it was sensitive to scrambling.

Aust and Huber (2003) found that in a "people present" vs. "people absent" discrimination, pigeons' performance dropped when scrambled and distorted photographs were displayed, but not as low as when "people absent" images were shown. Thus both individual people components and configurations of components were responsible for the pigeon's discrimination. However, the test does not convincingly show that pigeons saw the people stimuli as people. Just turning the human figure upside down had the same effect as scrambling the pictures severely. One should thus be vary of assuming that just because photographs are photographs they are treated in other ways than abstract stimuli.

But there might be good indications of recognition in other studies. Wilkie (2000) concludes that pigeons' responses to photographs of outdoor scenes corresponds to landmark use in pigeon navigation. This means that photographic scenes to some degree are seen as natural scenes. However, the transfer to novel views of the same scene is poor unless they are given many training views (Spetch et al., 2000). Correspondence might thus not be what would be expected in reality mode. Can pigeons still use this correspondence in an actual task? Cole and Honig (1994) found that pigeons could use information from photographs of a room to find food in that room[54], but they could not learn from the room to find a baited place in photographs. Similarly, Dawkins et al. (1996) could not find in pigeons any transfer from rewarded places to photographs of those places. Lechelt and Spetch (1997) found that pigeons did not transfer in any direction, although they could independently learn to use landmarks in both a real room and in digitised displays. Again, relations in the pictorial world seem to take on aspects of relations in the real world, but with no bridging between the two spaces.

Watanabe (1993) showed transfer between objects and photographs, and vice versa, for the distinction food vs. non-food. Processed in a reality mode photographs could just have been seen as further exemplars of foods and other objects. This analysis is made by Watanabe (1997) himself. He thus repeated the experiment and also found that pigeons could learn to discriminate between real objects and photographs.

One of the most, in their own judgement, convincing demonstrations of object - photograph equivalence, including differentiation, in pigeons, is published in Spetch and Friedman (2006). In order to exclude predisposed reactions they chose to look

---

[54] This was in a heavily reinforced recognition task and not a case of map reading.

at learned instead of spontaneous discriminations and therefore used nonsense objects. Furthermore, the stimuli was constructed and presented so as to require a global processing. The need for global processing would exclude responses based on local invariant features, such as colour, or memorisation of specific views. The photographic stimuli used were realistic renditions, including depth cues such as shadows, on a homogenous background. Both photographs and objects were displayed behind glass. Transfer from depicted objects to real objects, and vice versa, was found. But separate subjects were used in respective transfer group. Symmetric equivalence can therefore not be said to have been proven on an individual level.

More interesting is the claim made by Spetch and Friedman (2006) that subjects perceived a difference between objects and pictures. All pigeons got worse directly after transfer, which means that it was not an effortless transition. Furthermore, pigeons in a stable-contingency group remained above chance and reclaimed proficiency much faster than subjects in a reversed-contingency group, who had to relearn the positive stimulus altogether. This means that the birds perceived a likeness in the new stimuli to the stimuli that had preceded the transfer. However, claims such as "[…] both groups demonstrated that they perceived a difference between objects and their pictures" (p. 970) and "[…] positive transfer was unlikely to reflect an inability to tell the difference between the objects and pictures" (p. 971) gives the wrong impression. What was rather shown was that there was a perceived difference in the new group of instances of the positive stimulus. The specifically *pictorial* part of this difference remains to be proven. The same result could have been provoked by making some other transformation to the stimuli.

Reality mode accounts for the Spetch and Friedman (2006) results, but only because reality mode can work for the stimuli used. An important factor is the objects chosen to be represented in the pictures, and here simple but realistic computer renderings were used.

Pigeons have a different visual system than humans, and photographs are constructed for human vision. Given birds different perception of e.g. colour in real objects and of colours in pictures (Delius et al., 2000), the pigeons might fail to see any correspondence between more visually demanding stimuli, in the very same experimental setups. That the use of photographs of people, or transfer between photographic and real space, fails in certain studies is thus not surprising. Transfer from objects to pictures, and vice versa, usually breaks down the more complex, or refined, discriminations have to be (Delius et al., 2000). This breakdown can be seen as a failure of reality mode to kick in, and the limitations of working in surface mode in experiments that demands recognition.

6.3 Monkeys

Many studies use pictorial stimuli without a need for the subject to recognise the content as being similar to a referent. Such pictures are sometimes abstract (e.g. judgement of same/different arrays in Wasserman et al., 2001), but also photographs are used (e.g. serial list learning in Terrace, 2005). In the latter case one can suspect that the experimenters assume that photographs, in virtue of offering recognisable motifs to attention, add ecological validity to a task, or make it more intuitive etc. This is not necessarily the case. But here I will rather focus on tasks that require recognition in order to be solved as the experimenter intended.

6.3.1 Conspecific recognition and social responses

Most research that actually requires that monkeys recognise pictorial stimuli seems to have regarded social cognition. Pascalis et al. (2000) gives a review of studies where monkeys and apes show differential recognition of conspecifics, and even individuals, in photographs. Fourteen studies between 1974 and 1999 showed positive identification of conspecifics.[55]

An example is Demaria and Thierry's (1988) study of responses to photographs in stump-tailed female macaques (*Macaca arctoides*). The macaques looked longer at pictures of their own species overall, but they also preferred to look at males of rhesus macaques, even more than either males or females of their own species. Within their own species they specifically looked at other females with infants, and not infants or females alone.[56]

Among other groups of stimulus animals than primates, felines elicited the most looking time, especially if the eyes were facing front. No social responses towards the pictures were reported other than the note that facial expressions were too rare for analysis.

Kyes et al. (1992) found good evidence that long-tailed macaques (*Macaca fascicularis*) recognised photographs of a gorilla mask and an unknown human as threatening social stimuli, while being uninterested in random colours or photographs of apples. Pictures were presented as back-projected slides on one wall, which thus might have allowed grasping for apples, which the macaques did not do.[57] To the social stimuli dominate subjects responded with aggression and the subordinate subjects with submissive responses. Thus, the slides of the gorilla mask and the human were perceived as threatening.

While showing similar behaviour towards pictorial content as towards real objects is the defining character of a confusion response, fear responses might not be the most suitable ones to go by. It is likely that fear responses, or other highly emotional responses, are not sufficiently under control to be inhibited regardless of perceptual mode. This is one of the reasons it seems that reality mode can work concurrently with a pictorial mode. People with e.g. spider phobia can find photographs of spiders unnerving despite a mature differential attitude towards pictures in general. However, such people do not stomp on pictures of spiders. The most dominant

---

[55] The studies of Sackett (1965; 1966), in Cabe (1980), suggest that responsiveness to social stimuli in photographic colour slides is innate in the sense that even macaque infants reared in isolation respond to them. However, they can hardly be recognised *as* conspecifics if isolation has been total.

[56] For macaques, other females with infants are generally highly interesting (Demaria & Thierry, 1988).

[57] Projected images arguably facilitate reality mode processing since they do not allow physical handling the way e.g. a piece of paper does. Indeed, Beilin and Pearlman (1991) found this effect for 3 and 5-year-old humans who were queried about the nature of depicted objects.

animals in Kyes et al. (1992) not only displayed threat signals, but lunged towards the images of the gorilla mask and the unknown human.

An absence of "confusion behaviours" does not mean that processing has not taken place in reality mode. It can rather be that depicted animals do not give off those social cues that usually elicit responses. Initial confusion can be brief, and then wear off when no social interchange occurs. However, this can be difficult to separate from an overall decline in interest. Overall interest in individual pictures, as measured by looking time, often wears of spontaneously when the novelty is gone (Humphrey, 1974). A lack of confusion behaviours in studies like these might just indicate that the sensitivity, the setup, or the report of the study does not capture those kinds of responses.

Rhesus macaques are sensitive to a familiar human's gaze in photographs and scan the image differently, as measured with an eye-tracker, when the person's eyes are fixed on the viewer or averted (Sato & Nakamura, 2001). Faced with a gaze the subjects targeted the eyes of the depicted person. This was especially strong for faces that displayed aggression. Similar findings were made by Gothard et al. (2004) for conspecific stimuli, but they also found that for photographs of macaques showing emotions, the subjects scanned not only the eyes but all features that contributed to the uniqueness of that expression, such as the mouth.[58] There was a strong tendency to look more at novel pictures, regardless of the emotional expressions depicted.

Both aggressive faces and averted (!) gaze elicit more responses in the macaque amygdale, measured by fMRI, when looking at photographs of conspecifics, than do appeasing and neutral faces, and gaze that is directed towards the viewer (Hoffman et al., 2007). Face stimuli elicited more response overall in the amygdale than did scrambled pictures, i.e. noise.

Rhesus macaques not only target eyes but also follow the direction of gaze in colour photographs, regardless of stimulus species (Lorincs et al., 2003; but see Fagot & Deruell, 2002 for baboons). If the depicted individual looks to the right, the subjects are more prone to look at the right side of the picture as well. For monkey subjects this effect was not affected by the presence or absence in the picture of a visible target object to which the depicted animal's gaze was directed. Adult human comparison subjects on the other hand looked significantly more in the direction of the depicted gaze when target objects had not been masked out. Interestingly, only a minority of human subjects at all spontaneously followed gaze direction in photographs. This is clear evidence that rhesus macaques can recognise animals in pictures and respond to them as they would respond in real life. Given, of course, that monkeys follow gaze in real life, which this experiment was designed to investigate.

In a go/no-go task[59] rhesus macaques where reinforced to respond to complex photographic scenes in colour containing conspecifics, as opposed to scenes without monkeys (Yoshikubo, 1985). Pictures were back-projected on a translucent screen.

---

[58] It is possible that the value of the mouth is lost when viewing photographs of humans, or that differences in eyes are informative enough.

[59] A go/no-go task requires a subject to respond to a specific type of stimuli, and refrain from responding to other types.

The same 80 pictures (of with 40 were positive stimuli) were used in all sessions. The long-term recognition memory for pictures in macaques is comparable to humans' (Ringo et al., 1986). Therefore a generalisation test was necessary. Transfer to novel photographs containing rhesus macaques interspersed among familiar pictures proved problematic at first, but with further experience generalisation to novel photographs took place. These difficulties do not necessarily indicate an inability to recognise the content of the pictures, only that the subjects did not yet respond reliably in accordance with the "macaque present vs. macaque absent" rule. They unequivocally showed that they had learned to respond to the concept of "rhesus macaque" when they later, without further training, discriminated scenes containing rhesus macaques from scenes with Japanese macaques (*Macaca fuscata*). No social behaviours towards pictures are reported in this study but it is a clear case of picture use where reality mode processing cannot be ruled out. All subjects were experimentally naïve before the study. There is little reason to believe that they had the chance to learn about the true nature of photographs from the task they were performing.

To check which body parts pig-tailed macaques (*Macaca nemestrina*) used to discriminate conspecifics in tasks like the above Fujita (1993a) used a preferential looking task with masked photographs shown on a television screen. It was found that no single feature was critical but that the combination of masked face and tail lowered interest in the pictures markedly. Two of the three monkeys preferred to look at their own species rather than Japanese macaques. This was true also when stimuli were masked. Due to a small number of subjects the findings are difficult to generalise, but preference for looking at photographs of one's own species has been demonstrated for several macaque species (e.g. Fujita, 1987; Fujita et al., 1997). For some, but not all species, this preference seems to be dependent on rearing history (Fujita, 1993b).

The examples above of preferring to look at photographs of one's own species, specifically of females carrying infants, and showing aggressive or submissive fear responses to threatening faces, speak for a small difference between interests in the real world and in the pictorial world, and can be explained by a reality view of photographic stimuli. However, there are still reasons to be wary of the "Charlie Brown effect" in these kinds of studies. When pictures are used as object substitution, photos are commonly used, but there are a couple of studies that instead uses handmade stimuli in the form of drawings for the same purpose, in the below cases to simulate social stimuli.

6.3.2 Surface-mode in monkeys

Keating and Keating (1993) for example found that subjects that learned to sort faces of various primates, including humans, readily generalised to so called Identikit faces, which are composite sketches used to depict people from eyewitness accounts. Thus, the testing materials used were black-and-white shaded pencil drawings of human faces. The subjects were then trained to recognise one standard face. Eye-tracking revealed that the eyes, just as for photographic stimuli, attracted the most attention. However, removal of the eyes (brows, nose or lip) did not lower recognition, but modification of the eyes or brows did. Rotation, inversion and scrambling of the face disrupted recognition. It thus seems that the subjects used the spatial relations between features as well as the specific look of the eyes to identify the target face sketch. One finding is enlightening, however. One of the four subjects performed as well with a scrambled as with the standard face. The eye-tracking data revealed that by isolating an eye of the picture, regardless of where it was, the macaque could identify the target face.[60] It is reasonable that the subject has regarded the task as one of finding a particular eye rather than a face the whole time. This would mean that it might also had sorted faces in this manner during the rest of the study which implies that in the transfer from photographs to mug-shots it only had to recognise drawn eyes, and never whole faces. Idiosyncratic response patterns would speak against a reality mode based performance and for a feature mode analysis.

Dittrich (1994) used black-and-white drawings to test recognition of body shapes in long-tailed macaques (*Macaca fascicularis*). Dittrich's reason for using drawn stimuli instead of photographs was motivated by the ease by which individual features can be controlled between images. In addition Dittrich maintains that highly schematic stimuli has been used in studies of movement perception, and that infants are able to recognise caricature faces. That it might be hasty to generalise findings from the perception of movement to the perception of still images do not need to be expanded upon here, and the cited infant study (Tyrrell et al., 1987) is inconclusive. In a preferential looking paradigm one month old infants oriented towards novel individuals in photographs if they had been familiarised with a caricature, but they did not preferentially target a novel individual in the form of a caricature if they had been familiarised with a photograph.

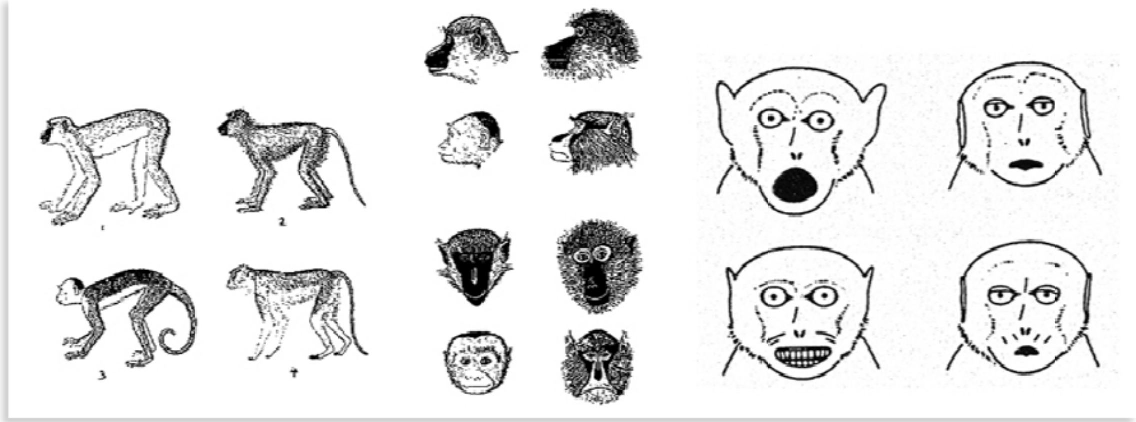


Figure 10. Drawn stimuli used in Dittrich (1994) (left), and Dittrich (1990) (right). From Dittrich (1994; 1990).

The macaques tested in Dittrich (1994) had to choose between four slide projections of monkey bodies. One of the species was chosen to be the constant target and

---

[60] Eye-tracking seems to be a most promising method for pinpointing individual strategies in experimental tasks.

the same four monkey bodies were used during the whole study, but with manipulations (*fig. 10, left*). During a training phase where pictures were held constant all subjects learned to discriminate the correct body picture in about 70 trials, much faster than for discrimination of geometrical or "non-natural" stimuli. In a transfer test reduction of picture size induced more errors than did rotation of the stimuli, but performance were still at 80% correct. When different features were removed from the bodies correct discrimination deteriorated differently for different features but remained above chance for most subjects and most features.

Taken together it was found that the torso can be judged to have been the most informative area. This strongly suggests that the subjects did not view the stimuli as pictures of monkeys. In a generalisation test to a frontal view of only the faces (the pictures had previously been presented in profile) three of four subjects responded above chance. This might, according to Dittrich, suggest that face recognition was involved, but the necessary controls of rotating other body parts or replicate the test with nonsense stimuli were not done.

The experiment was rerun, but this time with another of the four species as target stimuli in all manipulations. The findings were reproduced, but with one important difference. This time the torso was not the most critical feature, but the extremities.[61] Dittrich concludes that this difference was due to differences in the task, maintaining that in the second experiment the pictures were known and therefore had been comparable to discrimination of individuals, while in the first experiment the task had been to separate species. The former is done with a local strategy, and the latter with a global one, i.e. looking at body shape, according to Dittrich (1994). I would rather attribute the differences to the possibility that the subjects did not view these drawings as neither monkeys nor representations of monkeys, and simply found different parts of different pictures useful for discriminating them. That is, performance in a surface mode.

Dittrich (1990) also used line-drawings, but this time of faces (*fig. 10, right*). The task was for long tailed macaques to discriminate between facial emotional expressions. Again there was no detrimental effect by rotating the stimuli, or by changing colour, brightness or size of the stimuli. Again local features proved important, notably the facial outline followed by the eyes and mouth, but also configurational processing was involved since scrambled faces were less successfully discriminated. However, claims that recognition of emotional expression took place are not supported. No transfer trials to novel exemplars of the expressions were given, other than distortions of the four original stimuli pictures. Also, the local features that proved crucial for discrimination did not contribute to emotional information in the drawings. The absence of an inversion also speaks against recognition of faces (see section 6.3.4).

D'Amato and van Sant (1988) remains the pinnacle example of performance on a level that works very well in surface mode. On a task that the experimenters intended to work on a completely different level, capuchin monkeys (*Cebus apella*) used the frequent occurrence of reddish patches in human clothes etc. to, quite suc-

---

[61] Amusingly, the target stimuli was the patas monkey (*Erythrocebus patas*), the world's fastest running monkey

cessfully, form a “person versus non-person” categorisation. Red flowers and watermelons were thus included as “people.” Naturally, this categorisation broke down when colour was controlled for. This effect can occur in all forms of pictorial testing, depending on which discriminative cues the subjects attend to and are subsequently reinforced for. Even when seeing that a picture depicts a person, one can judge this information to be less relevant than the fact that it wears a hat in a specific colour. But the reason the monkeys settled to attended to colour strongly suggests that the photographs were not seen as depictions of real entities. This is a more parsimonious explanation than that the life-world of monkeys is one predominantly of colour, rather than a world of objects.

Recognising species, or species typical features, in photographs seem not to be limited to macaques but can be extended to New World monkeys. In Neiworth et al. (2004) cotton top tamarins (*Saguinus oedipus*) demonstrated apparent categorical recognition of individual species, as well as primates in relation to other animals. Complex colour photographs displayed on a television screen served as stimuli. The test method was the “preference for novelty” paradigm commonly used with human infants. After familiarisation with a category (specific species or primates in general) exemplars from another category (another species or nonprimate animals) were presented together with novel pictures from the old category. If the subject preferred to look at exemplars from the novel category rather than at novel pictures, familiarisation on a categorical level had taken place.

Complex pictures (i.e. with natural backgrounds) and a diverse mix of exemplars, as to counteract response to invariant local properties, were used. Therefore surface mode processing unlikely accounts for the results. However, the first alarm bell should go off when considering that it is assumed that these laboratory raised subjects recognised photographic animals that they had never experienced before. And how diverse were the visual patterns if considering combinations of local properties? When experiencing a novel pattern of colours and shapes, it is reasonable that interest is evoked. The only potentially convincing indication that recognition took place, was that when familiarised with four species of monkeys, the subjects preferred to look at non primate mammals rather than at a novel monkey species or photographs of apes. But species and their visual patterns can be expected to covary.

When mammal pictures were contrasted to reptile photographs, on the other hand, looking preferences did not deviate from chance. This was interpreted as an avoidance to look at reptiles, or possibly a continued interest in novel exemplars of mammals. But it could also be explained by the fact that the visual patterns as such are more varied when using pictures from such a broad category as mammals, and that they therefore never really got familiarised. A necessary control for this type of study is to include pictures of nonsense shapes and record the relative interest for nondepicting stimuli.

6.3.3 Food and picture attention

One need to go through the trouble of making sure that the animals use the intended information in pictures, even when they perform well with them or have shown to recognise intended content in other studies. To take an example, Hamilton and Vermeire (1983) used colour photographs to study face perception in rhesus macaques that had had their brain hemispheres surgically split. Using a go/no-go procedure the subjects learned the discriminations of individual target monkeys quickly and generalised to new photographs, however with great individual variability. "The tests of generalisation used new but somewhat similar photographs" (Hamilton & Vermeire, 1983). No difference for monkeys using the left or the right hemisphere in the task was found. It was thus concluded that face perception in rhesus macaques is either not lateralised, or the wrong stimuli were used. However, the stimuli were not suspected to be inadequate in terms of being photographs, but in terms of being used for discriminating the identity of individuals rather than facial expressions. The latter has been proven to be a lateralised process in intact subjects (Hamilton & Vermeire, 1983).

Different strategies for classifying faces of humans and baboons were found between the two species in Martin-Malivel et al. (2006). In a go/no-go task Guinea baboons (*Papio papio*) and humans were exposed to grey-scale cut-out photographs of human and baboon faces masked in various random configurations of noise. Baboons based their category judgement on the area surrounding the eyes, while humans used a more global strategy of taking into account also the lower parts of the face. In fact, the eye areas were the most discriminative parts of the masked pictures and a computer simulation performed very similar to the baboons. Martin-Malivel et al. (2006) therefore doubted that the baboons had used the pictorial stimuli as versions of faces. Similarly, in Martin-Malivel and Fagot (2001) the shape of human facial cut-outs had been the salient aspect that the baboons used for discrimination. Fagot and Deruell (2002) failed to elicit gaze following to pictorial displays in the same animals.

So far we can conclude that different features seem to be selected by the monkeys in different studies, but also that there is individual differences between subjects in the same experiment. A comparison with visual scan data for live stimuli in experiments comparable to the pictorial procedures is greatly needed in order to see whether eyes are everything a baboon looks at when discriminating faces. It probably is not. As mentioned earlier in this section, e.g. Keating and Keating (1993) and Sato and Nakamura (2001) found that also rhesus macaques attended to eyes more than anything else. However, e.g. Fujita (1993a) and Gothard et al. (2007) found that pigtail macaques and rhesus macaques respectively, attended also to tails and mouths, and Lorincz et al. (2003) reported that rhesus macaques follow gaze in photographs.

There is an important difference between these two groups of studies. In the ones where it has been found that only eyes, or other single features, have been focused upon the subjects have been tested in a generalisation task where they have been specifically reinforced to select a target category of photographs. In the second group, spontaneous responses in preferential looking tasks and during mere exposure have been measured instead. In the generalisation studies reinforcement does not only seem to select behaviour, but also attention. Even when subjects can actually recognise the content of photographs, due to at least an initial reality mode approach, they turn to localised strategies rather than continue to use those strategies that true recognition would allow, probably because the minimal strategy works just as well. But perhaps also because reality mode wears off when one finds out that the stimuli is flat and lifeless. Add to that unusual manipulations of scrambling, blurring, or cutting-out, and processing in reality mode is likely to fail altogether, especially when one is continuously given food for responding to pictures as patterns.

MTS paradigms, although responses are food reinforced, do not seem as detrimental to picture analysis as are discrimination tasks. For example Kanazawa (1996) found that Japanese macaques attended also to mouth and eyebrow cues when matching facial expressions in monkeys. This was not directly observed with eye-tracking or by way of masking however, but statistically inferred from the correct and incorrect responses to specific stimulus configurations. Why MTS might affect attention differently is probably because relational properties seldom get redundant when similarity judgements need to be made. Given the specific matching required of course, a sample is often too similar to both of the comparison stimuli for a local strategy to do the trick. Also, if matching on a local level, the relevant cues might very well change for each stimulus display and one has to attend to the whole pictures, all of them, to isolate the relevant information. Because of this, global processing will make itself useful.

There are cases when also reinforced go/no-go discriminations seem to uphold a global processing and not distil local strategies. In e.g. Jitsumori and Matsuzawa (1991) rhesus macaques and Taiwanese macaques (*Macaca cyclopis*) were tested on colour photographs displayed on a television monitor behind glass. The task was to discriminate photographs that were right side up (positive, “go”) from those that were upside down (negative, “no-go”). In order to judge which orientation an object has, one must be able to identify the object together with recalling its typical orientation. The monkeys performed well and transferred to novel photographs of full frontal views of humans, full rear views of humans, and full views of human silhouettes in black and white.[62] However, no transfer occurred to close-ups of human faces, or to human faces that were displayed far away. The faces far away could have been too abstract for recognition to occur. They were small round objects devoid of context. An observation was made that could explain the lack of performance with close-up faces. One of the subjects showed strong fear responses towards them and avoided to respond at all cost. It should also be mentioned that there was great individual variation for which pictures the four subjects could judge orientation, thus the animals did not seem to recognise the photographs as objects with ease. In the same study pigeons showed good transfer only to novel stimuli that were similar to the training stimuli. It is likely that they performed solely in surface mode.

There are further examples where successful demonstration of recognition in photographs has occurred despite heavy reinforcement. Dasser (1987), using both discrimination procedures and MTS setups, showed that long-tailed macaques could

---

[62] This is a further indication that silhouette information is possible to process in reality mode.

recognise individual monkeys from other monkeys in unique transfer trials, even when the bodies were viewed from different angles and when the faces were not in the picture. Stimuli used were back-projected colour photographs on a Plexiglas screen. Pictorial processing probably occurred in reality mode, although no social responses towards the stimuli is reported. Dasser does not exclude that the macaques failed to realise the representational nature of the photographs, but that they clearly perceived their motifs.

Macaques can also seemingly judge "sameness" and "difference"[63] in a categorisation task involving colour photographs of familiar animals and objects (Neiworth & Wright, 1994). Transfer occurred to novel pictures but was better for objects that the animals had interacted with, than for ones that they had only observed. This indicates that although the animals were food reinforced they continued to recognise photographs in accordance with their real life experiences of the depicted content. That is, they did not revert to localised, or "surface," strategies. Another reason for this might have been that the test aimed at category and not instance identity. In within-category trials different views of the same objects were used. Rotation of familiar objects in multiple photographs highlights their properties as objects. Such presentation has the potential to promote processing in reality mode and to ease recognition.

The effect of experimental procedure on the perception of pictures is only a tentative explanation. The reason I blame reinforcement is that with a strong incentive to solve a task that effectively gives you food, *any* strategy that delivers food will be learned. Recognising motifs, or act on them, might not be what is being drilled, depending on the task, and arguably how easily one slips into a surface mode. I second a call in Martin-Malivel et al.'s (2006) for further research on how animals actually perceive pictures in laboratory settings. But more specifically, to put emphasis on the laboratory part of the question, a comparison between different experimental paradigms would be especially interesting.

6.3.4 Inversion effects

The analysis above of the effects of reinforcement can be applied also to the mixed results regarding inversion effects in monkeys. Does a subject recognise a stimulus turned upside down as readily as one facing the typical end up? It is generally believed that inversion effects can be attributed to the fact that certain expert stimuli, e.g. faces, are recognised by configural processing, and that the relations between features of a face get different from those stored in experience when the whole image is rotated (Vermeire and Hamilton, 1998).

However, this can be argued to be true for many local features as well, it is just a matter of where the line for holistic processing is drawn. Local features can be processed configurally within its own boundaries. Responding to eyes instead of faces, for example, might very well entail configural processing of the iris and pupil's rela-

---

[63] See section 11.3 for discussion of this paradigm.

tion to the sclera, while ignoring the eye's relation to the face at large. Nevertheless, the result when encountering a view which is incongruent with the perceptual strategy you normally use is of course that recognition is slowed down.

Humans seem to have a strong inversion effect for faces. Presence or absence of inversion effects for facial photographs has therefore been used to argue for the presence or absence of face specific visual processing in animals. Results have been mixed, with neurological (e.g. Bruce, 1982) or stimuli centred (e.g. Parr et al., 1999) explanations offered. As argued above, local or global processing strategies as a result of procedural variables is perhaps also worth considering, and notably the effect of reinforced drilling. Vermeire and Hamilton (1998) suggest that the number of exemplars during discrimination training might affect the processing mode, with localised strategies being the result of small stimulus sets. This might certainly be the case for some studies, but the stimulus quantity alone is unlikely to explain the effect.

To consider a few examples of studies that have drawn negative conclusions regarding inversion effects in monkeys. Martin-Malivel and Fagot (2001), where baboons were found to respond to cut-out shapes rather than faces, found no inversion effect. Neither could Bruce (1982), Rosenfeld and Van Hoesen (1979) and Dittrich (1990). The latter with line-drawn facial stimuli (see *fig. 10, right,* p. 94). What these studies had in common was a testing paradigm based on heavily reinforced discrimination.

Studies of monkeys that instead have found evidence in the affirmative have relied on other types of paradigms, such as preferential looking (e.g. Tomonaga, 1994; Swartz, 1982) or MTS (e.g. Parr et al.,1999; Phelps & Roberts, 1994; Overman & Doty, 1982). A notable exception is Vermeire and Hamilton (1998) who found inversion effects in rhesus macaques using a go/no-go paradigm, but their monkeys were split-brain subjects and comparison to intact subjects is complicated. Furthermore, inversion effects could only be shown for about half of the subjects. They found that the right hemisphere was more sensitive to inversion of macaque facial photographic stimuli than the left one, but also that macaques can rely on processing of local features regardless of hemisphere when needed, i.e. when recognition (as in reality mode processing) is not possible. "[...] the fact that the same stimuli made of exactly the same features are subjected to different modes of processing depending on their orientation suggests that it is not merely sufficient that the stimuli have both component and configural properties available to be processed, but rather that the perceiving organism finds itself unable to extract configural information from an inverted face and resorts to an analytical, component mode of processing" (Vermeire & Hamilton, 1998, p. 1012). This flexibility is an important reason for why findings from pictorial experiments have to be analysed as pictorial problems before generalisations to real life perception can be drawn. True recognition must be teased apart from apparent recognition because mode of picture processing will affect the processes underlying performance on the task and the conclusions that can be drawn from it all.

There are reasons to believe that the inversion effect is dependent on recognising the depicted stimuli as real objects.

Overman and Doty (1982) not only found that recognition of inverted faces were more difficult than upright images for pig-tailed macaques, but also that the subjects displayed social responses towards monkey and human faces. This response seemed to have been strongest at initial exposure. Naturalistic images, like scenery, flowers, birds and insects, did neither yield emotional responses nor inversion effects.

Using a MTS procedure to study memory Phelps & Roberts (1994) demonstrated inversion effects for primate facial black-and-white photographs in humans and squirrel monkeys (*Saimiri sciureus*), but not in pigeons. Humans but not squirrel monkeys also showed inversion sensitivity for photographs of outdoor scenery. Thus it is likely that the squirrel monkeys did not recognise scenery in photographs.

Parr et al. (1999) showed that the inversion effect is not face specific in rhesus macaques. The effect was not elicited by human faces and abstract shapes but by cars and faces of an unknown monkey species.[64] So why did the inversion effect manifest itself for monkey faces and cars, but not for human faces and abstract shapes? Can it be that human faces were not recognised as faces in the photographs, and/or that attention to local features was selected for in the food reinforced MTS procedure used? According to Parr et al. (1999) discrimination of human faces was acquired quickly compared to other stimulus classes, which is supposed to be consistent with a configural, or global, processing strategy. However, fast acquisition could likewise occur for the typical human eye or another more local feature which is highly invariant between photographs of human faces. After all, the food reinforced training procedure gave no incentive to actually recognise the content of photographs.

Chimpanzees in Parr et al. (1998) showed a more expected performance with the very same stimuli and method (MTS), with inversion effects occurring for familiar face categories, like chimpanzees and humans, but failing for capuchin monkeys, cars and abstract shapes.[65] However, responses became notably less accurate across all categories during inversion trials, suggesting configural processing.

The real question is why cars did evoke an inversion effect for the macaques above. Further studies of this phenomenon must take into account whether the subjects recognise the stimuli as objects or not, and whether they retain this recognition across the whole experiment. Perhaps it will be found that faces happen to be extra salient motifs in pictures (several studies to this effect has been presented in this chapter) and are sensitive to inversion for this reason, as are other recognisable objects, although they might be novel, like cars. Studying the inversion effect with real objects ought to be a given control.

---

[64] That known objects, or "expert categories", are more sensitive to inversion has been shown for humans (Diamond & Carey, 1986) and chimpanzees (Parr et al., 1998; Parr & Heintz, 2006) but was not supported in Parr et al. (1999).

[65] The study fails to tease apart the face specificity of inversion from the familiarity effect since no pictures of well known non-face objects were included.

Chapter 7

Apes in indirect picture tests

7.1 Inversion in apes

Tomonaga (1999) showed that cut-out photographic human faces but not houses elicited inversion effects in chimpanzees. Similarly, in their study on recognition of rotated black and white photographs in chimpanzees, Parr and Heintz (2006) found an inversion effect for photographs of unfamiliar conspecifics' faces, but not for frontal views of houses. A sequential identity MTS procedure was used, which means that the sample and the match were the same picture, although it had been cropped differently to preclude matching based on figure shape. The photographs were displayed on a computer screen behind an extra front of Plexiglas, 30 centimetres from the mesh through which the chimpanzees controlled a joystick. In this study a significant linear impairment for five rotational angles was found, with the worst performance for full inversion. Recognition of houses also decreased with rotation, but non-linearly. Performance with chimpanzee faces was only significantly worse than for houses at the fully inversed rotational angle.

There was an assumption that blurring the pictures would increase sensitivity to rotation since the value of configural information rather than feature recognition would increase. Blurring of the comparison photographs (the sample always stayed focused) impaired recognition linearly, but only the maximum level of blurring decreased performance significantly. Blurring did not interact with rotational angle. This either suggests that blurring does not increase configural, or global, processing, or that the connection between sensitivity to rotation and configurally based visual processing is not as straight forward as had been assumed. Without a combination of eye-tracking and inversion in the same test, the claim that a lack of inversion effect is indicative of a feature based processing is unfounded. Gothard et al. (2004) could show that macaques scanned photographs of macaque faces in the same face-specific manner regardless of orientation, thus showing recognition. The same method should ideally be used with content that is not sensitive to inversion, using individuals that show inversion effects for e.g. faces.

If the houses were indeed processed on a non-configural level, the rotation data, according to Parr and Heintz (2006), might hint at which features were used by the chimpanzees. Performance was almost identical in the upright, 90 and 180 degree conditions, and only when displayed in a diagonal direction did the success rate drop significantly. However, rather than a feature based recognition, it is as likely that a configural one was used, only one not based on recognition of the content as such. After all, a house is a coherent figure against a ground and chimpanzees have been shown to have a global preference for coherent patterns (Fagot & Tomonaga, 1999; Hopkins & Washburn, 2002). The configural combination of local features was just one based on something that did not transfer well to diagonals, perhaps relations between specific squares and rectangles (windows and doors). Again, eye-tracking, and to a lesser degree masking trials, would add enlightening information. The chimpanzees in Vauclair et al. (1983) (see Chapter 10) had a poor reproductive memory for simple diagonal patterns as opposed to horizontal or vertical ones. It is possible that rotation to diagonal displays in a sequential MTS task pose a similar problem for working memory.

That being said, chimpanzees seem more sensitive to the (perceived) separation of elements than are humans, which in turn controls a local or global processing of geometrical patterns (Fagot & Tomonaga, 1999; 2001). If, after all, the features of the houses, due to not being spontaneously recognised as neither a house nor a coherent object, were seen as disjoint elements, the research of Fagot and Tomonaga would instead predict that houses were indeed processed on a local level, and support the conclusions drawn by Parr and Heintz (2006).

From a pictorial perspective a reasonable explanation for the difference between "expert categories" and lesser known categories in e.g. Parr's research on inversion effects is first of all a difference in recognition of the depicted objects. Only second to that can the influence of category expertise be studied. Chimpanzees with little or no experience of houses do not recognise houses in small black and white photographs. Houses do therefore not necessarily even look like objects in photographs, but patterns. The same can be argued for houses in Tomonaga (1999) and cars in Parr et al. (1998). However, macaques in Parr et al. (1999) did show inversion effects for the very same cars, which might mean that cars look like real objects to macaques but not to chimpanzees in photographs. Whatever the case, the expert category hypothesis, originally developed for humans (Diamond & Carey, 1986), cannot be explored in nonhumans until the relationships between different picture stimuli and the different analytical processes they evoke have been settled.

Tomonaga (2007) agrees that faces are special. Using a visual search task he demonstrated that a chimpanzee, Chloe, at the Kyoto Primate Research Institute significantly faster located upright human faces among inverted and horizontal ones, than the other way around. This extended also to dog faces and cartoon caricature human faces. Photographs and cartoons were all in greyscale. No upright superiority effect[66] could be found for photographs of chairs and human hands. (What is the upright position of a hand is not explained.) The only indication that caricature faces were perceived as (representations of) faces is that they grouped with the response time for realistic faces of humans and dogs, and not with hands or chairs.

In each trial the search array had been composed of four or ten identical pictures, of which one exemplar differed in orientation from the rest. The task was to pinpoint the oddly oriented picture. Had the pictures in the array been different also in content, Chloe would have needed to decode the content of the pictures before be-

---

[66] The upright superiority effect is the inverse of inversion effect.

ing able to judge which one was the odd one. Now invariant pattern cues could not be ruled out. However, with such a setup the response time would not be comparable across trials since one would not know what respective contribution the rotation and the picture content would have.

So why did the caricature faces stand out? We do not know the pictorial background of the 17 year old Chloe, but she had been with the institute since she was 4 years old and had in the past participated in many computerised matching and visual search tasks using pictorial stimuli. It is possible that the caricature faces were seen as faces somehow, but a direct test is needed. Basing this conclusion on a test that was designed to test the face specificity of the upright superiority/inversion effect is far too roundabout and uncertain. That said, if the upright superiority effect is indeed face specific, the face caricatures could show an inversion effect because they showed facial configurations, regardless of being recognised as faces.

To test whether face-like configuration, rather than faces themselves, invoked the effect Tomonaga (2007) tested extremely schematic facial stimuli containing a circle for head shape, three identical horizontal lines for eyes and mouth, and a similar vertical line for nose. This was compared to pictures where the internal features of the schematic face were scrambled. Chloe made significantly fewer mistakes finding the differently oriented picture when the internal relations were face-like than when they were scrambled, and the response times indicated a very strong inversion effect for the face-like stimuli. The face specificity of this effect is not entirely convincing, though, since the necessary controls were lacking. Trials with a figurative, but not face-like, configuration should be tested in addition to the arbitrarily scrambled configurations. It should be more difficult to find the oddly oriented scrambled picture than an orderly one, regardless if that order is similar to a face or not. This can account for the difference between face-like and scrambled stimuli when it comes to percentage correct responses, but it does not completely explain the inversion effect for face-like stimuli. If a figurative, or orderly, configuration was all that was needed, finding an inverted face among upright ones should be as easy as finding the reverse. But performance was different for the two conditions.

There was little doubt that Chloe used configural processing however. Her performance dropped severely when only parts of the face was shown, but remained good, as well as showing upright superiority effect, when the eyes plus one other feature (mouth or nose) was shown. Also eye areas on their own yielded good results, but when eyebrows were excluded her performance dropped significantly. Thus the relation between eyebrows and eyes were an important configuration for recognising orientation.

7.2 Spontaneous preferences

At the Primate Research Institute in Kyoto adult chimpanzees, regardless of rearing history, preferred to choose colour photographs of humans among other primates on a computerised free-choice task. They were food rewarded regardless of choice. Two of three juvenile chimpanzees also preferred to indicate pictures of humans, but not until 5 years of age. At four they choose without preference. The third juvenile chose chimpanzees significantly more often than humans or other primates, starting by the age of four (Tanaka, 2007a). The conclusion put forth in the report, that the chimpanzees preferred pictures of one species before the others (as a function of rearing history), just because they touched those pictures more often to receive their reward, is a bit strong. The chimpanzees did acquire a habit of selecting a specific category of pictures, but their reason for this is not transparent. This test does not even prove that the chimpanzees recognised the content as primates, but it is a most parsimonious conclusion given what we know from other studies that use photographs, as well as the subjects' extensive experience of being tested with pictures. That ten chimpanzees settled for the same category is not easily explained by them making a choice based on invariant features decided independently by each subject.

In a similar study Tanaka (2003) did actually invoke the conspicuous human gaze as a factor that five independent subjects could all converge on, but it still entails that at least the gaze was recognised in the pictures. Also *with* recognition there could be other reasons than species category that singled out humans, for example their colourful clothes. When colour was removed the preference for humans did indeed lessen, but not below significance. A finding in Tanaka (2003) was also that chimpanzee photographs were chosen significantly more frequently than gorilla photographs. However, this finding was not replicated in Tanaka (2007a). In both studies individual preferences varied substantially.

It seems that preferential choice is not a good method for studying picture recognition as such. Preferential looking is somewhat more promising but only conclusive if there is a stable preference that generalises to many different exemplars. In Fujita and Matsuzawa (1986) Ai the chimpanzee, at the time 7 years old, looked significantly longer at colour photographs that contained humans than at other complex photographic scenes that was devoid of humans.

In a study of the development of face recognition Myowa-Yamakoshi et al. (2005) demonstrated that chimpanzee infants between four and eight weeks of age showed a preference for looking at their own mothers' cut-out facial photographs. This preference disappeared at eight weeks but infants still continued to track pictures in the same manner as during preferential looking. Before 4 weeks of age they showed very few tracking responses. Human infants in comparable studies show discrimination between individual faces at about roughly the same age. Whatever they see in their mothers' faces they seem to be able to see in photographs at this age, or at least enough of it. Chimpanzees did not show recognition of photographs of human faces at this age however, as measured by preference for known versus unknown humans (Myowa-Yamakoshi et al., 2005).

When being between 10 and 32 weeks old the same three infants as above showed a preference for viewing photographs of human faces directing their gaze at the subjects, than faces that displayed an averted gaze or closed eyes (Myowa-Yamakoshi, 2003). This preference did not transfer to scrambled faces why it is likely that the faces were processed globally, and that photographs are viewed in reality mode by infant chimpanzees.

7.3 Categorical discrimination

As a way of exploring categorisation in the chimpanzee Viki, Hayes and Nissen (1971) report the continuation of the generalised discrimination problems presented in Hayes and Hayes (1953). Although not targeted at investigating pictorial competence per se, recognition was a requirement for successful performance.

Testing was done using a Klüver board as in the Hayes and Hayes (1953) study. Two wells were covered with pictures. One of the wells was baited with food. The same category was always rewarded and the presentation was randomised. The results given (*table 2,* p.108) are generalisation trials after Viki had discovered the categories to sort by, thus only showing her further application of a given discrimination rule. Pictures were only used once.

Four human subjects of the same age as Viki performed above chance on the same stimuli in all problems but the active versus the inactive people condition. Hayes and Nissen (1971) ascribe this to the limitation of static pictures to depict activity. Reading activity into pictures might be a late development in both humans (Friedman & Stevenson, 1980) and ape. It remains to be tested if this component of pictorial competence at all can appear in animals. Dynamic content of *some* sorts, e.g. the recognition of living individuals, seem to pose no problems in even reality mode picture processing.

Verbal reports by the children suggested that it was seldom categorisation as such that posed problems but to remember which option had been correct in the previous trials. This probably held true for Viki as well. In addition she tended to become careless when a task required close attention. "Caution in selection was not one of Viki's usual working characteristics" (Hayes & Nissen, 1971, p. 85).

There is little reason to believe that a chimpanzee, albeit raised in a human context, should form the same categories as an adult language competent experimenter. This alone is a reason to not draw too far-reaching conclusions regarding Viki's outstanding results. An additional concern is that Viki in other studies had shown a capacity for single-trial learning, as well as simultaneously retaining in memory up to 10 concurrent discriminations for object pairs (Hayes et al., 1953b). This means that she could remember which one of a stimulus pair that was correct on ten successive but independent problems.[67] In the picture categorisation study this competence of Viki's made it difficult to control for choice by invariant cues rather than choice based on concepts, even when switching types of pictures.

---

[67] Some chimpanzees in Hayes et al. (1953b) could retain 20 concurrent discriminations.

+ Target	- Target	n	% corr.	Type of picture
Animals	Objects	41	85	Realistic colour
Men	Women	65	67	No details given
Red	Green	43	74	Nonsense designs, line drawings, realistic with a predominant colour
Children	Adults	47	89	Realistic, mostly colour
Circles	Crosses	40	80	Drawings
Larger	Smaller	25	84	Nonsense designs (matched shape and colour)
Larger	Smaller	54	55 (n.s.) *89*	Nonsense designs (mismatched shape and colour) *With ratio > 2:1*
Quiet people	Active people	34	61 (n.s.)	Realistic pictures (e.g. sleeping, playing baseball)
Complete drawings	Incomplete drawings	37	79	Black-and-white drawings (e.g. dog without legs, hand missing finger)

Table 2. Generalised discrimination study on Viki (*Pan troglodytes*, 5 years) in Hayes and Nissen (1971)

Hayes and Nissen (1971) propose that an alternative explanation for Viki's results with the pictures might be that she continuously changed the cues she worked by, or learned concurrent ones, which could then be generalised to novel pictures. In for example the gender condition she could have started out by basing her categorisation on clothing cues in the full body pictures, then switch to hair length in the head shots, *independently* from the fact that certain clothes and hair styles have a gender component in common. She might never have thought of applying the gender concept to the problem at all, but still be rewarded with food.

Nevertheless, to single out "clothing" or "hairstyle" some sort of apprehension of the pictures' motifs is necessary, but not necessarily in a pictorial mode. Reality mode would work for the photographs and perhaps some of the line drawings (i.e. silhouette ones). Even surface mode would suffice if "clothing" and "hairstyle" were substituted with patterns and colourations. The complete versus incomplete drawings, though, seem to again demonstrate Viki's ability to recognise not only photographs but quite abstracted drawings as being depicting pictures. But unfortunately are descriptions of the stimuli and her performance on individual pictures not given.

A discrimination task was used by Vonk and MacDonald (2002) to test categorisation of natural concepts on "three different levels of abstraction" in a juvenile gorilla. The levels of abstractions: concrete, intermediate and abstract, mirror those of Roberts and Mazmanian (1988). In this classic experiment humans, pigeons and squirrel monkeys had to discriminate photographs of kingfishers from other birds, birds from other animals and, lastly, animals from non-animals. Human undergraduate students managed to discriminate on all levels of abstraction while pigeons and monkeys performed best on the most concrete, i.e. choosing kingfishers among two pictures of birds. With further training both species could also discriminate novel photographs of animals from photos of inanimate objects. Performance on the intermediate level on the other hand (birds from other animals) never exceeded chance.

Using the discriminations orangutans/gorillas versus humans, primates versus other animals, and animals versus foods/objects/sceneries, Vonk and MacDonald (2002) mirrored also the results of Roberts and Mazmanian (1998). Zuri the gorilla performed best on the concrete and abstract "levels" and more questionably on the intermediate one. Stimuli used were photographs, mostly in colour, of many different views of the animals and objects, shown on a computer touchscreen.

At the discrete level (orangutans/gorillas versus humans) discrimination could occur simply by way of recognising invariant visual properties of a single species. Not even recognition of the species would be necessary. But the authors, making a thorough analysis of properties across pictures as well as reinforcement history, could not find any causes external to the categorisation of the content itself that could account for the successful transfer of Zuri to novel pictures. For example, a test that controlled for the conspicuous colour of orangutans did not reveal an effect of colour on Zuri's discriminations. It is therefore safe to assume that Zuri recognised content in most of the photographs and made her judgements accordingly.

The question, however, is *which* content Zuri ascribed to the pictures. Zuri had limited experience of other animals, and no real life experience of orangutans at all. The pictorial versions of orangutans can thus count as a class on its own, recognised as a bona fide natural class due to the photographic quality of the stimuli, but not necessarily related to the world outside of computer screens. Zuri's fastest acquisition was in fact the discrimination between familiar foods and animals, the former being a category that she *did* have extensive experience with and that furthermore is more faithfully reproduced in photographs, in terms of size and static nature, than are animals, scenery etc. Worth noting is that food was the positive target stimulus, so in order to discover how to solve the task Zuri had just to continue touching that which she intuitively preferred, and realise that she was differentially rewarded for this.

Here two versions of reality mode processing become evident. One is taking the content at face value, as a novel real entity. This is likely to happen to pictures of objects and animal types that one has no real-life experience of. Since all experience is from the pictorial version some properties of the real instances is never evoked in the viewing mind, such as the movements of orangutans, their size, or smell. The depicted objects have, more or less, the properties displayed in the static image and nothing more. The second type, recognising well known entities, has the potential to evoke some of the knowledge of these entities which are not directly captured in the picture.[68] Expecting a photograph of an apple to smell, expecting a known human to react when you call to him or her on television, etc., are examples of this phenomenon. To a pictorial novice the case of well known objects affords a bigger chance to realise both the referential nature of pictures, and their differences from the real world. However, due to unfulfilled expectations, recognition of this type should also be likely to fail, or break down.

For Zuri, most difficulties were posed by the intermediate level of abstraction: primates versus other animals. This is because there are no single, or few, visual commonalities among primates that excludes all other animals, especially mammals. There is rather a mosaic of traits, unevenly distributed across primate species, that constitutes the category. Should one expect a 4 year old gorilla in a zoo nursery to have, or be able to discover and apply, a concept for "primate" (prosimian, monkey and ape) as opposed to "animal" (mammal, bird, fish, insect etc.)? Especially if she is working in a reality mode, where there is little room to infer extra information to the pictures. It is difficult to draw any conclusions about Zuri's categorisation abilities from these results. After all, birds in pictures do not fly, fish do not swim, and insects are not small, fast and crawling. If there is no real-life experience of the involved species, Zuri is even more confined to using only visual cues. To give Zuri a fair chance of categorising on an intermediate level, one ought to use real objects instead of photographs of living beings.

That said, from the data it is safe to say that whatever Zuri could retrieve from the photographs was enough to effortlessly solve the "concrete" and "abstract" discriminations, and with limited success the "intermediate" ones. Vonk and MacDonald (2002) conclude their study with, rightfully, questioning the value of dividing abstraction levels in terms of only the "breadth" of the categories that are to be discriminated. Ease of categorisation seems rather to be dictated by the ratio of feature overlap within and between specific categories. In this sense the "intermediate" level is often by far the most abstract one. (Compare "primates" with "mammals.") I would hasten to add that the feature overlap within and between categories is also dependent on the interpretation the subject is able to make of the stimuli that is being used.

That categorisation is dynamic and varies with the stimuli and subjects tested is supported by a replication of the Zuri study with six orangutans (Vonk & MacDonald, 2004). All but one subjects were adults, touchscreen was used, and except for black-and-white pictures intended to control for the conspicuous colour of orangutans, all photographs were in colour. In this study it was found that the discrimination of primates versus other animals was *not* significantly more difficult

---

[68] Even more properties can of course be inferred when viewing the picture in a pictorial mode, where imagination "fills in the gaps." It is possible that e.g. orangutans seem even *more* real in that kind of mode than in reality mode.

than orangutans versus other primates/humans, or animals versus non-animals/foods. In fact, animals versus non-animals was the more difficult discrimination for the orangutans, although, just as in Vonk and MacDonald (2002), food versus animal discrimination seemed to be easily acquired.

On all problems large individual differences were found. This might reflect not only individual differences in learning, but in interpreting the stimulus photographs as well. It is feasible that some pictures made sense so some subjects but not to others. Interestingly, the adult subjects did not perform unequivocally better than the 2 year old orangutan, which in turn performed better than the 4 year old gorilla. General exposure to pictorial material during one's lifetime, which is unavoidable in captive settings, did not seem to have influenced pictorial ability. Experience with the species and objects in the pictures did not seem to have affected performance *overall*, but is likely to account for some of the variation between individuals.

Using the discrimination task in a touchscreen setup as above, Vonk (2002) did also test concepts for social relationships. Two of the orangutans and the gorilla Zuri were tested on their ability to choose mother-infant pairs as opposed to individuals in other types of configurations, as well as single individuals. Positive and negative choice photographs depicted the same species. In this task orangutans and gorillas were used as stimuli in separate sessions, while unfamiliar species were placed in mixed stimuli sets. In training sessions all subjects performed on par with training in the previous studies, reaching criterion of 80% correct in 20 to 130 trials.

Only one of the animals, an orangutan, showed good transfer to novel photographs. The two remaining subjects were above chance but did not reach criterion in the first transfer sessions. In subsequent sessions all three subjects performed well on novel photographs of unfamiliar species. The orangutans, but not the gorilla, were thrown by a photograph of an adult male gorilla with an infant. It was categorised as a mother-infant pair. Zuri, although reared by humans, had experienced conspecifics of both sexes. This episode highlights the necessity of experience with the depicted stimuli, as well as being able to transfer that experience to its recognised version in a photograph. It was also found that Zuri sometimes preferred to choose gorillas of her own age, regardless of context. This effect did not transfer to other species. This is another example of how the real world can affect the pictorial world. Perhaps especially when viewed in a reality mode. When pictures are truly pictures it makes little sense to prefer to indicate one before another, especially in a non-communicative context.

7.4 Categorical matching

Vonk (2002) continued her investigation of social categories in the two adult orangutans and the juvenile gorilla by switching to MTS, more specifically a delayed matching-to-sample procedure (DMTS), which they had previous experience of.

This step was made in order to reduce the possibility of making visual, rather than conceptual, identification of the correct choice picture. When several different social concepts are tested at once, of which one is being determined by the present sample, and this has to be stored in memory before the choice stimuli appears, there is little use in relying on a purely visual strategy. Different features become useful depending on the specific discrimination that has to be made and, in addition, this has to be accomplished with novel photographs. This has implications also for the pictorial recognition that is necessary. In order to appreciate a social relationship in a picture one must recognise the individuals that form this relationship. Patches of e.g. red and oblong shapes could in theory form relationships in the discrimination version of the task described above. However, it is a farfetched assumption, and it is even less probable in a conditional matching task.[69]

The four matching principles in the present study were mother-infant pairs, social groups, mated pairs, and siblings. In each trial the sample photograph displayed one of these configurations and after a brief blackout period a matching photograph of the same configuration and a non-matching photograph of another configuration came onscreen. Species were mixed across the stimuli set, i.e. a group of gorillas could form a match with a group of birds to the exclusion of a sibling pair of chimpanzees. The subjects were, as usual, rewarded for the right choice with food.

All subjects performed above chance overall within 30 trials, which is exceedingly fast. However, they differed in the matching rules they were able to acquire. All three performed well on matching pictures that contained a mother with her offspring or two siblings. But only one of the orangutans could match on the concept "social group" and the second one on "mated pairs." That the gorilla was able to match only pictures that contained young individuals could perhaps be ascribed to her preference for such photographs rather than matching in accordance with social concepts (Vonk, 2002). Whether the sample and the match contained the same species or not did not seem to influence any of the subjects, which strongly suggests that matching did take place on more than on a purely visual surface level.

Although Vonk (2002), beside ages and sexual dimorphism, mentions the perceived activity in the photograph as a clue for interpreting the social relationship between the depicted animals, this might not be possible for all pictures and activities. For sure, a play activity hints to a sibling relationship and a mating activity hints to a mated pair, but this entails that an otherwise dynamic activity can be read into static pictures. That this can be done by a given subject is not necessarily the case, although it is certainly a possibility for many activities. In pictures playing is for example often accompanied by play faces, and, as shall be made clear below, reading emotion into photographs seems to be possible even for pictorial novices.

Furthermore, if one is able to identify individuals in photographs one is automatically inclined to make out what activity they are involved in. There is a drive for a complete perception, but what one comes up with might differ depending on which mode of picture perception one is working in. "Frozen in an awkward position" is perhaps a common conclusion within a reality mode. However, it is premature to view degree of perceived dynamic activity in pictures as a defining feature for any of the modes, at least for photographic stimuli, before studies targeting this specific question have been made.

One of the orangutans and the gorilla were indeed tested on further dynamic content in a third experiment in Vonk (2002). Asked to match in DMTS according

---

[69] One need not assume that the specific concept that is intended by the researcher, e.g. "mother," necessarily is the relationship perceived and acted on by the subject either.

to the concepts “sleeping,” “grooming,” “eating” and “playing,” the gorilla matched readily by the first session and for all categories. The orangutan took three exposures to the stimuli (there was no transfer test in this setup) and performed above chance only on “sleeping” and “eating.” “Sleeping” was also the most successful category for the gorilla. That “sleeping” was marginally easier than “eating,” “grooming” or “playing” might be due to the fact that one does not have to read very much into a picture in order to see the similarity between two scenes that depict sleeping individuals.

The photographs were in no way simplified by reduction of irrelevant backgrounds, or chosen for their prototypical looks. They were rather intentionally made more difficult to discriminate without thorough interpretation. They were for example balanced so that no single feature was unique for a particular category, and species were again mixed. One could thus expect to find playing individuals with closed eyes or grooming individuals with play faces. No data is given for performance on individual pictures so we do not know if certain manipulations rendered them more difficult than others.

It is noteworthy that overall performance with activity photographs seems to have been slightly better than performance in the experiment that measured concepts for social relationships. Perhaps this reflects that social concepts are more farfetched for a subject to attribute to pictures in an experiment than are individual activities. Whatever the case it is noteworthy that subjects without much experience with photographs, and definitely no social communicative experience with pictures, are quite able to attribute some dynamic content to them. This illustrates the power of pictures, i.e. to evoke one’s knowledge of the real world from exemplars that mirrors only some of that knowledge. But since these studies do not target the question of differentiation and reference, this power can be attributed to reality mode processing.

It can easily be argued that when matching photographs of different objects, or photographs of objects to different views of the same object, reality mode processing cannot be excluded. The task requires neither differentiation nor reference. The same is of course true for matching *identical* pictures where there is even no need to recognise the content of the pictures at all and pure surface mode processing would in theory suffice for matching to occur.

Categorical matching on the other hand, as shown above, can often exclude surface mode processing. A further example is Tanaka (2001). Five chimpanzees at the Primate Research Institute in Kyoto, experienced in both MTS and picture tasks, could categorise colour photographs of flowers, trees, weeds and ground (dirt etc.) on a categorical level. However, when the comparison stimuli were all (there were four of them) from the same category as the sample the subjects performed significantly worse than when only the match was from the same experimenter-defined category as the sample. This means that the subjects had problems, either conceptually or visually, to pinpoint among similar exemplars the one that was the closest match to the sample. However, these mistakes on a computer screen do not necessarily mean that the same mistakes would be made in real life. Distinguishing one type of plant from another is a crucial ability for a foraging animal. It might rather mean that the grounds for matching photographs did not work on a detailed, e.g. plant species, level.

Another experimental paradigm that address categorisation, often with the help of pictures, is to teach subjects the concepts of "same" and "different," and then let them apply these on groups of stimuli, be it objects or pictures of objects, creatures etc. There are two versions of this paradigm. Same can mean "perceptually identical," in which case there is no need at all to see what a picture depicts, only to discriminate patterns from each other. An example is Wasserman et al. (2001) where baboons judged arrays of clip-art to contain either the same or different images. Another sense in which "sameness" is used is to mean "categorically identical." In this version you must avoid using identical stimuli. We have seen several examples of this type of similarity in experiments already. (It will be further discussed in section 11.3.)

The problem with the concepts "same" and "different," is that one cannot be sure of what exactly it is that is judged to be the same or different in the view of the subject. There is also little reason to assume that the criteria for sameness or difference are stable across entire experiments. In theory the subject can use novel criteria for each set of comparison stimuli.

Brown and Boysen (2000) argue that categorisation experiments that involve nondifferential reinforcement more accurately reflect subjects' natural categories, while reinforced paradigms on the other hand can induce a specific categorisation during the course of testing. This is certainly true, but even without differential reinforcement the nature of the task is bound to structure the categorisation that is applied. Just being exposed to stimuli pitted against each other forces one to discriminate between them, and thereby categorise, in some way or the other. This becomes further removed from "natural" categorisation when pictorial stimuli are used, where properties that would perhaps otherwise be used for categorisation is not captured in the picture.

That a pictorial cat is judged to be similar or different from a pictorial tiger does therefore not tell us what the subject thinks about the relationship between real cats and tigers, only that they found some basis to judge similarity or difference in the discrimination made. That a specific pair consisting of a gorilla and a chimpanzee is unequivocally judged by five chimpanzee subjects to be "the same" in Brown and Boysen (2000) does not mean that this particular gorilla and chimpanzee would not be seen as very different entities in real life. In fact, the reason they were judged to be "the same" was, according to Brown and Boysen, probably that they were sitting in very similar poses. Other chimpanzee – gorilla pairs were not categorised as "the same." If the same two animals were to sit alike in a field outside of the subjects' enclosures they might not be judged to be "the same," but "different," or "the same" but for completely different reasons. Maybe their colour was the same at that occasion. If this discrimination had been preceded by several trials on species discrimination there is a chance that this criterion would transfer also to the next pair, but with nondifferential reinforcement it is just as likely that it would not. The gorilla and the chimpanzee that were sitting alike in Brown and Boysen (2000), had in fact been preceded by several trials on chimpanzee – gorilla discriminations where the subjects had, seemingly, responded categorically according to species.

The five chimpanzees (there was a sixth naïve control subject) in Brown and Boysen (2000) had learned the concepts of "same" and "different" previously to being tested with photographs of animals. They were for example able to judge Arabic numerals and arrays of dots as being either the same or different. It is not said if this is the only same/different training that they had had, neither what their previous experience with pictures were. In the present experiment they were required to judge if colour photographs of house cats, chimpanzees, gorillas, tigers and fish were the same or different, within, and across species categories. Seven images for each category were used, but two images were only pitted against each other once during the whole of testing. Two symbols represented the choices "same" and "different" respectively.

The subjects did not seem to respond on the basis of surface features, such as size, but since they performed on average in accordance with the experimenter defined species categories on "only" 69% of the trials, they are unlikely to have responded on the basis of species membership on all occasions. The implications mentioned above probably accounts for this.[70] But still, they must have made *some* assessment if we are to believe that they fully understood the concepts "same" and "different." If this assessment was always based on the animal content of the pictures, and never on surface features, is impossible to say. Even when animal categories were appreciated, we cannot be sure of which aspects of the animals that were used for the same/different judgements. After all, e.g. chimpanzees were judged to be "the same" to fish about 30% of the time. Even if content was fully recognised in the pictures we can unfortunately not conclude that anything but reality mode was applied to them. Colour photographs on a computer screen were used, and there was neither a requirement for differentiation, nor reference, inherent in the task.

7.5 Matching impossible content

A special case of MTS, which in theory might contradict reality enough for reality mode to break down, is matching different views of the same individual in a simultaneous matching task. In simultaneous MTS the sample image remains visible the whole time, while in sequential matching the sample disappears before the comparison stimuli is presented. If working in reality mode, matching photographs of the same individual in simultaneous matching is a strange occurrence. One version is to match identical photographs, but this poses less of a problem than different photographs since one does not really have to process the content of the picture. When matching different views, on the other hand, one has to identify all the stimulus contents and make the judgement whether they are the same individual or not.[71]

---

[70] A notable, and perhaps expected, discrimination that dragged down performance was that between tigers and house cats.

[71] Here it is assumed, however, that chimpanzees have different criteria for ascribing individuality to chimpanzees than for other objects. If the stimulus set had instead been apples, matching one apple to another view of the same apple could be based on similarity rather than identity, i.e. two differ-

However, it is a question for future research whether reality mode at all can break down due to “impossible” content. A first step in testing this could be to make a comparison between simultaneous matching of different views of known and unknown individuals, as well as objects. If this is specifically a test of the dynamics of picture processing in reality mode, one should also make sure to use pictorially naïve subjects since the whole matter is expected to work quite differently if someone happened to process in a pictorial mode. Then there would naturally be nothing strange with things e.g. being in multiple places at once.

The closest we get to such a test, albeit indirect, is Parr et al. (2000), who used black-and-white photographs of the faces of unknown conspecifics in a simultaneous MTS task given to chimpanzees, and a sequential version given to rhesus macaques. The pictures were displayed on a computer screen, encased in Plexiglas, outside of the cages, i.e. prime factors for retaining the illusion that one is viewing some sort of real scenes. Subjects indicated their choices by way of a joystick. Importantly, the subjects had experience in MTS, but neither of matching social stimuli nor any other “complex digitized stimuli” prior to the study. For the chimpanzees no training was needed to match identical photographs.

Two chimpanzee subjects performed above chance on their first trials, and the remaining three subjects on the second exposure to the stimuli. This unequivocal change in performance on the second trial should be considered with some worry. Since only 25 stimulus sets (sample, match, and non-match) were used, and food reinforcement was given on each correct trial, one-trial learning and choice by exclusion can unfortunately not be ruled out for the subjects that required a second chance. In discrimination tasks, one-trial learning as well as choice by exclusion is not uncommon in experienced learners (Harlow, 1951). When stimulus sets are kept constant (i.e. matches and non-matches do not switch roles), as in Parr et al. (2000), the second exposure can be solved as a discrimination task rather than as one of matching. Retaining at least 20 unique discriminations in memory is no feat for some chimpanzees (Hayes et al., 1953b). Nevertheless, two subjects did indeed perform above chance on the first trial. Matching of at least *something* in the pictures must therefore have taken place.

However, the critique above is not unimportant since none of the chimpanzees performed above chance on the first trial when the matching photograph was a different photograph from the sample photograph, i.e. a different view of the same individual. On the second exposure only two of the five subjects performed above chance. (One of whom had performed well on the first trial also in the identical-match condition.) Comparing the two conditions it is clear that matching different photographs of the same individual was more difficult than matching identical photographs. This implies that matching in terms of content, as in identity, was not as intuitive as matching based on complete visual correspondence. This can be due to the fact that recognising different views of strangers is difficult, or to the fact that the different photographs were not seen as different views at all, but different individuals. In the latter case the intended basis for similarity between the samples and the comparison stimuli suddenly becomes opaque. A comparison between matching

---

ent apples. Even though apples are probably more important to chimpanzees than to the average adult human, their identities are hardly as important as that of other chimpanzees.

photographs of strangers and matching of known individuals is needed to arbitrate between these interpretations.

In the final phase of Parr et al. (2000) sequential matching was used for both chimpanzees and macaques. In order to see how it affected the performance on the photographs used in the previous two experiments, different parts of the photographic face stimuli were masked For chimpanzees, only covering the eyes had a detrimental effect.[72] For rhesus macaques masking the eyes had to be coupled with masking the mouth to lower their success rate. Neither the chimpanzees nor the macaques were completely unable, as groups, to match in any of the masking conditions. This suggests that the subjects approached the pictures as global configurations where missing pieces were counterbalanced by those that were present. This conclusion is supported by the much easier task in Parr et al. (2006) where the requirement was instead to match a masked sample to its identical, but unmasked, counterpart. In this setup, masking the eyes had no detrimental effect on recognition. Gross pixelation on the other hand, as opposed to a mild one, did significantly impair recognition, as did manipulation of second-order relational properties, such as the spacing between facial features. There are factors involved in recognition of individual faces using a global processing strategy.

How did the macaques perform in the first two experiments in the Parr et al. (2000) study? They performed above chance on the fourth and sixth presentation of the stimulus sets respectively. However, since they were given sequential rather than simultaneous presentations direct comparison is problematic.

Using the same procedure as above Parr and de Waal (1999) compared different types of matching of black-and-white face photographs of conspecifics in chimpanzees. The task was to match two views of the same individual, mother – offspring pairs, or unrelated individuals. The mother – offspring pairs were further analysed according to sex of offspring. Naturally, matching unrelated individuals occurred at chance level, and so did matching mothers with daughters. However, matching mothers and sons occurred significantly above chance. This was unequivocally true for all five subjects. Best of all was performance on matching different views of the same stimulus chimpanzee.

Only second trial data and total performance for 600 – 650 trials are given in the report. Long-term learning effects could only be found for mother – daughter pairs. Thus the likeness of the daughters to their mothers seems to have had some impact after all. Since no learning effect was found for control trials it is unlikely that the subjects learned each response as a discrimination rather than as a match. It is likely that the subjects in Parr et al. (2000) likewise used matching strategies rather than relying on memorisation of the correct responses.[73] With regards to the condition with two views of the same individual the experiment can unfortunately not answer whether they were seen as the same individual or two different ones, but in any case the two views were responded to as significantly more alike than mothers and sons. It is likely that reality mode,[74] and not surface mode, accounts for the findings. Had

---

[72] For some reason, covering the eyes simultaneously as the mouth had less impact.

[73] It even seems to have been the same subjects in the two studies.

[74] Or pictorial mode, but this test cannot make the distinction.

simple feature matching been used it is a strange occurrence indeed that all five chimpanzees settled on features that were only shared between mothers and sons and not mothers and daughters.

To investigate why chimpanzee sons are perceived as more alike their mothers than are daughters Vokey et al. (2004) replicated the above study with human subjects. Using the same stimuli it was found that human subjects also more easily matched sons than daughters to their mothers. In fact, results for all conditions closely matched the results of the chimpanzees. However, in addition to the replication an analysis of the chimpanzee portraits were made which revealed that the distribution of characteristics in the pictures was biased between the sexes. This was due to how the faces were framed. The original pose, expression, and face-type of the stimulus animals probably accounted for this, and in interaction with the borders of the photographs an evident bias was created that was external to the appearance of the stimulus chimpanzees' faces as such. It seemed that mothers and sons just happened to appear in similar ways in photographs more than did daughters. Re-cropping the photographs close to the facial outline eliminated this differential effect between sons and daughters. Daughters became as easily matched to their mothers as were sons. That the ability to recognise similarity was retained is an important point because it shows that face properties per se, and not the framing biases, accounted for the positive performance. Rather than enhancing likeness, the framing had reduced likeness in the daughters relative to the sons.

That the interaction between the content of a picture and the boundaries of the picture itself can affect recognition in the negative is an important lesson for all studies that use picture stimuli.

The MTS paradigm is extremely open ended. Once a group of subjects are proficient matchers there is almost no limit to the kinds of tasks they can be subjected to, given that they understand the picture stimuli involved in a way proper to the task. In many studies it does not matter whether they view photographs as small semi-real events or representations of events far removed in space, and possibly in time. It does not matter because the questions that are studied pertain to perception and categorisation of the real world. In fact it might even be preferred that the subjects do not view the stimuli as much differentiated from the real world.

Lisa Parr (2004), for example, could study categorisation of emotional video clips in three chimpanzees in her Yerkes laboratory. The videos, depicting an unknown conspecific that displayed an emotional expression, with or without sound, were to be matched to static black-and-white photographs depicting facial expressions from the same category. The chimpanzees could also be played a vocalisation in isolation to be matched to a photograph. Or the sample could be a visual expression that was coupled with the vocalisation of a different emotional expression. This last condition was used to see which modality that had the more weight for discrimination in respective emotional category. One comparison stimulus that matched the visual and one that matched the auditory information was given, which meant that the subjects were non-differentially reinforced, i.e. there was no right or wrong response. The results showed that the three chimpanzees could match visual or auditory emotional information to static photographs. Again trial-one data for each of the 24 unique stimulus sets is not given. Since there were no learning trials and subsequent transfer trials, the first exposures are confounded with the subjects having to learn the specific matching rule. Thus trial-one data is not informative.

When visual and auditory information were mixed in the sample videos, the subjects utilised different information depending on the emotional expressions involved. Auditory information was preferentially used for choosing pictures of pant-hoots and play faces (laughter), while screams (fearful faces) were discriminated using visual information. However, there is great variation across subjects for how these preferences play out. I suggest that some of this variation might be attributable to an occasional problem of translating isolated video or auditory information into static pictures. When both visual and auditory information is available a clearer picture of the event to match is attained. Thus, for the visual stimuli, the problem is the interpretation of the sample video clips, and when matching vocalisations to photographs the problem is reading sound into picture. However, perhaps the most parsimonious explanation is that multi-modal samples just leave less room for lapses in attention. That would explain the heightened difference between the three subjects when the sample was unimodal.[75] Whatever the case, in terms of pictorial considerations, the more "real" the sample is, i.e. multimodal movies as opposed to single information channels, the more homogenous the responses seem to be.

Recognition of emotional expressions in photographs, coupled with MTS, can be utilised to query subjects about their attitude towards certain stimuli. Parr (2001), in a way, did exactly this. She let chimpanzees in her laboratory, at the time experienced matchers but naïve to using emotional stimuli in MTS tasks, categorise movie scenes of syringes, chimpanzees being injected with needles, and chimpanzees showing agonistic responses towards veterinarians. As matches Parr used photographs of fear expressions. Neutral and vocalising faces were used as non-matching comparisons. She called the procedure "matching-to-meaning." She also tested the categorisation of positive video clips, i.e. of the testing apparatus (!) and favourite foods, which should be matched to joyous expressions versus identical non-matches as the ones above.

There were learning effects, but the three subjects performed significantly above chance on the first session with all discriminations, which totalled 28. A session comprised two exposures to the stimuli and first trial data is not given. However, after seven presentations of the 28 discriminations, performance reached a criterion of 85% correct for two consecutive sessions, while a control condition with arbitrary matches remained at chance level. Thus, learning effects alone cannot account for the results.

Given that the subjects were naïve to emotional stimuli in photographs it is unlikely that a pictorial competence as such was formed during the experiment or was crucial for performance. Rather, the performance depended on recognising the content of the movie clips and photographs at face value, as real scenes. In fact, if one is not viewing them as cases of real scenes, judging the emotional value of them would be a very different feat. Understanding the task as "choose the pictures that

---

[75] They performed better when both video and audio were present in the sample, even when incongruous, than with unimodal samples.

represent what the movies represent" is very different from "choose the pictures that show what the movies show." Nevertheless, the fact that the pictures *did not* show what the movies showed in Parr (2001) seemingly places this experiment on the border to reference. Parr (2001, p. 227) herself notes: "But because the subjects were not physically participating in the emotional situations [...] the selection of specific facial expressions may be considered representational, in that they were used as markers of emotional valence." Since movies and photographs did not show the same thing the commonalities between them had to be inferred. However, the *pictorial* part in this reference is not necessarily different from finding a commonality between real events and real emotional expressions. We can thus have reference, to emotional valence in a movie, without differentiation between individual pictures or movies and that which they depict. A photograph viewed in reality mode does not *stand for* laughter more than a laugh does.

Mediation through one's own emotional reactions to the video stimuli can greatly help in finding the crucial commonality to base one's matching upon. In the same study Parr (2001) measured peripheral skin temperatures of the subjects. These corroborated the finding that the subjects indeed reacted emotionally to the stimulus movies, but only to movies of other chimpanzees being injected with needles, and syringes on their own. Conspecifics engaged in general agonism did not evoke a significant response as measured by skin temperature.

For social stimuli (colour photographs back-projected on a screen two meters from the subject) Boysen and Berntson (1986) could measure decreased heart rate in a juvenile chimpanzee when viewing favourite caregivers, and in Boysen and Berntson (1989) acceleratory heart rates when viewing an aggressive known conspecific. Response to other familiar individuals was minimal, whereas the heart rate for an unknown chimpanzee unexpectedly decreased.

Finding physiological emotional correlates when viewing pictures in a reality mode is expected. When viewed in a pictorial mode, on the other hand, more pictures can be expected to remain at a safe differentiated distance. Some pictures, though, can bridge differentiation and reality responses will kick in. Examples would be to feel distressed when watching a distressing picture, or aroused by pornographic pictures. Leaving aside a potential complementary part played by imagination, a photograph can be expected to evoke these reactions more easily than a pencil drawing. Also, getting scared by a pictorial tiger is most certainly less common than being scared by a pictorial snake or spider. The threshold for physiological responses can thus be expected to vary across what is depicted and how it is depicted. Individual variation can likewise be expected to be large, but individuals that exclusively view photographs from a reality perspective most probably place themselves at one extreme of this distribution. Unfortunately it is impossible to say where the subjects in Parr (2001) and Boysen and Berntson (1986; 1989) fall on this scale without further research.

7.6 Pseudo-pictures

While the studies of Parr and colleagues have focused on pictures of unknown individuals, Boysen and Berntson found strong evidence for recognition of known individuals. From a pictorial perspective, in terms of reference and differentiation, this is interesting since known individuals almost never are in two places at once. Also, they are seldom reduced in size, transformed in colour and more or less flat. Is reality mode this flexible?

Kojima et al. (2003) showed that a chimpanzee, Pan in the Kyoto laboratory, was able to match vocalisations of well-known individuals to their photographs. This means that she indeed recognised individuals in photographs as someone she knew. Only one photograph was used to designate each comparison individual but performance was above chance from trial-one. A notable exception was that Pan could not match her own vocalisation to her own photograph either than by exclusion. If this was due to an inability to recognise her own vocalisation, her own photograph, or both, is an interesting but unsettled question. Otherwise her audio-visual matching was by no means confined to chimpanzee stimuli but was applicable to photographs of known humans and inanimate sounding objects as well (Hashiya & Kojima, 2001a; 2001b; 1997). Birdsong was more difficult though (Hashiya & Kojima, 2001a). Again, whether the sound, the photographs, or both, is to blame for poor performance is not possible to say. A further implication was that Pan habituated to the photographs and performed best at novel presentations (Hashiya & Kojima, 1997). This can probably be attributed not only to the reinforcement history of individual photographs, but also to the deterioration of attention to pictorial displays that food rewards tend to cause (see Gardner & Gardner, 1998).

Bauer and Philip (1983, in Bovet & Vauclair, 2000) likewise showed that chimpanzees that could match photographs of familiar conspecifics could transfer this ability to match vocalisations to the vocalisers' photographs. However, this study has been criticised for not using novel stimuli in critical trials (Boysen, 1994). Such controls were included in Boysen (1994) and resulted in excellent transfer for one of four chimpanzees tested on auditory to visual matching of human caretakers. With photographs depicting conspecifics three subjects transferred well to novel auditory and visual stimuli. However, for this set of novel stimuli strict trial-one data is not given.

In Izumi and Kojima (2004) Pan could also match vocalisations to movie clips of known individuals. She based her response on the identity of the vocaliser even when the non-matching clip more closely matched the particular vocalisation type than did the matching individual (who had a passive expression). She could also judge which of two movie clips of the same individual most matched the vocalisation she had heard. Thus, besides recognising who the vocaliser was, Pan could both identify individuals and their emotional expression in silent movie clips.

For the non-social domain, Pan has also shown that she can perceive a correspondence between real objects and their photographs when it comes to learned *relationships* between those objects (Tanaka, 1996). After learning which objects form a group, Pan could select photographs from the said group when shown one of the objects, even though the pair-wise matching of those particular objects had never been specifically reinforced. That is, after learning that A – B form a group, and that B – C form a group, she deduced that a *photograph* of C forms a group with the *object* A.[76] The learned group membership of real objects had thus transferred to photographs of those objects.

So why not grant the chimpanzees in the Yerkes and Kyoto laboratories a pictorial mode competence? Matching known individuals poses some implications for picture processing argued to occur in reality mode. Reality mode must allow for reducing someone in size, perhaps transform colour, remove the third dimension, and so forth.[77] It is perhaps these types of "problematic" relations between the real world and the pictorial world that highlights sources for differentiation and helps the transition into a more pictorial mode. But differentiation is only one piece of the puzzle. When matching individuals to their vocalisations there is no referential requirement involved. Matching infants to mothers, or body parts to familiar group members, as for Java macaques in Dasser (1987; 1988, in Bovet & Vauclair, 2000), are also examples of tasks that can be solved without really having to move beyond reality mode, although it must be explained how the identities of children and mothers and parts of bodies of someone you know can be retained in size-reduced, flat versions. Especially when all body parts were attached to your friend last time you saw him or her.

But one does not really have to map the pictorial world onto the real world if one's visual and auditory systems react and recognise the input directly, regardless of the relationship to the non-pictorial world. The only requirement is that one does not become confused by the potential oddity of the situation, and retains the identification of the individuals in the pictures and sounds. One case of "magical thinking" is thus the acceptance of the way things present themselves, even when they are incongruent with one's experiences. If this requirement is fulfilled, incorporation of new experiences into one's expectations on the world will occur spontaneously. The fact that a known individual seem to appear in two places at once will not become a problem. But, as mentioned, this is not synonymous with a fully pictorial competence that allows differentiation and reference in the same act.

There is a possibility that "pseudo-pictorial" concepts can manifest themselves in tasks where confusion has worn off while the identity of the content is retained. But reality mode processing is still the source of recognition. A prediction would therefore be that for processing in pseudo-pictorial mode the possible generalisation to abstracted pictures is more limited than it would be in a fully pictorial mode. Rather than through the construction and application of a general picture concept, "pseudo-pictures" are incorporated as further exemplars of real-life categories. Reference, the sign function, is embedded in the e.g. matching task rather than attached to a general picture concept. Such a task is probably even pivotal for retaining recognition after initial "confusion" wears off, if the picture is not to become decontex-

---

[76] Such transitivity is one defining feature of equivalence relations (e.g. Sidman & Tailby, 1982).

[77] The flatness of pictures is perhaps less prominent when working on a computer screen than with physical pictures.

tualised and meaningless.[78] Another prediction would therefore be that pseudo-pictures might only work in certain tasks, and only work for certain categories. A subject who can match individuals to their kin in a computerised MTS task might not necessarily be able to communicate with photographs, or the subject would perhaps continue to bite pictures of preferred food (without pretending).

Pan, the Kyoto chimpanzee who successfully identified individuals in photographs, had difficulties imitating the pointing out of facial features on a colour photograph (Kojima, 2003). This was attributed to a limited body image in chimpanzees, but implications from using pictorial stimuli cannot be ruled out. In a sequential MTS format task the sample constituted a human face with a human cut-out hand touching the face in a specific spot with the index finger. Then the hand disappeared and Pan, at the time 7 years old, had to touch the same spot in the very same photograph. In the first experiment two spots marked the matching and the non-matching areas respectively. It took Pan 30 sessions to master the task. The number of trials in each session is not specified but in the other experiments in the study they are more than 50. Pan thus perhaps needed over a 1 500 trials to grasp the task. Most failures were due to pointing to the non-matching spot. When the pre-printed spots were removed in the second experiment Pan needed 18 sessions to reach criterion. The same facial photograph was still used as sample and the correct area indicated with a photographic hand. Especially poor was the pointing to areas of the face that had no clear landmarks, such as the forehead and cheeks. When the comparison photograph was displayed on another part of the screen than the sample photograph Pan's performance dropped and she again needed 30 sessions to perform well. Likewise when the sample or comparison photographs were rotated her performance dropped. The exception was the nose, which remained cantered regardless of rotation. If the sample and the comparison were rotated in the same manner performance did not drop.

The above results suggest that Pan did not parse the pictorial face into subcategories that could be retained in working memory. In a further experiment her response transferred to comparison photographs of two other humans than the sample, but when the comparison was a chimpanzee performance was severely impaired, except for when the correct area was an eye. The role played by the pointing finger in the sample is also questionable. Had Pan understood the application of a hand to a face in a photograph as a bona fide point, the task ought to have been easier.

In order to see what factors contributed to Pan's rather poor performance a simultaneous matching task as well as a task using real stimuli would be needed. Also, making sure that Pan understands the pointing gesture and can imitate its use is essential. Kojima (2003) combined these three controls into one and tested if Pan could indicate on the computer-screen face where the experimenter pointed at his own face. "Do-as-I-do imitation" of a model is difficult to learn as such, but possible, at least when the response is made on one's own body (Hayes & Hayes, 1952; Custance et al., 1995; Miles et al., 1996). Responding on a pictorial representation requires that it is seen as such a representation, or at least as a comparable object.

---

[78] Within a pictorial mode, on the other hand, context is always present in virtue of a picture being a picture and the expectations that this entails.

And then it is still subjected to the limitations found in live imitation. Kojima (2003) notes that when he taught chimpanzees to imitate pointing to their own body in earlier research, they tended to touch previously correct areas adjacent to novel target areas. Pan needed 19 sessions, which included correction trials, to reach criterion in the live model to photograph condition. Applying a *stand-for* relation to the pictures would arguably decrease the vast amount of learning in these tasks.

In the same way that a "symbol" can be used in a pseudo-symbolic way in (failed) ape language research, a picture can be used in a pseudo-pictorial way. Just as we expect a language competent ape to understand that also novel expressions are words, although the meaning is not yet clear, the picturehood of pictures is expected to precede their content.

Understanding a picture *as a picture* entails perceiving a commonality between all, or many, pictures. Experiencing the appearance of many pictures is thus crucial for forming a picture concept that spans across all those media that can contain pictures. This does not mean, however, that a chimpanzee that has only worked with pictures on a computer screen cannot acquire a picture concept; it just means that it might be limited to computer screens. When this ape has to transfer her ability to new media, it entails forming the picture concept again by discovering that marks on this new surface can do exactly what marks on a computer screen could do.[79] While the dynamics between expression and content might have to be learned again, due to media-specific transformations, the referential part of the equation should arguably not have to be learned again. When we learn a new language we do not need to learn how words relate to the real world all over again.

The expectation on a true pictorial or symbolic ability is thus that it will transfer to novel contexts relatively effortlessly, whether drilled pseudo abilities have to be built from the bottom up all over again. The animal never really "got" the word or the picture concept.

7.7 Referential use of pictures

In a series of tests aimed at measuring memory for specific feeding events the adult male gorilla King responded to the researchers' queries using pre-printed colour drawings (further nature unspecified) and photographs (Schwartz et al., 2002; 2004; 2005). The drawings had been associated with food as part of King's enrichment and communication with his caretakers prior to the testing described in the studies of Schwartz el at. If shown a specific fruit, or verbally asked for it, he was able to chose the correct picture from his stack of five wooden cards and hand it back through the bars of his enclosure. Whether the likeness in the pictures had any part in his acquiring of these fruit names is not clear. Since they apparently were heavily drilled they could just as well work as learned associations, i.e. conventionally rather than iconically. In Schwartz et al. (2002) King also learned three novel cards, containing names written in bold letters, to represent three experimenters. He learned

---

[79] Given that obstacles due to prominence have been transgressed.

to associate these during several weeks of training to both the presence of the people intended and just their names. When tested after extensive drilling he reached a 90% criterion in 30 trials (Schwartz et al, 2005). Since he was able to learn arbitrary connections between his name-cards and people, there is no reason to ascribe him an ability to recognise a visual likeness between his drawn picture-cards and fruit. After showing that he mastered the name cards, testing his memory for events could start.

King was tested on his ability to remember what he had been given and by whom, for both short (7 minutes) and long (24 hours) retention intervals (Schwartz et al., 2002). In order to return a card following the question "what did you eat?" or "who gave you the food?" King needed no training.[80] King was allowed one correction trial and was rewarded with verbal praise and food for correct choices. He performed very well on both "what" and "who" questions for both short and long retention intervals.

It can be argued that returning a picture card was equivalent with answering a question, or at least making a comment, or indicating, what had happened on a previous occasion. There thus was clear reference in his use of the cards. However, the iconic role, as opposed to conventional, is not addressed by the procedure in this test. It could easily have been tested, though, whether he could make proper statements also with novel pictures. Nevertheless, King is still showing that the cognitive underpinnings for referring to things not present, by way of an external medium, are there also in non language-trained apes. This is a most important requirement for a pictorial competence that is not tied solely to the expression in front of one's eyes, as in surface or reality mode processing. Even when fed many types of foods and encountering different people during the course of the long retention intervals, not to mention during sleep, King was able to ignore these events and single out one specific occurrence of food and person in time (Schwartz et al., 2002).[81]

Schwartz et al. (2004) introduced novel photographs in a test on remembering events, people and objects. Events and objects were all unusual and could be such things as skipping rope, playing an instrument, bouncing a basketball etc. Familiar and unfamiliar demonstrators were used. For each episode King had three photographs to choose from, presented to him five minutes after the event. However, rather disappointingly for our concerns, the photographs only depicted portraits of people against a white background, or a single object against a white background. King thus did not need to read actions and events per se into the photographs, but only identify the matching individual or object. This is not trivial, though, since experience of a real object that have never been seen before can be perceptually far removed from its static depiction which also is novel.

King could correctly recognise novel objects in photographs (50%), as well as identify well known people (60%) and strangers (55%), significantly above chance level (33%). When trying to explain these significant but rather poor results it is impossible to separate lapse in memory from lapse in recognising people and objects in their photographs. Perhaps King just lacked motivation or adequate understand-

---

[80] Before the study King seemed to already respond adequately to a fair amount of spoken English.

[81] Gorillas are otherwise believed to be sensitive to interfering information during retention intervals, just as human eyewitnesses are (Schwartz et al., 2004).

ing of the task. Furthermore, he was more often correct for some people than others. This can have reflected his recognition of these people in photographs, but also other factors such as his attitude towards and willingness to cooperate with them (Schwartz et al., 2004). Findings like these are difficult to interpret, also from a pictorial perspective. It is impossible to say whether King acted as he did for communicative purposes (i.e. referentially), displayed an attitude towards the photographs as such (which might have been contaminated with properties of the real), or just did not want to cooperate. In studies where interaction is open-ended it seems that apes sometimes make up their own rules. A further habit of King's was for example that he occasionally placed the correct choice card under his leg and returned an incorrect one. He never withheld an incorrect card in this manner. Thus in 24% of his incorrect choices King could in fact have known the right answer. What is clear, though, is that King could use also novel photographs in a seemingly referential task.

Schwartz et al. (2005) extended the findings on King with King showing that he remembers the order of events and where they occurred. On a pictorial note only photographs of places were novel. For the ordering of events (feeding events) the drilled fruit drawings were reused. Photographs of three places had been taken from King's view inside his cage. Whether King used these photographs correctly from trial one is not reported but it is unlikely that he would be able to bypass similarity and learn the designated locations through associative learning. There are just too many things that could be mislearned in associating an external event with a specific card.

That King, given that the photographs were not learned associatively, could relate a photograph of a location to an event in memory, thus strongly suggest a referential and differentiated view of the pictures. The pictured is used to make a statement about the referent, not the other way around, and it is an apparent difference between the two. Very magical thinking would indeed have to be evoked to explain how the location can be in two places at once, and furthermore be heavily reduced in size. That said, there is no reason to doubt the magic of photographs when they apparently work. The low level of correct response (45%; chance 33%) might possibly reflect these representational problems. Where an event has taken place is otherwise bound to have a salient place in a foraging species memory system (see e.g. Gibeault & MacDonald, 2000; Menzel, 1973). In fact, King's performance significantly declined during the course of the 60 trials. During the first 20 trials he had been 60% correct. Rather than blaming motivational factors it is possible that the photographs started to lose their meaning with a few incorrect responses. This would have been facilitated by their unsure connection to the events to be "commented" upon.

The perhaps most well-known ape example of using photographs to "comment" on an event is a test by Premack and Woodruff (1978a) with the chimpanzee Sarah.[82] Inspired by the debate regarding insight in Wolfgang Köhler's chimpanzees, Premack and Woodruff (1978a) wanted to test Sarah's *knowledge* about problem-

---

[82] Sarah is a chimpanzee trained in symbolic representation using tokens (i.e. magnetic plastic chips) and has been involved in cognitive testing, notably matching tasks, since infancy (Premack , 1976; Premack & Premack, 1983). When tested she was experienced with photographs and television.

solving situations rather than her own *performance* in such tasks. This shift was important in order to exclude trial-and-error contribution to insight solutions (see e.g. Chance, 1959). The problem-solving situations which Sarah was to analyse were enacted by people. This test is famous because it would also lay the foundation for an explosive interest in "theory of mind" research in apes and humans, starting with Premack and Woodruff (1978b). Interestingly, but largely ignored, the test is not only about reading intentions into agents, but also about reading intentions into pictorial media.

Sarah was shown one of four video clips depicting a well known human that encountered a problem. The movie was then paused and Sarah was presented with two photographs: one that depicted the person solving the problem, and one depicting the person involved in an activity that was incongruent with the actions just seen in the movie. On trial one she was correct on three of the four problems. Overall performance was correct choice on 21 of 24 trials using the total of four movies and four photographs.

Four further problems were presented as movies and four novel photographs were developed for the choice task. This time, however, the photographs were of single objects which never appeared in the movies, but would help in the solution of the problems displayed. Sarah was correct on 12 out of 12 trials, which corresponds to correct choice on all four problems at the first trial. Since depicting single objects, there had been no need to read activity into the photographs in these problems.

Although Premack and Woodruff (1978a) made sure to include the same objects in all four of the first movies, so as to exclude a response made on visual matching between the frozen video image and the photographs, Savage-Rumbaugh and Rumbaugh (1979) evoke exactly this as a possible explanation for Sarah's performance. Furthermore, there was no independent evidence presented that showed that Sarah actually viewed the movies as problems to be solved, and consequently that the photographs depicted solutions to these problems.

If ordinary MTS competence was applied to the task, the dynamic content of the photographs, i.e. depicted action, could have been bypassed altogether. In fact, 3.5 year old human children who were shown Sarah's stimuli did exactly this. They matched salient objects in the movies to photographs that contained these objects rather than matching "solutions" to "problems." Sarah had not performed as these children, however (Premack & Premack, 1983). For the second set of four problems Savage-Rumbaugh and Rumbaugh (1979) suggest that MTS, but of an associative type, could account for the performance. Given the possibility for trial-one learning the low number of problems, only 8, is also a concern.

Premack and Woodruff (1979) maintained that Sarah did not have a direct experience of the associations that would be required in order to solve the tasks on matching grounds, and that further tests (i.e. Premack & Woodruff, 1978b) confirms that Sarah indeed understood the videotapes and did not merely match photographs to the frozen video image. Furthermore, on the problems where Sarah failed she could easily have applied a matching strategy and succeed, but still she continued to fail after several attempts.[83]

---

[83] There is reason to believe that this failure was for conceptual reasons since the movie showed one of those problems that also Köhler's chimpanzees had great difficulties solving.

In Premack and Woodruff (1978b) the test was given a twist. Sarah received three choice images after viewing new recordings of the former problems. Each recording came in two versions: One containing an actor that Sarah liked, and one with a person that she did not like. The three photographs were this time of the correct solution, of a devastating mishap, or of an irrelevant action. Sarah chose correct solutions for the actor she liked, mishaps for the one she did not like, and rarely chose the irrelevant photograph. However, rather than choosing the mishap that pertained to the video, Sarah sometimes preferred to chose an unrelated punishment for the person she did not like. The positive photographs were all reused from the earlier experiment.

Another question is whether Sarah saw a correspondence between the scenes in the film and the scenes in the photographs. In a condition that was intended to test attribution of mind states to the actor, Premack and colleagues wanted to see if Sarah considered the attentional direction of the actor in the film clip during the time one of four opaque boxes was baited. Which box would he subsequently choose? Sarah was given a choice between photographs of the correct box and of an unbaited one. She did not base her response on the attentional state of the actor. However, she did let the favoured actor chose the correct box and the disliked actor the wrong box. Again she apparently chose according to what she *wanted* to happen (Premack & Premack, 1983). I have not seen the choice photographs, but if no food and no actors are visible in them, she showed that she saw the boxes in the pictures as corresponding to the boxes in the film clip, a clear referential use of the pictures. However, if the food rewards were visible in the opened boxes this claim cannot be made.

Among the experiments reviewed above it seems clear that the use of pictures in referential tasks, rather than ones that depend on visual or categorical matching, are the strongest cases for a pictorial mode processing of pictures and their relation to the rest of the world. However, this relation might not be what one would expect in a pictorial mode building on truly pictorial concepts. The notion of pseudo-pictures was used to explain cases where pictures in a reality mode were used in matching tasks despite the fact that they displayed "impossible" content. Pseudo-pictures can also be used to refer to external entities and events. Some apes discover that they can use pictures to perform certain tasks that are required of them, such as answering "what" and "who" questions. But the relation between the picture and the external world that they discover is not that pictures themselves are about the external world, but that they can be used to comment on the external (or inner) world. Thus the picture – referent relation is one of *usefulness*, going from picture to the world, and not one of *depiction*, going from the world to the picture. They discover that they can use what they recognise in the pictures. Remember from section 5.1 how the chimpanzee Viki used pictures of beverages and cars to request drinks and car rides. She was highly limited to these items however, and attempts to get her to communicate with a wider assortment of pictures failed. If a photograph is viewed in reality mode, and hence is an object of sorts, problems pertaining to prominence can be expected. In order for one object to stand for another object, all of one's attention cannot be caught up in the first one. If I get heavily involved with the expression in front of me, to the point that I might even test its realness by touching it, express my liking by kissing it, or my distaste by hitting it, my attention to a potential function of the same picture is greatly hindered.

It would be highly interesting to see how e.g. Kings seemingly referential use of pictures would generalise over contexts, and furthermore, interact with degrees and kinds of iconicity.

Chapter 8

Enculturants and anecdotes

Apes that are raised in close proximity with humans and their ways of life tend to perform differently in experimental tasks than do individual raised exclusively with conspecifics. The process that makes apes in human settings "atypical" is commonly referred to as "enculturation." However, there is no convincing theory about what enculturation really is (Tomasello & Call, 2004). Somehow the interface between ape and human is enhanced, of which communication is a central part, but also other social aspects, like the ability (Call & Tomasello, 1996) or motivation (de Waal, 1998) to imitate. Enculturation is thus some kind of socialisation process. But does this process yield a cognitive change or a change in interaction patterns? On what level is the interface between ape and human refined?

In an older theory of theirs, Call and Tomasello (1996) opted for a more cognitive change taking place in the enculturant, one of starting to view others (humans) as intentional agents. However, according to Bering (2004) enhancement is restricted to interaction with objects. Tomasello and Call (2004) extends this to communicative behaviours: Enculturated apes understand and use human signals more than do non-enculturants. In light of new evidence regarding intentional agency they also update their older theory and posit that enculturation does not create new skills, but is rather a matter of modifying "existing social interactional and attentional skills" (p. 214). They leave it at that.

When comparing infants reared in a human responsive environment, as opposed to standard laboratory care in peer groups, at the Yerkes facilities, Bard and Gardner (1996) found differences in personality rather than in cognition. Infants that had received the more interactive care displayed among other things better persistence in attaining goals, longer attention spans, and higher levels of manual contacts with objects. These early differences can set up continued developmental trajectories differently between individuals. Variation in personality and cognition later in life should not surprise. However, differences should also not be taken for granted, since convergence on similar outcomes through different routes is also common in development (e.g. Spencer et al., 2006).

Enculturation is not simple mathematics, adding one ability to another, but is the effect of sharing one's life for a prolonged period of time. "Enculturation occurs in labyrinths of life, not in referential triangles," in the words of Segerdahl et al. (2005, p. 132). What takes place is the learning and anticipation of another beings responses in a shared and reoccurring social, physical and, perhaps, mental environment. The outcomes are consequently expected to be different in different environments, which have been pointed out by Call and Tomasello (1996). In a milieu where a human controls access to everything, as in a classic zoo or laboratory setting, getting what one wants out of the other, pushing the right buttons so to speak, is probably a large part of enculturation. It will yield pointing behaviours, attention-getting behaviours etc, which might or might not boost the development of reading attentions and intentions. In a more free interaction, as in a home-raising situation, there is perhaps no limit to the synchronisation that can be achieved between human and ape worlds (and minds). This is what has happened in the most successful ape language projects. Future research will have to settle exactly what happened.

That enculturation exists at all suggests that great apes are culturally predisposed and that cognitive variation can be one of experience. I will leave further discussion about the sources of enculturation and only make use of a fact that Call and Tomasello (1996) thinks might contribute to the perceived difference between home-raised/cross-fostered apes and their wilder counterparts: that there just are more, and closer, observations of enculturated apes. In this chapter I will account for anecdotes and data from a few such projects. Anecdotal descriptions from careful observers are the next best thing to long-term personal experience when it comes to glimpse the full scope of animal potential.

Several zoo keepers (pers. com.) report that apes in their care are fond of looking in magazines, but they also have difficulties pinpointing exactly what it is in the magazines that elicit this interest. As much time can apparently be spent browsing through a furniture catalogue, as one that contains animal pictures. Sometimes the flipping of pages itself seems to be an enjoyment. However, there seems often to be a noticeable difference between magazines with and without pictures. To take a published example, Hoyt (1941) supplied a home-raised gorilla, Toto, with picture magazines. These she would browse and study by the hour, spending long times at preferred pictures. But given ordinary newspapers she would rather crumble them up, tear them to shreds, or stomp on them joyfully. The desired effect seemed to have been the noise it made.

8.1 Gua

Winthrop and Luella Kellogg began their cross-fostering experiment in 1931 (Kellogg & Kellogg, 1933/1967). It continued for nine months, during which time they made comparative tests and recorded developmental sequences for their own son, Donald, and a female chimpanzee from the Yerkes breeding colony, Gua, who they raised as a second child. At the onset of the study Donald was ten months old while Gua was seven and a half.

The rationale behind the endeavour was to explore the age old question of the respective roles of nature and nurture in human ontogeny, seemingly fuelled by the case of “feral children” who were unable to readjust to civilised society after being found in forests, caves, and wolf dens (Benjamin & Bruce, 1982). Kellogg believed that those children had started out with the same potential as other children, but that the fully human trajectory was forfeit due to early experiences in a different direction. He saw a chance of settling the subject by cross-fostering apes, a suggestion originally made by Witmer in 1909 (Kellogg & Kellogg, 1933/1967). If a child could not be put in the forest as a scientific experiment, the forest had to come into the home and live like a human. By way of analogy, if a chimpanzee can develop human traits in a human environment it is not far-fetched to assume that a child does the same for the same reasons.

If differences are found when two species are reared under exactly the same conditions (Gua was on no accounts to be treated differently from Donald), innateness *of some sort* would likely be the explanation. However, today we know that the interactions between hereditary components, environments, and the experiences of the organism are complex in the extreme (Moore, 2003). This is true for humans as well as chimpanzees. If Gua fails where Donald succeeds we cannot therefore automatically assume that this is because of the constraints set by Gua's inherited nature (or facilitated by Donald's). We can only say that Donald clicked and Gua did not under those particular circumstances.

Kellogg and Kellogg (1933/1967) were well aware of the fact that nature and nurture interact, and that it must, but still they thought that it made sense to ignore the most evident aspect of nature, that Gua and Donald had different bodies.[84] By having to adapt to a human culture through being treated as a human, Gua was forced into the bodily and social mold of a human baby. This substantially changes the meaning of the words: "[...] full opportunity to acquire a complete repertory of human reactions [...]" (Kellogg & Kellogg, 1933/1967, p. 15). The "opportunities" offered by being encouraged to walk on two legs or wear shoes are not transparent.

If the fact that Gua was a chimpanzee had also been taken into account and catered for, surely she would have had *more* developmental opportunities. As a result, the comparison between her and Donald might have turned out very different. For example, should the observation that Gua did not learn to speak during the study[85] be attributed to the fact that she was born with a chimpanzee anatomy, or to the fact that she was not given more suitable means adapted to that anatomy, such as a lexigram board? Ironically, Donald and Gua could have turned out more similar if they had not been pressed into the same mold. Since they shared environments and thus were faced with the same problems, and if they had been allowed to take individual routes to develop solutions for those problems, the likelihood that they would converge on the same abilities would be higher than if one of them were hindered from accessing his or her full potential.[86]

Kellogg and Kellogg (1933/1967) did show that early environmental opportunities as well as the biology brought to the table were crucial and interactive factors in ape (and by analogy human) development, and they did it in an experiment that every-

---

[84] They describe these differences in great detail however.

[85] Donald did also not learn to speak during the study, which might have contributed to its early end (Benjamin & Bruce, 1982). However, he readily imitated several of Gua's vocalisations.

[86] That human infants often take different developmental routes to the same behavioural end points has convincingly been shown in the works of e.g. Thelen and colleagues (see e.g. Spencer et al., 2006).

one had theorised about but no one turned into reality. However, the scale on which to actually compare Gua and Donald will remain a superficial one as long as one looks at surface behaviours and neglect detailed analysis of underlying interactive processes. It will have very little to do with the nature versus nurture issue which Kellogg set out to address. Nevertheless, the Gua experiment remains valuable as a purely descriptive study of human and ape development (in a particular household), and the potentiality of the chimpanzee brain (constrained in baby clothes).

If Kellogg and Kellogg could do the experiment all over again they would choose two younger subjects of a more equal age, raise them in a larger family with many siblings, continue the study for a longer time, and they would continue to treat the subjects as anthropomorphic as possible (Kellogg & Kellogg, 1933/1967). Almost 40 years after the study, when revisiting the question of cross-fostering experiments, W. Kellogg maintained that the ultimate purpose of cross-fostering was to "determine the genetic limitations of the animal when it is given the enriched environment of the civilized household" (Kellogg, 1968, p. 489). Furthermore, he still believed that apes would not learn language from mere exposure to speech, but he recognised the fact that a change in modality might help. Consequently he saw a promising future in the work just started by Allen Gardner and Beatrix Gardner (e.g. Gardner et al., 1989) with teaching sign-language to the chimpanzee Washoe (section 8.3).

Among the many observations described in detail in Kellogg and Kellogg (1933/1967) responses towards pictures, movies (see section 9.3) and mirrors (section 9.4) are of special value for this thesis.

Gua showed a spontaneous interest in pictures at the age of 10 months. She pointed to various parts of them and often tried to pick up depicted content with her lips. Especially coloured pictured received this treatment, and that which she touched with her lips were "often vaguely similar to drawings of fruits or other edible objects" (Kellogg & Kellogg, 1933/1967, p. 92).

A longitudinal comparison between Donald's and Gua's reactions towards the same ABC book, containing coloured drawings, was made when Donald was between 14.5 and 18.5 months and Gua between 11 and 16 months old (see *fig. 2,* p.19). It is clear from the description that Kellogg and Kellogg (1933/1967) believe that it is the vividness of the stimulus that elicits reality responses in the subjects. The more like a real object the picture seems, the more manipulative behaviours are directed towards it. Without the full data set we cannot know if this is what actually happened or just what Kellogg and Kellogg though they saw because it confirmed their theory. However, from the book at large it seems clear that the authors strived towards an objective assessment.

At the beginning of the period Donald focuses on turning the pages of the book, only once in a while stretching his hand towards the depicted shapes. This changes after 15 months of age and he spends more time looking at the pictures. He often rubs the pictures with the palm of his hand or his fingers. Sometimes he tries to pick a depiction up with a precision grip between thumb and index finger. This seems to be most common for brightly coloured shapes that stand out the most from the background. At 17 months he seems to turn the pages in order to view new pictures, not for turning's own sake. His right hand is constantly in contact with the pictures and his fingers follow his attention. Sometimes he scratches a picture with a fingernail "as if trying to remove a three-dimensional object" (Kellogg & Kellogg, 1933/1967, p. 93). At 18 months he spends even more time looking at pictures and he points at them, especially following a point of the parent to the same area. He still seems to react to the object properties of some pictures, but he does not scratch but instead feels in the air above them as if to checking for a raised surface.

Gua's development seems somewhat reversed from Donald's according to Kellogg and Kellogg (1933/1967). Her interest is in the pictures from the start. She "points" to many of them and she tries to pick them up with her lips, even some of the brightly coloured letters of the alphabet. She makes no attempt to turn the pages of the book on her own accord. The behaviours are similar at 12 months of age but her interest for the pictures is judged to be perhaps somewhat stronger. She looks fixedly on shapes pointed out to her by a parent and copies points towards pictures. Her lips are sometimes extended to only parts of a picture, especially if that part is a round, symmetrical, or brightly coloured shape. As an example, after mouthing a pair of round pink pigs she tries to scratch them from the page. When this fails she slaps the page. It is thus evident that she approached the shapes as potentially graspable objects. At 13 months Gua starts to turn the pages of the book herself. To her manipulative actions she adds rubbing pictures with her knuckles. The development of her page-turning ability continues at 15 months, but she still grasps at pictures with her lips. By 16 months she has supplemented mouthing with a full grasp of the hand, still targeted towards the drawn shapes.

At the end of the comparison Donald showed a larger interest in pictures than did Gua. We learn from this account that reality mode is not limited to photographs, but to any shape with an apparent pop-out nature. It is interesting that the experience with turning pages and finding out about the nature of paper do not seem to affect the grasping actions towards the pages, even after months. Unfortunately we are not informed about what other experience Donald and Gua receives with other pictorial material during the course of this longitudinal study. They are hardly isolated from it. For example, we learn that Gua on one occasion bites the picture of a cracker on a wrapping paper.

Kellogg and Kellogg (1933/1967) analyses these confusion behaviours within a larger framework of learning about two-dimensionality. The default mode, they maintain, is three-dimensionality, and only eventually does an infant learn that some instances of objects lack or have a very different third dimension. Gua, for example, had to learn that you cannot collect spilled milk with a grasping action. But there are other reasons than perceptual that can account for the fact that an infant chimpanzee can happen to grasp at puddles of milk, such as poorly developed motor scripts for liquids. Furthermore, they did not grasp at everything two-dimensional. The actions towards the pictures probably happened just because they seemed object like.

Did Gua at all recognise the content of the ABC book or did she just go for what looked most like graspable objects? In a verbal comprehension test as part of a larger test battery, Donald (17.5 months) and Gua (15 months) were shown a small white card with four "printed crude ink drawings": A cup, a dog, a house, and a shoe.

Donald could point out the dog on command, but no other objects. Gua in turn could indicate both the dog and the shoe. We are not given information of attempts to preclude cueing in this test, nor other details, but the conclusion of Kellogg and Kellogg (1933/1967) is that this event might illustrate Gua's "superior ability to perceive printed pictures" at this age. Needless to say there is a possibility that the performance in such a limited data set might be attributed purely to chance. Of the 58 instances on which Gua adequately responded to verbal commands during the whole nine months of observation, the word "shoe" can namely only be found in the example above. The list excludes all verbal communication that occurred in the household that was not directed to Gua, as well as those commands that were not adequately responded to by her, but from the data given one cannot conclude that Gua knew the word for shoe. "Bow-wow" (dog) has a bit more support.

8.2 Christine

Christine, the baby chimpanzee described by Hess (1954), enjoyed looking at pictures in magazines and books, and proved clear recognition from the age of 12 months. To explore this behaviour Hess used pictures that were likely to interest Christine. She showed her pictures of other apes and asked "Where is Christine?" The chimpanzee would point to the ape pictures several times in a row, then eventually point at something completely different when asked one time too many. Christine's first points were always at the correct picture though, Hess reports.[87] In a group picture of ten people Christine singled out the one person she knew, pointed it out and kissed it. This response was repeated towards the picture on several consecutive days. The most stable performance seems to have been with pictures of cats. Cats, in any form, were one of Christine's special interests. If several pictures were placed in front of her Christine would always pick out the cat one, make a noise of pleasure, kiss it, and "scratch on the picture as if to pick the kitten out" (Hess, 1954, p. 12). Photographs of kittens dressed up in clothes evoked the same response. The most valuable observation in Hess' report is that the chimpanzee also picked out drawings of cats.

Hess recognised that the best evidence for Christine's pictorial competence coincided with the evidence for her word competence. Pictures were used to test words, and words were used to test pictures. When given the command "Where is X?" Christine is said to have been able to reliably pick out horses, chickens, and cats from novel pictures (further nature unknown). As a check for habitual response the novel items were presented together with one known picture. She did not choose the well known picture, implying that reinforcement history did not interfere with her choice of pictures. In fact, when asked for pictures that could not at all be found among the presented ones she refused to pick anything and looked around for the missing ones, including turning over the other stimuli to see if the sought pictures were pasted on their backs. However, in all forms of setups, if a cat was present in the picture set she naturally chose her favourite motif, the cat. If encountering a cat

---

[87] She could also point out Teddy the anteater, and her friend Georgie the lamb, when asked to.

picture in a magazine she would typically tear the picture out and carry it around, kissing, hugging, and biting it. The pictures disintegrated within minutes, which is said to have frustrated her.

Hess (1954) studies are informal and do not include much detail nor any statistics, but raises interesting questions nonetheless. The kissing of pictures, for example, was it a comment directed to Hess, some form of thinking aloud, or involvement fuelled by a reality response? It is clear that the pictures took on a nature of their own, which could have been different from both fully real objects, and from pictorial objects. Perhaps the term pseudo-picture is useful again. Unfortunately Hess does not report differences in responses to drawings and photographs when it comes to e.g. kissing.

8.3 Washoe

The choice of teaching sign language to a chimpanzee by Gardner and Gardner (1969) was made due to the failures of Hayes and Hayes to teach vocal words to Viki. Viki acquired only four utterances that approximated the sound of English words. That said, Viki did use vocal sounds more adapted to chimpanzee production skills in meaningful communication with humans. Conventions such as clapping one's teeth together to request a car ride, or saying "ch" for a drink, developed spontaneously in the household (Hayes & Hayes, 1954). The language competent bonobo Kanzi also proved, years later, that vocal expressions used in communication with humans can contain specific semantic content and is not an impossible modality for language research (Taglialatela et al., 2003). Although Kanzi's ability might have been acquired as a result of upbringing (Hopkins & Savage-Rumbaugh, 1991), the possibility of learned referential vocalisation in apes outside of human culture is still understudied (see e.g. Slocombe & Zuberbühler, 2005), despite the fact that its feasibility was proposed a century ago (Garner, 1896): "So far we have only taken the first step as it were in the study of the speech of apes" (p. 74).

Because ape vocal communication gives the impression of being exclusively involuntary, and is not easily decoded by most human ears, the relative ease by which chimpanzees use gestures in a voluntary and transparent way, for example when begging, stands out. Gardner and Gardner (1969) were thus convinced that American Sign Language (ASL) was the right modality for language training. 40 years earlier Yerkes (1926, cited in the original English in Yerkes and Sykes Child, 1927) had made the prediction: "I am inclined to conclude from the various evidences that the great apes have plenty to talk about, but no gift for the use of sounds to represent individual, as contrasted with racial, feelings or ideas. Perhaps they can be taught to use their fingers, somewhat as does the deaf and dumb person, and thus helped to acquire a simple, nonvocal, 'sign language'" (p. 54).

In order to not only elicit requests for food items and favours, a rich material and social environment was made available to Washoe, in which two-way conversation would make sense. Unfortunately, however, spoken language was used minimally around her. This was not a decision based on ignorance, but in the best interest of Washoe as a communicator on equal terms (Gardner & Gardner, 1969). But it was perhaps an unfounded concern since comprehension of spoken language has proven to be an integral and possibly fruitful part of interaction in both Project Koko (e.g. Patterson & Linden, 1981) and the bonobo research of Savage-Rumbaugh and colleagues (e.g. Savage-Rumbaugh et al., 1993; Brakke & Savage-Rumbaugh, 1995). Since it makes no sense to exclude comprehension from a definition of language, understanding spoken language is as much the language of these apes as are their lexigrams or signed gestures. Williams et al. (1997) report that the type of language-use the apes had been involved in during their formative years greatly affected their comprehension of spoken human language. Where Kanzi, Panbanisha (*Pan paniscus*) and Panpanzee (*Pan troglodytes*) understood spoken language, Sherman and Austin fell short, although they often gave the impression of understanding since they were apt at using contextual cues. The two latter had received a trial-based training, while the former had been continuously included in everyday discourse about the shared life of humans and apes at the laboratory. Consequently, the language of Sherman and Austin can thus be argued to be very different from the experimental generation that followed. Likewise, such differences might also pertain to picture use, but since detailed information on the specific use of pictures in respective project has not been published, a comparison cannot be made here.

The reason for initially choosing chimpanzees for language studies was not primarily their genetic closeness to humans, but the intensity of their attachment to human caretakers. Sociability was seen by Gardner and Gardner (1969) as *the* crucial aspect of language learning. Later, however, Gardner and Gardner's methods would be successfully applied to both gorillas (e.g. Patterson & Linden, 1981) and orangutans (e.g. Miles, 1990; Shapiro & Galdikas, 1999).

Washoe was caught in the African wild and was less than a year old when the project started in 1966. Within two years she had learned about 30 signs, defined by strict criteria, confirming Gardner and Gardner's hunches about the appropriateness of a signed language (Gardner & Gardner, 1969). A year later the signs readily used by Washoe amounted to 85 (Gardner & Gardner, 1971), and still a year later to 132 (Gardner & Gardner, 1989). Imitation and molding of her hands were used in training and real objects as well as pictures of objects were included as conversation topics from the beginning. Some entities were trained more with pictures than real instances, such as animals. "Dog" and "cat," for example, are reported to mainly have been taught by using photographs. Miniature replicas were also used, but far less successfully than photographs (Gardner & Gardner, 1971). It was judged that this effect was due to lack of realism, although the miniatures had been selected just because of their striking realism, from a human perspective. (Miniature replicas will be further discussed in section 9.1)

Given her ready transfer between photographic instances and real versions of various sorts, it is clear that Washoe recognised at least photographs. This was confirmed beyond doubt when she was formally tested in a blind procedure on naming novel back-projected photographs at the age of 4 years (Gardner & Gardner, 1984). The testing with novel photographic pictures was indeed successful and Washoe performed correctly in the 70-90 % range. In a later project three other sign-language trained chimpanzees (Moja, Tatu and Dar) were tested in an equivalent paradigm when they were about 4 years old respectively. Only Moja performed relatively poorly, being able to name about half of the slides. Interestingly, Gardner and Gardner (1984) attribute Moja's performance to the poor quality of the photographs, which had been prepared by "an inexperienced photographer."

It was necessary for Gardner and Gardner (1984) to carefully chose good testing exemplars with regards to how the depicted objects were presented in the picture, in terms of perspective etc, and they "[…] had to learn to look at the slides with the eyes of our subjects" (p. 386). The implications of this fact did of course not concern Gardner and Gardner because their test pertained to testing vocabulary, but the need for such choices also speaks for a reality mode processing of pictures. Naturally, Washoe can only name objects that she can identify with certainty, and with real objects she can always change her view to turn an atypical view into a typical one. With pictures she is stuck with one view. However, this is only a problem if one tries to make sense of pictures as one makes sense of the real world. As argued in this thesis, truly pictorial competence is about interpretation. The ability to fill in information, recognise atypical views, renderings etc., is something that a pictorial mode implies. A judgement of likeness through the effects of e.g. resemanticisation is possible only because the system actively tries to make sense of a scene. My guess is that when the content did not hit Washoe directly, she did not try to interpret the pictorial view.

Also in free response photographs had to have the content blown up and the background kept as plain as possible, otherwise attention and naming would get caught in non-intended content. This suggests that certain conventions regarding the focus of a photographic display were not in place in Washoe and the rest of Gardner and Gardner's subjects. It can also be that Washoe turned to naming backgrounds when she failed to recognise the foreground, which is nevertheless an unconventional way of reading pictures. Although this interpretation does not exclude a pictorial view of pictures, it is suggestive of a more face-value mode, like reality mode.

But even though these chimpanzees were possibly working outside of a fully fledged picture concept, they learned the names of objects and animals from photographs and generalised this to the three-dimensional world. Arguably the name training they received helped with bridging the two types of instances, but it remains a speculation.

Looking at pictures, an activity Washoe enjoyed almost from the first day of her training (Gardner & Gardner, 1985), remained a favourite activity for the rest of her life. "[...] She spends her free time looking through books, magazines, and catalogues (especially shoe catalogues). She apparently likes to look at them by herself but doesn't mind signing about the pictures with friends" (Friends of Washoe, 2004).

8.4 Lucy

A lesser known replication of sorts of the Kellogg and the Hayes rearing experiments with Gua and Viki respectively, is Maurice and Jane Temerlin's upbringing of the chimpanzee Lucy (Temerlin, 1976). From the second day after birth, and for the following decade, Lucy was raised in the home of the Temerlin family together with her older human foster brother. Maurice Temerlin was a Freudian psychotherapist, which influenced the study of Lucy a great deal. This will be apparent below. To say that Lucy was raised exactly like a human child is not true. The fact that she was a chimpanzee changed many things in the household. A concrete example is that the house and its furnishings were modified to be chimpanzee proof. Since the age of 5 years Lucy was also part of Fouts' sign language research,[88] which is an experience quite different from learning to express one's language as a human child from a much younger age. Lucy acquired a vocabulary of just above 100 signs. Her human upbringing was thus far from typical, and she probably got away with a lot by virtue of being a chimpanzee. Lucy never got fully toilet trained, masturbated in public, and was allowed to consume alcohol. She became a habitual drinker, spending much time on the couch with illustrated magazines and a glass of gin, straight or with tonic. Browsing illustrated magazines was a favourite pastime.

But although being raised in a human home from birth into adulthood and always being fond of browsing magazines and looking at pictures, Lucy showed some extraordinarily non-differentiated behaviour towards photographs. Savage-Rumbaugh (1986), who worked with Lucy in the beginning of her career, writes: "When shown a photograph of a luscious ripe apple, for example, she would attempt to take a bite of it. In fact, there were few magazines to be found in the house which, on pages depicting delicious fruits, did not bear the marks of having been 'tasted' by a chimpanzee" (p. 300). Lucy also responded to cat and dog photographs, and to photographs of people engaged in hugging, kissing or dancing. More so, when in oestrus, and allowed to browse pornographic magazines, Lucy got most excited by encountering nude males in the pictures. She would stroke and scratch depicted erect penises, and only penises (Temerlin, 1976). She also performed sexual acts on the photographs by rubbing her genitals against them, again seemingly targeting the penis of the depicted models.

Now, if these behaviours should be analysed in terms of participation with the depicted scene, but with a full understanding that it is not real (i.e. Premack, 1976, see section 9.3), it is unclear to me why the outlet for Lucy's excitement had to be targeted towards the pictures as such. Could she not masturbate as she usually did,[89] or perhaps signal sexual interest in the bystanders? Similarly, if she found photographs of fruit to whet her appetite, why not go and find a fruit? Because these events were not about penises in general or fruit in general, but about specific penises and specific fruits, i.e. the penises and fruits directly in front of her.

The question whether Lucy viewed pictures as pictures was again highlighted in an episode where she was browsing a magazine that included an article about herself. Her cat had died three months earlier, and when encountering a photograph of herself with her cat she stopped browsing and repeatedly signed "Lucy cat" while staring intently at the picture (Temerlin, 1976). This transfixion went on for about 20 minutes. No social context is given for the account but it seems that Lucy signed aloud for herself. Also, no manipulation of the picture is described. It thus seems

---

[88] Fouts' (see e.g. 1997) language research builds on that of Gardner and Gardner (e.g. 1969).

[89] Temerlin (1976) is full of examples.

that the photograph was indeed viewed as referring to her own cat, and there is no overt indication, other than Lucy's transfixion, that she was confused by the fact that her cat was in the picture, "alive" and well. She did not seem to have tested the nature of the picture. However, her transfixion, which was described as "thoughtful sadness" by Temerlin, could in fact have been a case of "thoughtful sadness and something like confusion or wonderment" if rated by other observers.

8.5 Nim

Nim, the chimpanzee studied in Terrace's sign-language project, is said to have learned several of his signs from photographs (Terrace, 1980). Nim was apt at naming photographs and could browse a magazine or book for a particular picture that was asked of him. He also spontaneously named pictures when seeing billboards or browsing picture books and magazines. He enjoyed such browsing and often signed to himself when doing it. Generalisation to real instances of the depicted entities is not reported however.

A year after the work with Nim had ended Terrace reunited with him to see what he could remember from his training. Terrace (1980) describes the following episode which took place after naming some objects: "His interest in an old picture book was considerably greater. When I showed him the book he looked delighted and made the quiet "oooh" sound that I had often heard him make when he was relaxed and interested in something. Without hesitation, and without my even asking him to identify the pictures, he signed *toothbrush*, *hat*, and *dog* to the appropriate pictures" (p. 231). This account serves to illustrate not only Nim's fondness of picture books, but his view of them. They clearly served the purpose of being material in naming tasks. In the episode described Nim is not necessarily communicating to Terrace what he sees in the book, but rather responds to the drill of naming pictures, which he has gone through countless times. This indicates that pictures, as appearing in at least books, have a special status which other objects do not have. Such a status requires differentiation of some sort, but not necessarily a full one which entails pictures being *about* objects. In Nim's favour can be said that he is not reported to have acted on pictures, other than kissing a photographic face of a favourite caretaker.

8.6 Koko

The gorilla Koko, born 1971, is part of a language project since 1972, which continues to this day (e.g. Patterson & Linden, 1981). The modality chosen for communication with Koko was American Sign Language (ASL), modified to gorilla anatomy and motor control. Using similar methods and establishing the same kind of data as Gardner and Gardner (e.g. 1971) would make a direct comparisons between chimpanzee, gorilla and human children possible (Patterson, 1980). However, the research has not been limited to use and understanding of ASL. Spoken language has been used in parallel with ASL since the beginning and Koko's understanding of language spans both modalities. In fact she performs best on comprehension tests when she can view signs and hear speech simultaneously (Patterson & Linden, 1981). However, Savage-Rumbaugh et al. (1986) objected that at the time of the claim that Koko understood speech, no tests had been made that precluded contextual cues. Routine, intonation, glances, and gestures had proven crucial for other apes' ability to comprehend speech, at the exclusion of understanding spoken words as such (Savage-Rumbaugh et al., 1986). But in an updated view Savage-Rumbaugh et al. (2006) grant Koko comprehension of novel spoken English conversation.

The claims for the size of Koko's signed vocabulary has varied with criteria, from "over 1,500 words" (Patterson & Gordon, 2001) to "hundreds of signs" (Patterson & Cohn, 1990). The actual number of words is of no importance for this thesis, but it is important to acknowledge that Koko's language abilities is a result of being immersed, for a lifetime, in culture co-constructed with a few close caretakers, rich in linguistic and material elements, including pictures, mirrors, movies and toys.

However, the development of Koko's pictorial competence is only briefly touched upon in the only book about her first nine years of life, although they play a major part in her daily life, as well as in language training and testing. The information given reads: "[...] we had to teach her the great preponderance of her signs by showing her referents that were pictures or toys. Almost from the beginning of the experiment, Koko was forced to distinguish between representation and reality. The first few times Koko saw pictures of food in a book, she tried to eat the pages. But quite early she came to understand that pictures stood for something else not present. Indeed, the very constraints of her upbringing may have helped in this process of analyzing her world" (Patterson & Linden, 1981, p. 134). The constraints referred to is the fact that the infant Koko was confined to a zoo nursery and not allowed to mingle in the world.

How can we know that only because Koko stopped responding to photographs as reality did she start to view them as representations at such an early age? A somewhat circular argument is presented above. How can words learned from pictures be about the real world if the pictures themselves are a substitute for the real world? If no perceptual contact between picture and real world can be catered for, this connection has to be mediated by other means, for example through words. Might the infant Koko in some cases perhaps have learned the names for *picture* versions of objects that she later had to generalise to real life exemplars? In this case viewing pictures in reality mode would actually help with transfer from picture (photographs) to the real world, should real exemplars happen to be encountered eventually. It is reasonable that Koko only gradually came to realise the difference between pictures and reality, as she came into contact with more real exemplars. That picture-food was not edible was only one piece of the puzzle. As her vocabulary grew, further examples of objects appearing in the real world as well as in pictures could be pointed out to her, and an overall conceptualisation of pictures could develop

Unfortunately, there are no published direct tests of Koko's pictorial competence. Only indirect indications have therefore been collected in the following section.

On the pages of the book about Koko's training (Patterson & Linden, 1981), in a paper on her early language abilities (Patterson, 1978a), in films (e.g. Schroeder, 1978/2006), and in numerous pictures on the website of The Gorilla Foundation (2007), Koko has demonstrated an ability to comment on the content of photographic and cartoon-like stimuli. In Patterson (1978a) Koko can be seen browsing a magazine, looking at a page and holding her left index finger at her mouth. The caption reads: "Koko frequently signs to herself while playing alone and while looking through books and magazines. Here she signs 'toothbrush' (the index finger moved back and forth across the upper teeth) in response to a toothpaste advertisement" (p. 87). The photographic advertisement depicts a child holding a toothbrush and a tube of toothpaste. Interestingly, Koko signs about a toothpaste advertisement also in the television documentary *A Conversation with Koko* (Brennan & Visty, 1999), as well as seemingly mouthing the picture. Another instance of commenting pictures is pictured in National Geographic Magazine (Patterson, 1978b), in Patterson (1980), and again in Patterson & Linden (1981), which shows Koko looking in a home-made picture book made out of wood, containing animal photographs presumably cut out from magazines. On the open pages in her lap a lion and a tree frog with large eyes can be seen. Koko herself sits with her index finger pressed against her own eye, thus signing "eye." Another example in Patterson (1978b) is a rich interpretation of an event where Koko during an interaction with Patterson views a picture of a chimpanzee. What Patterson says or does in the interaction we do not know, but she points to the teeth of the chimpanzee. The caption reads: "Inspired by a grinning chimp, Koko enthusiastically signs, 'Teeth', before a plexiglas mirror […] She thus demonstrates that she recognizes not only the picture as something related to the real world, but also her own imitation of the picture" (p. 442).

Although the actual published examples of Koko commenting photographs amounts to a handful, and the history of her interactions with these pictures is never given, there is no reason to doubt that Koko can recognise objects in photographs.[90] The reason that published examples are few might just be an effect of the behaviours being common and thus perhaps believed to be trivial. Everyday examples can for example be found on the KokoPix photograph blog at the project's website. For January the 29th 2007 the entry reads: "Koko and Penny look through some issues of National Geographic Magazine with articles about mountain gorillas. Koko points to one photograph of a mountain gorilla family and signs 'gorilla there'" (The Gorilla Foundation, 2007). However, by now it should be clear that naming photographs is not enough to satisfy the criteria for a referential understanding of pictures. Reality mode would in theory suffice to account for adequate performance in this task.[91] But a clear sign of differentiation suggests itself in the naming examples, which is that pictures are conversation topics, while real instances of e.g. gorillas presumably are not, at least not in the same way. But again, it remains to be demonstrated that the direction in the relation between picture and referent goes from real gorillas to their depictions.

---

[90] Limitations, if any, in this ability remains to be settled, both for Koko and for other subjects that seem to readily recognise objects in most photographs.

[91] One can presume, given Koko's language training based on pictorial exemplars, that her gorilla concept and naming extend to small two-dimensional versions.

8.6.1 Koko's naming of drawings

More promising than the naming of photographs, as indicative of a fuller pictorial competence in Koko, is her ability to adequately respond to the content of non-photographic pictures, like line-drawings and cartoons, especially if the pictures can be shown not to have been used in conversation before. These requirements are unfortunately not fulfilled for the case presented in Patterson (1978b) or Patterson and Linden (1981) where Koko is commenting that the cat mother is angry, the kittens are crying and that something is "bad" in connection to a story read to her by Patterson. The story is illustrated with colour drawings. But we do not know if Koko is commenting the pictures or the telling of the story, and we do not know if this is the first time the book is ever read to her.[92]

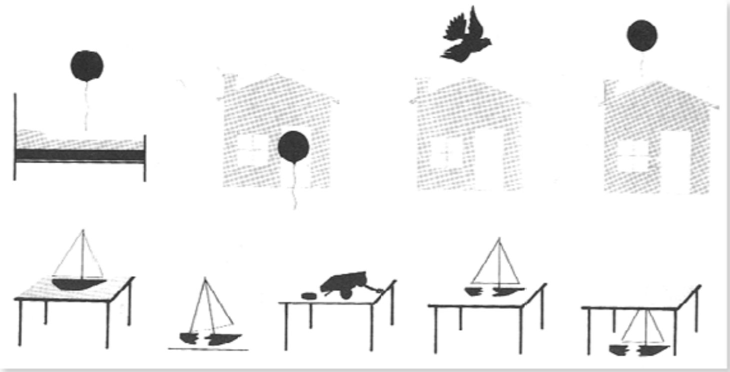
In Patterson and Linden (1981) a second anecdotal illustration of Koko responding adequately (from the point of view of a scientist) to a line-drawing of unknown history is given. "As part of a vocabulary test, Penny asks Koko to find 'crying' and Koko points to a picture of a child crying" (p. 56). The point of this illustration is to demonstrate Koko's linguistic capabilities. If the pictorial part of this or other events would have been unusual for Koko, a point would probably have been made about it.

Koko is subjected to non-photographic pictures (as well as photographs) in the regular testing of her vocabulary. The only readily available published data for such a testing session that involves pictures, where the types of picture are specified, and where they furthermore might be suspected to be novel to Koko, is a test using the Assessment of Children's Language Comprehension (ACLC) material (Patterson & Linden, 1981). This test was given to Koko at the age of 4.5, with potentially impressive results from a pictorial perspective. It is reported that Koko did not receive any drilling or training before the test, which can be taken to mean that also the test material, i.e. the pictures, were novel to her. However, it can also mean that only those aspects intended for testing, i.e. linguistic comprehension, were not preceded by drilling, which excludes the pictures as such. This is even likely since one would like to make sure that the subject understands the testing material. A couple of the pictures used are shown in Patterson & Linden (1981) (see *fig. 11*). They described different objects, attributes and relationships between objects. In the vocabulary part of the test Koko had to point to the appropriate item in a picture, and in the comprehension part she had to choose which scene among four or five that depicted a specified relation between objects. For example: "point to the bird above the house" or "point to the broken sailboat on the table." The latter example thus implies, pictorially, recognising a broken sailboat, a table, and the relation between the two. Interpretation in relation to the non-matches had to occur. The complete test contained 40 picture cards of with 30 depicted scenes like this.

---

[92] In *A Conversation with Koko* (Brennan & Visty, 1999) the very same book is described as one of her favourites.

Figure 11. An example of the drawing stimuli used to test the gorilla Koko's (4.5 years) receptive English vocabulary. From Patterson and Linden (1981).



Koko performed significantly better than chance and "slightly less accurately than educationally handicapped children" (Patterson & Linden, 1981, p. 103). The results were difficult to interpret, though, because Koko performed steadily across problem difficulty from a linguistic perspective (the number of critical elements in the request), but she made several errors on the whole. Patterson did not entertain the idea that perhaps the pictorial stimuli as such were the cause of these errors. She attributed the performance to attitude factors, such as boredom with the task, although Koko did receive motivating edible treats for her work. Unfortunately it is not stated whether the pictures were reused or not. The 10 vocabulary pictures were only used in one condition, but the remaining 30 were used in nine different conditions. Information on how the pictorial stimuli were distributed among these conditions is not given. If pictures were reused, trial one data is needed, otherwise we do not know if there were learning effects for the pictorial content.

Also at age 4.5, Koko seemingly outperformed human 6 year old children on same – different judgements, as well as detecting "flaws in a series of incomplete or distorted drawings" (Patterson & Linden, 1981, p. 127). The picture-completion problems included drawings of hands missing a fingernail, or a cat with only one set of whiskers.[93]

And when she was 6 years old Koko received, among other tests, "form B of the Peabody Picture Vocabulary Test," which (to my knowledge) contains drawings of people in action. Patterson admits that Koko was given the same test many times, in order to acquire repeated measures of her performance, but also that familiarity did not increase her scores, which were very good from the start. On the contrary, Koko became worse with repetition, seemingly due to boredom with familiar tasks (Patterson & Linden, 1981).

A trial from a task similar to the ACLC described above, probably using pictures from the original set, can be seen in the documentary *Koko: A Talking Gorilla* (Schroeder, 1978/2006).[94] At the time Koko is about 6 years old and she pinpoints the "bird in the basket" without a flinch.[95] The three non-matches seem to be a bird outside of a basket, something in a cart, and lastly an empty basket. When requested

---

[93] Viki, the chimpanzee in Hayes and Nissen (1971), also showed that she can discriminate complete from incomplete drawings in a similar manner (see *table 2*, section 7.3).

[94] This documentary is a valuable record of Koko because it contains long unedited episodes of interaction.

[95] She performs her indication by kissing the image, an important reminder that not all mouthing is investigation.

to find the alligator in another vocabulary trial she kisses, then points to, a line-drawing of an alligator. Non-matches are a mouse, a kangaroo and a goat. In a productive vocabulary (and pictorial) test[96] she is asked to name line-drawings depicting a tree, a hat and a knife. She fails on the knife. The potential difference between “receptive” and “productive” pictorial comprehension will be expanded upon in Chapter 13.

As previously, Koko’s history with the pictures in the above episode is unknown, but her behaviour towards pictures often suggests that it is not merely a case of memorising previous answers or conversations. An example from the above documentary (Schroeder, 1978/2006): Koko, lying down by Patterson, browses a picture book that contains drawings. She starts out by viewing the first page for a while and taps a few times with her index finger at the sea lion enclosure in the middle of a zoo, an action that is difficult to interpret. She then turns the pages quickly with her lips, seemingly glancing at each page. She stops and points to a group of assorted flowers. As spectators we cannot know why she does this but the voice of Patterson hints to us that they look beautiful to her or Koko. Then there is an editing cut, and in the next frame Koko points to a red square and signs red, her favourite colour. She seems to address this to Patterson. Then, on the same page, she points to another square-like image, but yellow, with a small spider depicted in its centre. The tapping of Koko’s finger is targeted at the spider. Koko emphatically hits her hands together which Patterson translates “that is a scary bug, scary bug spider” while Koko sits up.

It is difficult to judge what modes of pictorial processing are operating here, colour is a feature easily processed in surface mode and it could have been colour and not “flowers” that were extracted from the flower picture etc. Bugs and spiders are one of those things that often look quite bug-like even as drawings.[97] However, there is an element of involvement, if not captivation, in all of the latter three actions towards the pictures, which hints to us that Koko actively acts towards, and perhaps about, these pictures. Furthermore, she is in no way guided in her responses by an interaction with Patterson, other than perhaps by the memory of earlier rewarding interactions around the same or similar pictures. The script she is going through of pointing and naming and socialising with a caretaker might all be part of a drill she has developed through her countless sessions with picture material, but her actual choice of pictures to respond (adequately) to does not seem to be part of a stereotyped drill (perhaps with the exception of the sea lion). Her reactions seem to unfold spontaneously as she browses through the book. Unfortunately, a few seconds of data from an edited documentary do not allow for the full necessary analysis. However, the point is that *how* an animal, or human, behaves in a task, be it informal or experimental, is sometimes just as important as the figures describing the result. Had we seen the same actions in a human child we would not doubt that the child ac-

---

[96] The terms “receptive” and ”productive” are borrowed from language research where a difference is made between e.g. receptive and productive vocabulary. In this thesis I use ”receptive” when the investigated picture is on the choice-item side in MTS, and ”productive” when the picture is on the sample side in MTS, because that is how they appear in language testing.

[97] A drawing of a bug encased on the backside of a laminated drawing of an apple received close visual inspection by a bonobo in a test session (see section 13.6). The apple received less attention.

tively interprets the pictures. Human children also develop habitual social responses towards specific pictures in picture books (i.e. "stable reading routines" in Fletcher & Reese, 2005), but that does not mean that they cannot be stereotypical and recognise the content of pictures at the same time. Habits do not need to be a sign of drilled, rote learned, behaviour.

Is it possible that Koko's performance is due solely to a good memory for paired picture – referent associations? When attempts were made to introduce a voice computer to Koko she learned to use it quickly and transferred her words to the arbitrary geometrical coloured shapes on the keyboard (Patterson & Linden, 1981). She soon typed out requests like "want apple eat" and "want drink sip." If Koko can learn to ascribe meaning to new symbols easily (unfortunately we do not have a learning curve on the ability) perhaps she can also learn the meaning of pictures in this way, i.e. is as symbols rather than icons, bypassing visual similarity. By applying iconicity between pictures, rather than to the real world, learned meaning has the potential to generalise to novel pictures as well. If this is a possibility the prediction would be that shapes that have a tendency to remain the same in different pictures would be more frequent in Koko's comments on drawings, than shapes that have a large variability.

I have for example noticed in photographs and movie clips of Koko on the project's website (The Gorilla Foundation, 2007) that several comments of hers towards drawings on greeting cards, patterns on clothing, and even engravings on a cufflink (which was actually of a sun), are examples of flowers. A flower, as it is typically portrayed in handmade pictures, might be one of those stable shapes that easily transfers and is recognised across depictions (as well as potentially becomes confused with certain suns). That said, looking for depictions in cufflinks and clothing, besides on paper, speaks for an ability of Koko's to spontaneously look for iconically based meanings in objects, even though it might be limited to certain motifs, like flowers. Another interpretation, though, is that when flower-like motifs hit Koko, she comments on them. This would be the very opposite of looking for faces in clouds. A targeted study of this ability is greatly needed.[98]

8.7 Ossy

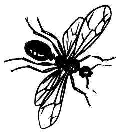
The infant orangutans in the care of Harrison (1964) all showed an interest in pictures, especially of human faces, in magazines. This was especially true for one of the animals, Ossy, who at the time ought to have been between 6 months and a year old. He targeted with kisses the eyes, noses and mouths of even very small pictured faces. It did not matter for him if the pictures were upside down or upright, further implying that the typical orientation of pictures is a conventionality.

---

[98] An example of another improvisation, which suggests that at least language-competent apes spontaneously appreciate iconicities in unusual places, is the lexigram name "mushroom" given by the bonobos at the Language Research Center to a female visitor with a very unusual hairdo (Segerdahl et al., 2005). According to the authors the list of such metaphors could be made long.

Harris experimented with different pictures but what apparently stood out enough to report, except the response to faces, was Ossy's vivid interest in pictures of leaves and flowers. These he poked with his fingers or tried to bite. He could do the same to flowery patterns on clothes. One would presume that fabric would facilitate differentiation, since clothes are a well known material, but if there is any pictorial content, viewed in reality mode, that would make sense on a body it is foliage. Draping oneself in leaves and flowers is not at all a farfetched behaviour. Given Ossy's young age it is implausible that he was engaged in pretence. Small faces on paper, or flowers on a skirt, are apparently stimuli similar enough to one's real world experience to make it worthwhile to investigate them further with hands and mouth in a way typical for objects. That said, a problem is that Harris only reports what stood out to her. Ossy's response, or indifference, to other pictures than faces and foliage are an important piece of the puzzle. And did he ever poke and bite pictures of faces, or kissed leaves?

Sometimes observers are criticised for offering mere anecdotes, which is held to be unscientific. But the problem with anecdotes is not their subjectivity, but in being highly interesting and suggestive, they often create more questions than they answer (which is as it should be).



Chapter 9

Pictorial sister competencies

Because pictures come in many forms there is an apparent problem with describing the understanding of them as understanding a single class of artefacts. That an individual learns how to decode e.g. drawings does for example not mean that it necessarily stops responding to photographic media as if it contained properties of the real. A general picture concept develops, if ever, through experience with the media and the analogies between these. If one is to claim that apes can be pictorial, or that they cannot be, one must therefore cast the net wide and look at all domains that contain iconic artefacts.

In the previous sections the focus can be said to have been on "marks on surfaces." But icons are not necessarily pictures. There are several other types of iconic media in the life of apes in human environments. This chapter will, in turn, consider toy replicas, scale models, video recordings, and mirrors. These media do not only share their iconic basis for meaning with pictures, the three modes of picture processing can likewise be applied to them. If apes differ in how they appreciate these media it can teach us more about the factors involved in iconic meaning making in nonhumans. If they can understand any of them in a fully fledged referential manner there is increasing hope that this would be possible also for pictures.

9.1 Replicas, toys, and pretence

Washoe, in her sign-language training, was often trained using photographs instead of real instances of the intended referents. This was done out of convenience, for examples when objects were too large, dangerous or expensive. Animals and cars were such categories. But the trainers wanted Washoe to learn about as many instances of a category as possible and therefore included also small replicas in the training. These were chosen to be as "realistic" as possible (Gardner & Gardner, 1971). However, a striking discovery was made in a test where a photograph or a miniature replica was placed in a box and Washoe was required to name, with the appropriate ASL sign, the item displayed. For the categories "bird," "cat," "cow," "dog" and "ride" (car) Washoe was quite good at naming the photographs, but with the miniature replicas she only got 3 out of 14 right. Furthermore, in half of her errors she had signed "baby" instead of the correct sign. When making an error with a photograph of these objects she signed “baby” only once, and in general she did not sign “baby” during testing.

When again tested with miniatures for “bird,” “cat” and “dog” Washoe could name 6 of 10, and on all 4 mistakes signed “baby.” On 7 out of 8 photograph trials from the same categories Washoe named them correctly, and on her one error she did not sign “baby.” Washoe had a correct notion of “baby” in that she correctly labelled photographs of human babies and dolls, and her use of “baby” did not seem to have reflected size, because photographic instances of objects were also reduced in size (Gardner & Gardner, 1971). However, her notion of doll (i.e. one meaning of “baby”) could have extended to all instances of small, three-dimensional, animal-like objects. This label could have overridden the specific animal types the miniatures represented. Alternatively, Washoe could have recognised such items as some kind of animals but not readily make out which.

Besides suggesting that miniatures can be less real to life than photographs, this finding serves also to prove the revealing power of productive naming. Had Washoe instead been asked to find, receptively, the e.g. “cat” among three miniatures, she might have fared significantly better because she would then be looking for the most cat-like object. In productive testing, on the other hand, she had to interpret the object totally on her own accord and not be helped by context.

There are only a few studies that have used model replicas that require recognition in testing. Murai et al. (2005) could for example not find any habituation on a categorical level in infant chimpanzees’ (1-2 years) spontaneous exploration of miniature models from the categories mammals, furniture, and vehicles. However, they found support for novelty preference on a “categorical” level, but this does not imply recognition. It can have taken place either in a surface mode, or through the forming of new categories for the objects involved. It is implausible that the replicas were recognised as miniature versions of objects the subjects had very little experience with. Such experience is arguably more necessary for replicas than for photographs since a photograph affords properties of the real that pertains to the referent, which a replica does not. From a reality mode perspective a photograph can thus be recognised as something real but novel, while the model remains just an object.

The gorilla juvenile Zuri in Vonk and MacDonald (2002) seemed to categorise photographs of animal statues and sculptures as objects rather than animals. If this was because she did not recognise their similarity to animals, or that fake animals did not pass as animals in her understanding of the task, is impossible to tell.

9.1.1 Dolls

Remember the examples with Köhler’s chimpanzees and stuffed animals from the first chapter. Stuffed animals scared the chimpanzees very much, more so than did real animals. Köhler found out that the more realistic the stuffed animal looked, the stronger the response. These findings are suggestive of responses made in a reality mode. It was a fear of the unknown, according to Köhler (1925/1957), but equally important was the likeness to the known.

But stuffed animals do not only evoke fear, they can also be enjoyed in play. Using objects in play is a common animal behaviour, not the least in ape play. A toy is an object used in play, whether it enters into play accidentally or is sought out intentionally. Toys can also be provided. Manufactured toys can be specifically designed for exploration, sensation etc. of the toy as such, or they can be designed for pretend play. In that case the toy refers to a model entity and the practices coupled with such an entity. A toy car or a doll, are typical examples. Such toys are almost without exception iconic in one way or the other. An important question is what role iconicity plays in ape actions on toys. If pretence alone was the driving force in play, would not a piece of arbitrary wood or a rock be a perfect doll for an ape?[99] There is reason to suspect that iconicity does indeed influence play behaviour. In enculturated subjects it seems that the idea of nursing an object, or pretend to bite someone with it, is evoked more by objects with mouths (e.g. dolls) than objects with for example wheels (e.g. cars). Drinking pretend tea out of miniature cups (see Patterson, 1978b) is arguably more common than having pretend tea out of rocks. However, I will argue that it is not always clear whether pretence is involved in actions like these, but that using iconic objects in proper pretence is a necessity for inferring a referential use of them in play.

“Throughout her life, Koko has had a variety of dolls, stuffed animals, and pictures of animals to play with” (Patterson & Linden, 1981, p. 134). Since toy animals and dolls have been an integral part of Koko’s (and other language trained apes’) upbringing and language training, it is difficult to exclude that dolls have formed categories of their own, together with appropriate actions on them, through imitation and interaction. Such categories could have very little to do with them being representations of real beings. The similarities between doll bodies and human or animal bodies has also been pointed out during such training, perhaps even more so than in picture discourse, and it is not farfetched to imagine that doll and human categories would have substantial overlaps. When such likeness has been established doll “eyes,” “mouth,” “hands” etc. can borrow properties from their live counterparts, but this does not necessarily make the doll stand for a real body. It very much *is* a body, just a lifeless one. Any “life” that is then blown into this body might not be collected from experiences in the real world, but with experiences from the doll world, i.e. what has been modelled and picked up in interaction.[100]

Showing maternal behaviours towards objects is perhaps not the best indication of pretence since such behaviours are shown by individuals that have no experience of such behaviours (either than towards themselves) and therefore seem likely to contain instinctive elements. Using a dead animal (Savage-Rumbaugh et al., 1998), or “stones, shoes, balls, pieces of cloth, and even a rubber doll” (Gómez & Martín-Andrade, 2002, p. 259) as if it were a baby does not mean that it represents a baby in any way. In fact, behaviours that involve object substitution are suggested by Gómez & Martín-Andrade (2002) to be cases of using objects *instead of* as opposed

---

[99] One reason is of course that it is difficult for a bystander to interpret “doll behaviour” with a piece of wood. An example of such a case is given in Wrangham and Peterson (1997).

[100] Lyn et al. (2006) suggest that all pretence in apes, and young children, is the outcome of scaffolding processes before it can become an independent and self-initiated act.

to *as if*, following Mitchell's notion of "schematic play" (1990, in Gómez & Martín-Andrade, 2002).

As real babies enter the lives of these individuals maternal behaviour can be calibrated to the proper entities. Some objects are just particularly suitable substitutions for the real thing. Alive kittens, cats, rabbits etc. have by several species of great ape repeatedly been treated in ways similar to caring for ape babies (e.g. Hess, 1954; Patterson & Linden, 1981; Temerlin, 1976). Is it appropriate to talk about pretence, or a *stand-for* relation, in those cases because kittens are not apes? Certainly not. Likewise, stuffed animals and other inanimate objects seem to sometimes be good substitutions for babies and evoke maternal care as well (*fig. 12*).
Another example of substitution without the need for pretence is generalising a ball game that was previously performed with tennis balls, to apples (Gómez & Martín-Andrade, 2002). Although physical similarity is to blame for this generalisation, the apple does never need to *stand for* a tennis ball in such games.

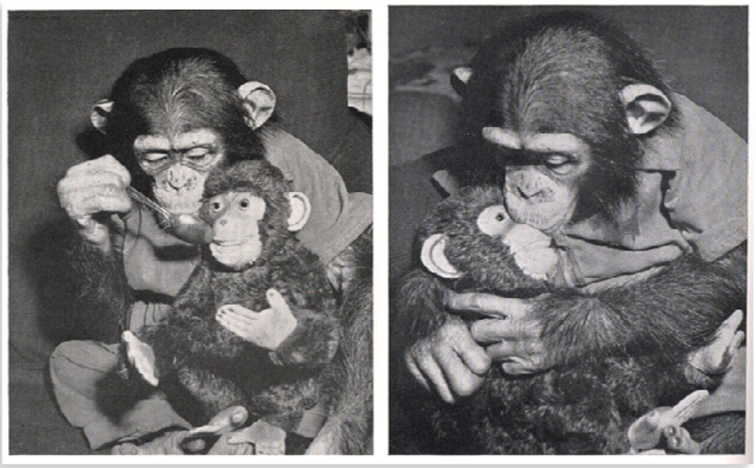


Figure 12. Christine, the chimpanzee in the home of Hess (1954) "feeds" and kisses a (doll) chimpanzee. From Hess (1954).

Also, objects can be played with in a manner that is neither *instead of* nor *as if*, although it might *look* like it. A bout of object play can look like baby pretence because it is affectionate, or involves actions such as cradling. Taking a fondness to an object and keep it with oneself for a prolonged time is not automatically doll behaviour. What can be best described as "being involved in a little project," like carrying around a seemingly arbitrary object, is common ape play behaviour. Such an example is given in Wrangham & Peterson (1997) for a young male who was observed carrying baby-sized logs with him on two separate occasions, even taking a nap with one of them in his nest. Independent observers judged the behaviour to be reminiscent of carrying a doll. However, carrying a baby-sized log is very likely to remind of the way a baby-sized baby is carried, especially if it is coupled with gentle rather than rough play. No further relation between log and baby is necessary.

However, turning the argument around, if indeed this was a case of doll play, did the chimpanzee chose baby-sized logs because they were baby-like,[101] or did the weight and size of them induce such play? The latter case can be said to be equivalent to "pretend" eating off photographs but not off arbitrary media. The likeness to a baby is secondary to searching out such a likeness. The difference is significant and

---

[101] This would rather be a case of secondary iconicity, i.e. expecting a likeness in a differentiated object before such likeness is found.

highlights the need to separate pretence that contains imagination from pretence that does not (Mitchell, 2002). When a doll can put things in its mouth but not be imagined to swallow or actually eat we have reason to question the imaginativeness of the doll play. The playing individual does not fill in what is not there. That might be why iconicity is potentially a powerful scaffolding factor in pretend play. The step to imagine that a doll with a mouth can swallow is arguably smaller than the step to imagine that a stone can swallow.

Koko often signs to her dolls (or herself) in private play, but sometimes she also lets the dolls sign. She occasionally lets them sign for example "drink" by moving the doll's hand to its mouth (Brennan & Visty, 1999; Matevia et al., 2002). In the cited example this was made in response to the question "where does the baby drink?" (Koko signed "mouth" directly after signing "drink.") But when she does such signing spontaneously, is it pretence or a routine that she has picked up somewhere? It does by all means not need to be a blind routine. Koko can very well understand that she is making the doll sign "drink," if nothing else because her caretakers have commented the event innumerable times, but how many other signs can she make her dolls sign? "Eat," "drink," "more" and "mouth" is mentioned in the Brennan and Visty (1999) documentary. That is, how open ended is her doll repertoire? A prediction would be that a repertoire is more open ended if dolls are seen as representations than when they are "just" dolls, because in the latter case only doll-specific interactions have taught Koko about dolls. Interestingly, Koko is said to only make her ape dolls sign, and not her human dolls or other toys (Brennan & Visty, 1999). A possible line of investigation could be to see whether caretakers also have treated ape dolls differently from other dolls in the past or if this is a discrimination that stems from Koko herself.

That dolls' likeness to real faces is often appreciated is evident from the literature, which is full of apes that kiss dolls, makes dolls kiss or bite others (including other dolls), feed dolls, cradle dolls, and put dolls to nipples (in the case of Koko). It is thus clear that at least mouths are parsed. But why this oral interest? Why not make dolls walk around a bit? Fight? Have sex with each other? These are all common ape behaviours, but they are perhaps not the games humans engage apes and dolls in. Furthermore, they are behaviours that require more of an attribution of intentions to the dolls. Mouths, for example, can afford biting and feeding without such attribution. But why would the doll want to walk around a bit? The modelling of contact actions such as feeding and biting is perhaps for the same reason more transparent than many other types of pretence picked up in interaction with humans.

Koko parses hand and mouth on her dolls of various looks and species. Alligator mouths are always easy to spot, and they are all scary until you cut the teeth out. It is striking that the actions on dolls are quite proper and very habitual. For example, Koko lets them often nurse nipples, they all get kissed on the mouth, and they all bite with their mouths. These might be limited cases of pretence, but they are good cases of appreciating (primary) iconicity. Identifying a vast range of mouths, hands, eyes, and other anatomical features of dolls is an impressive feat.

Other examples are the chimpanzee Sherman, who made King Kong dolls bite his fingers and toys (Savage-Rumbaugh & Lewin, 1994) and the bonobo Kanzi makes toy dogs or gorillas bite him or others (Savage-Rumbaugh et al., 1998). Also, Kanzi carries around dolls as if they were younger companions that he can hug, play bite, tickle, and share food with. However, he is said to grow tired quickly at such games, since the dolls do not play back (Savage-Rumbaugh et al., 1998).

Koko treats also very small dolls as babies, cradling and nursing them. Similar behaviours have been reported for the chimpanzee Viki who kissed miniature dolls whose mouths were minute, opened and closed doors and drawers of miniature furniture, and scribbled in a miniature notebook with a pencil the size of a match (Hayes, 1951). Although claims of pretend play has been made for Viki (see Mitchell, 2002), size reductions in themselves do not necessarily turn an object into a model.

"Scale errors" is a common phenomenon in children's play (DeLoache et al, 2004b; Ware et al., 2006) which entails making striking misinterpretations of the sizes of miniature toys. Children 18 – 30 months old for example try to sit on tiny furniture or fit themselves into small cars. Although scale errors have been discovered in the context of play the errors themselves are not pretence. The expressions of the children tell that they are quite serious in their intentions. This might be one of the many effects of reality mode processing, i.e. generalisation without differentiation between two classes of objects. Scale errors have been attributed to a planning-control mismatch, where planning of action is based on the expectations on a previous larger version of an object, while motor execution adapts to the actual smaller version (Glover, 2004). However, scale errors do not only involve one's own body, but also applies to external object relations like that between dolls and beds (Ware et al., 2006). The scale error phenomenon is a strong argument against Viki's and Koko's miniature use as automatically being one of pretence.

Given the striking phenomenon of scale errors, it is not surprising, when it comes to photographs, that children and animals can act on a two-dimensional surface as if it contained graspable properties. This acting out in a reality mode can be dependent on a similar dissociation between the affordances of one's object recognition, and the affordances given off by the actual flat surface. The grasping hand movements do adapt to the flat surface, but it does not seem as if this manual experience feeds back and updates one's expectations. Hence the persistence of the grasping children in e.g. Pierroutsakos and DeLoache (2003) (see section 2.1).

9.1.2 In the eye of the beholder

The sign-language trained chimpanzee Washoe was also particularly fond of dolls among the toys that were made available to her, and she kissed, fed and bathed them (Gardner & Gardner, 1985). Washoe first bathed a doll of hers spontaneously when she was 2 years old, and it was interpreted that she imitated the way she herself was washed (Gardner & Gardner, 1969). However, episodes when she cleaned other objects are not reported. We can thus not assume neither that the perceived likeness to a baby of some sort accounted for the behaviour, nor that there was an *as if* rather than *instead of* relation involved. A similar problem is evident in the following account: "Washoe, Moja, Pili, Tatu and Dar signed to friends and to strangers. They signed to each other and to themselves, to dogs and to cats, toys, tools, and even to trees" (Gardner & Gardner, 1989, p. 24). How would we be able to judge whether toys have a special status among these items?

Interpretative problems also occur when we try to make sense of Koko signing "hat" when a caregiver puts a strawberry stem on the head of an alligator (Matevia et al., 2002). This might be a creative response that goes outside habitual play with alligators. We know with some certainty that she can parse the alligator's head. But we must also make certain what "hat" means to Koko. Is anything put on a head a "hat," or are real hats more specific than that? If the strawberry stem was put on the alligator's back, would that elicit "hat" as well? Overextensions are otherwise not uncommon in ape language-use (e.g. Miles, 1990; Patterson & Linden, 1981).

Koko also has a fear of toy alligators, regardless of material, although she has never seen a real alligator. If they are broken and the lower jaw is missing her fear vanishes. Patterson and Linden (1981) suggest that it is the teeth that are dangerous. Rubber alligators and snakes are often used to bite other people in play. The fear of toy alligators seem to have became a pretend fear with age, but nevertheless they are scary things. The exact nature of the fear is difficult to interpret though. Koko can refuse to touch a toy dinosaur and show signs of "true" fear, while later the same day joyfully play with it (Matevia et al., 2002

Rubber snakes also have a special role for the bonobos and chimpanzees in Lyn et al. (2006), being toys bordering on the real. The excitement displayed by the apes is very real, but is the cause for this excitement the rubber snakes as such or what they allude to? The difference between these snakes and Koko's alligators is that the bonobos have most certainly seen real snakes, a common animal on the grounds where they resided at the time.

There are of course other, non-trivial, indications of pretence that accompanies the action as such. The context in which it occurs, the mood of the situation, and the particular way of execution are all dynamic aspects of the event which can be indicative of pretence for someone who knows the individual well. Such nuances can be very difficult to convey to someone without extensive experience of the individual ape. If an animal would "try to eat" the food off a picture in a general air of silliness (or otherwise atypical attitude for real eating), the action being initiated by the ape itself and unfolding independently of feedback from a social partner, I would not invoke reality mode as an explanation for the behaviour.

If the ape performs novel actions on a replica toy which entails evoking properties that pertain only to the referent (e.g. ability to swallow in a doll), without expecting that the toy will actually fulfil these properties, it can be said that the toy has been used as an icon, i.e. an iconic sign. The novelty requirement is to preclude that the action is learned from previous interactions. Our concern here is thus how the apes spontaneously handle iconic toys independently of social scaffolding, and whether the actions can be said to pertain to the referent, or just the toy as a class of its own. If the ape is told that a doll is a baby, or suggested that the doll can be fed, the resulting actions on the toy can be guided by iconicity (e.g. finding the mouth), but cannot be said to be an act of pretend "eating," and consequently not be said to refer to a referent with certainty. Sticking things in mouths of dolls can be done without reference to living, eating bodies. For example, when asked to feed the doll in one of the examples in Lyn et al. (2006), a grape that Panbanisha puts in the doll's mouth falls to the floor. Panbanisha then pushes the doll's head to the floor to seemingly continue the play act of eating.[102] However, is mouth-to-grape contact important to Panbanisha because Panbanisha *pretended* that the doll had not finished the grape, or because the doll had *indeed* not finished the grape? Why did the doll's mouth have to be in contact with the grape to pretend-eat it? When it fell to the floor, could that not as well have signified that it was eaten, since the doll could not *really* swallow a grape? These are questions for future empiric consideration. It should be stressed that these suggestions are not specific to ape research but to pretence research with children and other animals as well.

Eating "imaginary" food out of a bowl with a spoon (Savage-Rumbaugh & Lewin, 1994; (Matevia et al., 2002), or drinking from an empty cup does not automatically suffice to evoke pretence as an explanation. Going through habitual manoeuvres with objects can often be ascribed to routines, or schemas, that are heavily linked to the objects (Gómez & Martín-Andrade, 2002). However, when the imaginary food gets physical properties, as when Austin rolls imaginary food around in his mouth (Savage-Rumbaugh & Lewin, 1994), imaginary liquid creates slurping noises in Koko's mouth (Matevia et al., 2002), or Panbanisha chews imaginary food that she has grasped from a photograph (Lyn et al, 2006), it is questionable that it is just a matter of a continuation of a drilled eating schema. The most convincing case is perhaps when the imaginary food is not directly consumed, but placed somewhere (in this example by Kanzi), and can stay there for several minutes until it is either moved again or "consumed." Sometimes the imaginary food is quickly retrieved if somebody gets too close or tries to steal it (Savage-Rumbaugh et al., 1998).

Kanzi's and Panbanisha's games with imaginary food are often started by seeing food in a photograph or on a television screen from which it is grabbed (Savage-Rumbaugh et al., 1998; Lyn et al., 2006.). In the developmental account given by Lyn et al. (2006) the first "pretence" in young apes at the Language Research Center seems to be revolving around eating food off photographs. It is possible that the start of this development is a reality mode induced testing of the pictures, which is then moulded over time in interaction with the caretakers into a social practice and play. Human imitation of the practice, as well as initiation, and comments such as "are those good M&Ms?" seem to be common for imaginary-food episodes. Such interactions are probably rewarding and the ape will initiate the game again when it encounters photographs. The step from investigating photographs to grab invisible objects from them is not dependent on a transition into a pictorial mode, and it might be the reason we see this "representational" use of photographs at an early age, as the first sign of pretend play.

Kanzi is also said to enjoy feeding imaginary food to toy dogs etc. In a study of comprehension of novel sentences in spoken English (Savage-Rumbaugh et al., 1993) Kanzi responds to the request "feed your ball some tomato" by looking for a

---

[102] The non-language competent controls in Lyn et al. (2006) on the other hand showed very little proper response to language mediated interactions with a human with doll. They did not feed dolls or otherwise showed that they recognised their status as replicas, but hugged, slapped and bit them irrespectively of scaffolding.

ball with a pumpkin face embossed in its spongy material, and orients the ball so that he can touch its mouth with a tomato. Worth noting in regards to language comprehension is that Kanzi had never been required to feed balls before. In addition, the face on this particular ball had never been pointed out to him, nor had he acted on it as a face (Savage-Rumbaugh et al., 2006). The question then, is whether Kanzi had spontaneously appreciated its likeness to a face all along, or whether this became apparent to him only with the scaffolding help of the request. It should be said that the face depicted in Savage-Rumbaugh et al. (1993) seems to be a very striking face in marked three dimensions, but it is not realistic looking in that it is exaggerated and cartoonlike. It is a caricature of a face. Who knows how many resemblances Kanzi perceives in surfaces in his surroundings, but which goes unnoticed by bystanders because he does not spontaneously comment on them. A study in Part III tries to address this question by having him view non-realistic pictures *before* being scaffolded by language in his interpretation.

The sign-language trained orangutan Chantek also fed his toys, as well as signed to them (Miles, 1990). Again, it is unclear if the behaviour is directed towards the toy because it is a learned behaviour, because the toy itself is incurred with imagined animate properties, or because it is perceived as a replica of an animate object. Chantek's feeding his toy animals started before he was 2 years old. At the time he scored on the "Bayley Scales of Infant Development test" the equivalent of a human mental age of just above one year.

Christine, the infant chimpanzee described by Hess (1954) did not receive language training. Still she fed her dolls, especially her favourite one, from a tender age (see *fig. 12*). Whether imagined food was involved is not possible to tell. The chimpanzee Nim, who did receive language training, fed his dolls real food. Terrace (1980) makes an interesting comment in connection to this, namely that Nim at least must have understood that there was no danger of losing one's precious food when offering it to inanimate mouths. This suggests a possible empirical investigation to see if apes are reluctant to put their favourite food in the mouth of dolls. There *might* be a risk of losing it after all.

Savage-Rumbaugh et al. (1998) comment that the bonobos' imaginary food play lacks the elaborate structure of children's tea parties, but this, I suggest, is because such play does not refer to tea parties but are games developed with specific objects in specific contexts. This can be limited to making slurping noises when pretending to drink from a cup or bowl (see e.g. Matevia et al., 2002).

Much behaviour on objects in human children, which would usually fall into the category of pretence, seems to be heavily dependent on a scaffolding interaction with a more advanced individual. When the supporting roles of imitation and language are removed children below the age of 2 years do not seem to use replica toys as representations. Neither when used in a communicative setting, nor when allowed to freely play with them (e.g. Tomasello et al., 1999). Although 3 year olds are more creative on their own accord in pretend play with objects (not necessarily iconic ones), they are also significantly more affected by scaffolding language than are younger children. Thus, less ability *and* less effect of scaffolding is seen in younger children, while larger ability *and* larger sensitivity to language scaffolding is seen in older children. Importantly, both groups are heavily influenced by the iconicity of the toys, especially the younger children (Striano et al., 2001). Lyn et al. (2006) report that the apes at the Language Research Center were also heavily influenced by scaffolding in their pretend play.

It also seems that where very young children fail with replicas, they can succeed with iconic gestures (Tomasello et al., 1999). This is explained by Tomasello and colleagues by DeLoache's dual-representation model, i.e. that the strong object status of the toys themselves hindered a representational view of them. The children did indeed often reach for the toys, which they did not do for the gestures. When it comes to the iconic nature of the gestures Tomasello and colleagues make the important point that iconicity per se may not pay a role in the children's performance in this case, because the iconicity pertained to the *use* of the objects. Thus the movements intended to stand for e.g. a comb were also the ones that a child has extensive experience of when encountering combs. Combing movements happen to occur together with combs. On the same note, when a child encounters a comb the child might make a combing motion in its hair without signifying the comb with those movements, but signalling recognition of what it is used for. The iconicity is then only in the eye of the onlooker.[103] Some pretence can be analysed in the same way, i.e. that the movements that are made with an object does not need to signify anything else but the habitual movements connected to them.

In a picture in Patterson and Linden (1981) Koko signs "toothbrush" towards a toy banana and then uses it to brush her teeth. In this particular case her actions are interpreted as a case of mental transformation in play, i.e. pretence. It is a strong case since plastic bananas are not toothbrushes. Koko is claimed to have shown several substitutions like this. Such behaviours do not require a referential view of the object used, but it is an act of reference in that a second object is intended. However, in order to fully understand this behaviour we also need to know if Koko calls many other (oblong) things toothbrushes, if she can pretend the banana to be other things, like a telephone, and so forth.[104] In other words, Koko does not need to see that it is a plastic banana that she uses for a toothbrush. All she needs to see is that the object looks somewhat like a toothbrush and can be therefore used as one. (For this argument it would have been more convenient if Koko had chosen something smaller than a banana, but we cannot assume that she did not see something toothbrush-like in it. Apparently she did!) Using one well known object to stand for a functionally different one is judged to be difficult for young children because they cannot inhibit their sensory-motor schemas for the first object (Tomasello et al., 1999). The most convincing demonstration would thus have been if Koko first had named the banana "banana", then pretended to eat it, followed by naming it "tooth-brush" and pretend to brush her teeth with it.

---

[103] An account of the development of primate mimetic cognition can be found in Zlatev et al. (2005).

[104] The assumption that Koko is a frequent pretender (Matevia et al., 2002) would probably mean that if Koko was to bite a food pictures it would be interpreted as pretence. This has not been reported for the mature Koko, but as an infant she did bite pictures (Patterson & Linden, 1981).

It seems likely that several of the apes above do indeed engage in pretence and imagination, but this does not equal seeing toys as representational. That one pretends to feed a doll does not necessarily entail that one *also* pretends that the doll is more than a lifeless object with a mouth. Until more suggestive evidence than sticking real or imagined food into the mouth of a doll, or forming doll hands into a limited number of signs, is published, no definite conclusions can be drawn.

Even though they use the same words for real instances, pictorial, and replica versions of a particular entity we cannot assume at face value that the ape mean the same thing when it names a real cat and a plastic cat a "cat." A stuffed toy dog has more in common with other stuffed toy dogs than to real dogs. Which is the actual referent when saying that the toy dog represents a "dog" for the ape? One word can pertain to several separate categories.

Koko has the word "fake" in her vocabulary, which she sometimes uses towards toys (Matevia et al., 2002), hinting at the fact that she indeed sees e.g. cats as a single category, but that there are "fake" ones. Besides language, iconicity is probably a powerful factor when it comes to bridging two categories. It is worth emphasising that the particular language trained individuals described above seem to be able to instantaneously parse and recognise the features of a doll, and perhaps other toys as well. These apes need not *learn* to find head, extremities, eyes and the like on novel dolls, be they dogs, gorillas, dinosaurs, or babies. If they can do this with toys, and they see toys as representing a category that extends beyond toys, it seems likely that they are able to repeat this feat with other iconic media that entails abstraction, such as drawings.

9.2 Models and maps

In section 2.2 it was mentioned that human children could not readily use a model room as a representation of a real room until the age of 3 years, but that when hindered to interact with the model, or when it was put behind window glass, the task became easier. Dual representation was blamed. The model is both an object in itself and simultaneously stands for a different space. This finding was further corroborated by using photographs instead of models, which allowed children as young as 2.5 years to successfully retrieve hidden objects. Furthermore, when the children were hindered from doing perseveration errors by allowing only one trial per room, even 2-year olds could find an object that was indicated on a photograph or on a video screen. However, it was also mentioned that rather than poor inhibitory control the cause for perseveration errors might be a poor conceptualisation of the relationship between the model, or photograph, and the room referred to. That is, perhaps there is a failure to recognise that the real room and the model are being updated irrespectively of each other, and in different ways. The situation in the real room is updated by a person hiding the toy in a novel place, while the model or photograph is updated by a new narrative context, or by novel actions on its miniature elements. These actions must every time be seen as corresponding to the actions in the room that is out of view. The relationship between the model and the real room must be retained as well as updated, between trials.

When chimpanzees are shown a familiar room or space on television in which a caretaker hides a reward they will not spontaneously use this information when later searching for the item (Premack & Premack, 1983). It was hypothised that it was the lack of the third dimension that hindered chimpanzees from making the connection; hence the model room paradigm was created. None of the juveniles in Premack's laboratory was able to use a miniature room with miniature furniture to find a reward in a large room with large but otherwise identical furniture. However, when using a one-to-one scale, by constructing two identical rooms, the subjects found the hidden items instantaneously.[105]

The problem in this setup is to make sure that the chimpanzees are aware of the fact that there are indeed two rooms. Even then, the ape does not need to know that one room signifies the other, only that the same event tends to occur in booth rooms. This is not a large mental leap if an overall equivalence has been established. Premack and colleagues then successively reduced the size of the furniture in one of the rooms and, with initial drops in performance, the subjects soon performed as well as with two rooms identical in size. However, transfer only occurred for two out of four subjects when the model, or map as it had by then been distilled to, was moved to a different testing location. In addition, none of the subjects could retain performance when the map was rotated. When a novel map of a different room was introduced, the apes also performed without using information from the map. The conclusion will thus have to be that the iconicity of the map, or model, was bypassed in favour of other associations between the two spaces. This end result does not exclude, however, that iconicity might have played a part somewhere along the way.

The failures of Premack and Premack (1983) have later been contested using Premack's own subject, Sarah, as well as other chimpanzees, then at the Ohio State University Chimpanzee Center. In a 1:7 scale model of a familiar room Kuhlmeier et al. (1999) hid a miniature can of soda. An adult female, Sheba, succeeded in finding the real soda can in the larger room, while an adolescent male, Bobby, did not. As a pre-test training phase the chimpanzees had received intimate experience with the testing material. The model had for example been placed within the larger room and the miniature furniture had been placed beside their larger counterparts. The chimpanzees had also been allowed to participate in a training hiding event where the miniature can was placed in the model and the real can in the real room, in full view of the subject. This was all in accordance with the experience that human children in DeLoache (1987) had received.

On the first trial where the can was hidden in another place than the one used for familiarisation, and the subjects were only allowed to see the event in the model, they both fetched the real can directly at the correct spot when allowed into the larger room. However, during subsequent testing only the older animal, Sheba, successfully retrieved the can from the first place where she looked. (Seven out of eight

---

[105] Human children find low size ratios between spaces easier than bigger ones in object-retrieval tasks, but identical spaces thwarts performance severely, arguably because it creates conflicting memories for what took place where and when (DeLoache & Sharon, 2005). Also, without a clear difference the direction of a *stand-for* relation is opaque.

correct searches.) With the limited training offered it seems unlikely that Sheba had learned the corresponding places without attending to physical similarities, but this cannot be ruled out without specific controls, especially since there were only four hiding places. Introducing a set of novel furniture would be such a control. Sheba also performed above chance when only the correct miniature furniture was shown independently from the rest of the model, so just rotating the model, as in Premack and Premack (1983), might not have posed her any problems.

An interesting finding was that while Bobby remembered where the miniature can was hidden in the model after unsuccessfully looking for it in the real room, Sheba did not find the hidden miniature can in the model at a level above chance in the first place where she looked. Might this indicate that the two spaces were not seen as being updated independently from each other? Could Sheba in fact have believed that her actions of taking the real soda might have removed also the miniature one? Or perhaps she only distrusted the experimenter who could have tempered with the model while Sheba was in the room looking for the can. Bobby's failure in the real room was attributed to strong learning effects from previous trials. He always started out his search in a specific place and continued in a set pattern.

Sheba's performance seemed to transfer to individual colour photographs of the four pieces of furniture as well as to a photograph of the entire room (Kuhlmeier et al., 1999). However, she was only correct on about half of the trials for the individual furniture photographs. For this reason trial one data (all places were run twice) are needed to completely rule out learning and choices by exclusion. On the photograph of the complete room Sheba was correct on six out of eight trials, which implies, even if one would invoke learning, that she parsed the photograph into four relevant parts that corresponded to the four places in the real room. To accomplish this with a high success rate without perceiving the photograph - room correspondence seems unlikely.

An important difference from the previous setup was that the experimenter pointed to the photograph of the correct hiding place before she went into the room to hide the soda can rather than putting a miniature can there. That Sheba was successful either means that she understood the role of the point, which is interesting in itself (see Tomasello et al., 1997), or that the role of the iconicity of the miniature soda can itself was superfluous. This could have been the case also in previous conditions. Marking the spot with an arbitrary object or pointing might have sufficed.

Bobby did not perform better with photographs than with the model. If compared to 2.5-year old human children it was thus more than just dual representation that posed him difficulties. Recall that young children performed better with photographs than with models, in accordance with the idea that seductive object properties of the medium might obscure the message. However, it is not necessarily true for all subjects that one reduces confounding object cues by switching from models to photographs. If photographs were viewed in reality mode there would still be a need to counteract prominence effects, as it would imply one object standing for another object. It might work with 2-year-old human children just because they know more about pictures than did Bobby, in order to be able to differentiate it sufficiently from the real world. That said, the case can also have been the reverse.

Bobby might have had difficulties appreciating the similarities between photographs and real furniture.

To compare Sheba and Bobby to other chimpanzees, five additional subjects, among them Sarah, were tested (Kuhlmeier et al., 1999). This time a model with the scale 1:7 of their outdoor enclosure was used. Homemade miniatures represented four large familiar play items where a bottle of juice could be hidden. The spatial relations between objects in the two spaces were identical. A familiarisation phase similar to the one in the first experiment was used where the chimpanzees could view the paired objects together, as well as one sample hiding event in full view.

This time each hiding site was used five times and the potential for learning the connections between sites by rote learning was therefore even greater than before. But since the subjects were allowed to make exhaustive searches and the fruit drink always was acquired in the end it was not overly costly to use a search strategy that was *not* informed by the model. In other words, there was no real need to learn what was not spontaneously obvious. Three subjects, among them Sheba and Sarah, performed above chance in this task, looking in the indicated place first in about half of the trials. All the unsuccessful subjects were looking for the juice bottle in a more or less rigid search pattern. Making the search more costly might have reduced the reliance on this strategy. Interestingly, one of the subjects had a very high success rate if her first visited spot was excluded from the analysis, i.e. she chose the correct spot among the remaining three at her second try. This illustrates how persistent perseverance errors can be.

To somewhat disrupt set search patterns the experiment was rerun, but this time the spatial layout of the objects was shifted between trials (Kuhlmeier et al., 1999). Sheba and Sarah were still correct.[106] However, now the third previously successful subject fell short. In his place the female who was correct only on her second choices now performed above chance on her first. This manipulation can be seen as equal to the one of Suddendorf (2003) who found that when using several rooms instead of several trials, to counteract perseveration effects, the age at which human children could succeed in an object-retrieval task was pushed back to 2 years.

The call for costly search patterns, as well as transfer trials, was recognised by Kuhlmeier and Boysen (2001). The seven chimpanzees in the above study were re-tested in a procedure where they were only allowed to search one location at the time, then returning to the model for a correction trial. In addition, a novel set of miniatures and real objects were introduced in a limited number of transfer trials to control for learning. No correction trials were allowed for these. With the new procedure all subjects now performed above chance after training trials. Six of the chimpanzees performed above chance in transfer trials. Confusingly, the chimpanzee that failed was the third successful female in Kuhlmeier et al. (1999). For a second set of novel items she succeeded though, but this time another subject had reverted to chance search patterns. This was attributed to frustration with changes in the setup.

---

[106] Sarah finally got her revenge on Premack and Premack (1983), where she failed with rotated models.

The results strongly suggest that the formerly unsuccessful chimpanzees had finally either started to pay attention to the model hiding event, or had started using this information in their search strategies. In either case it is clear that information gleaned from the model affected their search behaviour. That something had clicked was evident in that only one chimpanzee reverted to an inefficient strategy when the old procedure was reinstated. Solving these tasks without applying a *stand-for* relationship, that furthermore hinges on iconicity, is improbable given the transfer to novel replicas. However, although no learning effect was found between early and late trials, without strict trial one data non-iconic strategies that bypass recognition cannot be entirely ruled out.

An issue is for example whether the objects in their totality, on a categorical level, were perceived as iconic, or only parts of them (e.g. colour). Rules such as "when the pink object is indicated look at the pink location, i.e. the plastic slide" cannot be controlled in the above setup. Kuhlmeier and Boysen (2002) turned this interjection into a new experiment. Besides colour and shape cues they also controlled for positional cues.

In the first experiment three conditions were tested. One in which the spatial layout of the objects differed between the two spaces but colour and shape cues were present, one in which the spatial arrangement was constant and shapes corresponded but colours differed on respective items, and lastly one in which shape was the incongruent variable between the two spaces. The two chimpanzees who failed in the last part of Kuhlmeier and Boysen (2001) also failed in this test. The other five succeeded above chance in all three conditions. Although only moderately successful, no learning effects could be shown across the eight trials. This suggests that the miniatures were perceived in a dynamic fashion, where all three variables, i.e. colour, shape, and position, were informative when involved in all three possible constellations, and the subjects had no difficulties switching between these. Although an impressive performance this might sound off an alarm bell. Is this really congruent with viewing the miniature objects as smaller versions of the larger objects? When effortlessly switching to a condition where shape is redundant, for example, and only position and colour corresponds to the larger space, does this mean that the miniature objects had never been perceived as replicas, but only as conglomerates of cues?

When using four identical hiding locations, and thus test positional cues in isolation, only two subjects performed above chance.[107] Similarly in Boysen and Kuhlmeier (2002), although seemingly using the old method that also fostered perseveration errors, individual subjects could not reliably solve the task when colour and shape information were removed. Colour and shape were unfortunately not tested in isolation in either study. Although the spatial layout can be said to form an iconic relationship between the model as a whole and the larger space, pictorial queries would benefit the most from a condition where shape is tested independently from both colour and position. In addition perhaps relative size between objects should also be controlled.

The important lessons from Kuhlmeier and Boysen's research are at least two. Firstly, some individuals can apprehend a correspondence between two spaces that

---

[107] The adult males Darrell and Kermit.

are similar but not identical in terms of visual features. Secondly, apes need a good reason for utilising such a correspondence. This "reason" needs to be discovered by the apes themselves and is dependent on the requirements of the task. In the way the subjects in Kuhlmeier and Boysen's studies read the situation, as competent and logical problem solvers working towards a goal, the informative value of the model was often superfluous. When the information got salient *from the point of view of the subjects* it was integrated into their problem solving. An iconic competence emerged.[108] If abilities are seen as being assembled in the task at hand, rather than residing inside a head, iconic abilities can emerge independently in several types of endeavours, such as gestural ones (e.g. Tanner & Byrne, 1996), the use of models, pictures, etc. When it comes to pictures, transition into a pictorial mode can be such an emergence.

A "general" ability can rather be seen as a result of learning many instances where a similar strategy is useful. In an object-retrieval task, for example, why should we expect a chimpanzee to immediately understand more models than the one it is trained on? If the chimpanzee cannot cope with a novel or changed model, does this disqualify the ability as a "true" ability to understand models? If a gorilla can understand some pictures, in a proper referential manner, but not others, does it have a pictorial competence or not? Abilities are often not black-and-white, but we tend to talk about them as if they were. Anthropocentrism, as well as anthropomorphism, have given roles in this problem, but I would like to repeat the quote that introduced Part II of this thesis: "The program we have described avoids the question of whether an animal other than man can acquire language. As comparative psychologists we must reject this question. It is like the question of whether an animal other than man can have thoughts. It depends on the definition of language rather than on the observations of what animals do" (Gardner & Gardner, 1971, p. 181).

9.3 Video

When the chimpanzee Gua and her human foster brother Donald are roughly a year old they are shown a film of themselves, projected as a small image of only 15 cm in height. They watch intently for several minutes. A couple of months later they are shown a second movie, this time of a recorded testing session. This time the projection is about half the size of the real event. When a piece of apple is shown in the film Gua climbs up on a desk to the projection. She touches the head and face of the image of Donald. Then she touches the apple with her lips. "She seems indeed to be trying to pick up the apple" (Kellogg & Kellogg, 1933/1967, p. 92).

Video has often been used by zoo and laboratory personnel to enrich the long hours of their captive charges (e.g. Bloomsmith & Lambeth, 2000), and it seems that some enjoy watching television and even develop individual preferences for movie types. But it also seems that many apes pay a most fleeting interest in such enrichment,

---

[108] This is perhaps how the process would have been described in the framework of Rumbaugh and Washburn's (e.g. 2003) "rational behaviourism." Köhler might have called the process insight.

even when raised in human homes (Savage-Rumbaugh & Rubert, 1986). This lack of interest can be all-inclusive, encompassing stimuli that would otherwise be considered highly interesting, such as movies of conspecifics (e.g. Temerlin, 1976). "Given the apes' lack of overt behavioral response to video images, it is difficult to arrive at any firm conclusion about how they process moving images. Do they interpret them as representations of reality? Do they think that there are Lilliputians inside the box doing things with one another, do they think that what they see is really happening – but only in a different location, or do they not even 'think' about any of these things at all?" (Savage-Rumbaugh & Rubert, 1986, p. 300).

Showing movies has not only been used as entertainment, but also to familiarise animals to future events or to elicit interest in sometimes uncommon behaviours, such as reproduction (e.g. Maple & Hoff, 1982). Whether this is a successful method awaits investigation, but a measurement of brain temperature in chimpanzees that viewed video clips of play, scenery, and aggression, recorded a significant increase in temperature in the right hemisphere when the most negative movies were shown. This presumably indicated an emotional effect of the video clips (Parr & Hopkins, 2000), and would suggest that apes can appreciate movies not only because they display colourful moving patterns, but actual content. However, the movies in this study were not muted. Sound recognition can stand for some, or all, of the effect.

The study above is an example from experiments where video is used in order to present a stimulus that is intended to be as real as possible. It can be seen as an extension of the typical photograph experiment. The added information of movement seems indeed to be a highly salient cue for eliciting some level of recognition. D'eath (1998; Cabe, 1980) report studies where even lizard and fish react to video presentations of conspecifics, and jumping spiders react to filmed prey.

However, it is important to make sure at what level recognition of video stimuli occurs. The first reason to believe that a stimulus is not always perceived as intended by the researcher is that video monitors are designed for human vision. Species differ in their colour vision, depth perception, visual acuity and the threshold at which a series of still images are perceived as a continuous flow (D'eath, 1998). If generalisation from video stimuli to the real visual world is to be made, it is pivotal that the video stimulus adequately captures what it is intended to capture. Luckily, old-world monkeys and apes seem to have a comparable vision to humans. This means that when they watch movies they can be expected to recognise what is going on in the scenery before them just as well as humans are, although they might need familiarisation and training with the apparatus etc.

Clear cases which can be attributed to reality mode views of video are such studies where e.g. macaque monkeys in Capitanio et al. (1985; Plimpton et al., 1981, both in Bloomsmith & Lambeth, 2000) responded socially to videos of conspecifics. Or studies where monkeys have shown sensitivity to threat signals, can recognise sex, or learn to avoid snakes by seeing fear responses towards snakes in video uptakes of others (see D'eath, 1998; Bovet & Vauclair, 2000). It is clear that the subjects recognise the content of the televised images in these studies, and that they tend to react to them as to real events.

But movies are not by default "more real" than other stimuli. Sometimes video stimuli fail, and there are reasons to believe that this is not always due to visual or technical issues but to the nature of the content itself. A peculiarity with video is that although it looks so real, it is not interactive. That a social stimulus does not react to your actions towards it is a great give away that seem to have hindered performance in several studies (D'Eath, 1998). That this incongruity has a frustrating effect and breaks down performance, and perhaps even recognition, strongly hints at a reality based mode of viewing the video content, but also that with added naturalness comes added expectations. With an appreciation of reality come expectations that are coupled with that reality. If these are violated, frustration, confusion, ignorance, and so forth can result.

This idea is perhaps supported by the findings of Washburn et al. (1997), who, in a study on the reinforcing value of being shown video clips in a joystick computer task, found that rhesus macaques preferred to reward themselves by viewing blank video rather than video containing face stimuli. When they did chose to view social clips they chose footage of themselves rather than known or unfamiliar conspecifics. This was probably not because they were particularly self interested, but because they had extensive experience with reflective surfaces they were used to the unimposing view of their own face. Social stimuli, on the other hand, can be arousing and not be felt as rewarding at all, especially if it acts out of the ordinary.

The assumption that there is symmetric equivalence between real world events and events on a television monitor is sometimes well founded, but can also result in premature conclusions. Cacchione and Krist (2004) tested 10 experimentally naïve chimpanzees on a video task and measured their looking time for various possible and impossible events. The subjects let some impossible object physics pass while they reacted to others. They were sensitive to the amount of support a banana needed in order not to fall over an edge, but they did not react when an apple was resting on a vertical surface. In this case the subjects were all naïve to video tasks. One of them even had to be excluded because he did not at all pay attention to the screen. Had more experienced subjects been used the equivalence assumption might have been more problematic. One would not be able to say with certainty whether the impossibility of the apple was accepted because it transpired on a television monitor, or if it was allowed on film because it would have passed in real life. Did the result say something about chimpanzees' expectations on fruits and their substrates, or chimpanzees' attitude to what can transpire on television monitors? There are potentially all sorts of reasons *besides* physics and real-life expectancies, to why some events draw more attention than others, especially if video events are seen as very different from real events.

The unfolding of events captured in video can both help recognition, by adding e.g. movement cues, but can also confuse since one has no control over the perspectives taken. Especially in edited material it might be difficult to follow the coherence in the "visual narrative."

For four adult chimpanzees tested at the Kyoto Primate Research Institute only one, Ai, could intuitively match video clips to video clips in a computerised MTS task at first presentation (Morimura & Matsuzawa, 2001). Chloe (from the picture experiments) came close second. All chimpanzees were experienced matchers, but only Ai had participated in tests using video stimuli. The remaining two subjects needed up to 98 trials before they could match according to criterion. The clips used were continuous footage containing minimal compositional changes. If further film-specific transformations would be included, such as sudden changes in perspective, understanding the coherence of a depicted event might become even more difficult. But this would only be a true problem if the content of the movie was indeed perceived as a continuous, unfolding event. In the present case the matching clip was identical to the sample one, so the actual "narrative" of the video event could in theory have been bypassed and still allow successful matching, based on some sense of overall recognition, or recognition of a particular detail.

Morimura and Matsuzawa (2001) also let the subjects match photographs to video clips, similar to Premack and Woodruff (1978a; 1978b) described in section 7.7. But since the photographs were freeze frames from the videos themselves the critique that was sometimes directed towards Premack and Woodruff is possible to apply also to Morimura and Matsuzawa. That is, that successful matching still does not entail that the subjects comprehended the events on the movie screen as a coherent event. However, when a movie clip had included a sudden compositional change, i.e. a cut to a new scene, the ability to match significantly decreased if the matching freeze frame had been taken from the early part of the movie, prior to the cut. A clear recency effect could be shown. But when the clips were continuous no such effect was found. This means that the clips, when divided by scene changes, were stored as a list of discrete items, while continuous movies were not. This suggests that movies indeed are seen as continuous events, but that cuts can disrupt the flow of such "narratives." A future experiment with subjects that are used to watching heavily edited movies, deep in story, would be most interesting. Will they be able to bridge scene changes and see the movie as one event, or do they watch such movies as a collection of many disjoint events? Understanding the conventional use of camera work would be greatly helped by viewing a movie in a pictorial mode rather than a reality mode.

The young chimpanzee Viki enjoyed going to the drive-in cinema with Keith and Catherine Hayes, although her attention frequently wandered from the screen, and once commented "cup," one of her very few words, when a toast was made in a movie. She thus indicated that she recognised onscreen events, or at least cups (Hayes, 1951). Interestingly, she especially enjoyed watching colour cartoons, which has a great potential to require a pictorial mode in order to at all evoke any interest. However, since she not even said "cup" in response to these we do not know what she actually saw in them. Other language-competent apes, such as the gorilla Koko or bonobo Kanzi, also can comment events that happen on screen but there are no published descriptions of doing this for cartoons that they see for the first time. If apes are able to decode, in whatever mode, static cartoon material, adding movement to the images probably further helps the recognition process, in line with the idea of wholes and parts affecting each other. It can be because of this that Sherman or Austin commented "carrot" on their lexigram board when they saw a Bugs Bunny cartoon (Savage-Rumbaugh, pers. com.). A static picture of the very same carrot might have been seen as something completely different, or just a blotch of two-dimensional orange colour.

9.3.1 Live television

Premack and Premack (1983) note that it is only with great difficulty an ape learns to use video information to look for objects. The cues that greatly help in this task can be explained by the television-as-window illusion. As expected from a reality-mode perspective, cues that strengthen the illusion of television as a window seems to enhance performance. The abstract to Menzel et al. (1978) reads: “Infant chimpanzees, after watching a small black-and-white closed-circuit television picture of a familiar caretaker walking out into an outdoor field and disappearing from sight, were more successful in finding the person than if they had been given no such cue; and their performance approximated that which obtained after they had seen the same scene normally, via direct perception.” The closed-circuit video image gave live feedback that corresponded to the hider leaving the room, appearing on screen and hiding. Importantly, the television screen had been arranged so as to be in line with the view that the chimpanzees would have had if they had been watching the events through a window (Menzel et al. 1978, in Poss & Rochat, 2003).

In a different search task Poss & Rochat (2003) hid rewards in one of three distinctive containers, either in view of the subjects or on live monitor broadcast. Chimpanzees, orangutans, and human children (2-3 years) participated in the study. It was found that apes were better than children when hiding took place on the television screen. Since filming took place behind a curtain directly behind the monitor there was no control for the window illusion, other than a familiarisation phase with the monitor and watching a live filming of oneself. This does not necessarily exclude a “magical window” attitude towards the monitor. Since the children had previous experience with television this probably explains why they were outperformed by the apes. Their experience of television as differentiated from what takes place in the actual room hindered children from making a connection between the monitor image and the hiding event which they knew took place behind the curtain. Cues that highlighted the hiding event, both visual and verbal, helped 3-year olds to better their performance. Simultaneously hearing the voice behind the curtain and seeing the actions commented taking place on television probably helped them to attend to the television – reality connection. If they had instead been made to believe that the television is a window of sorts, they might likewise have solved the task.

Troseth et al. (2006) found that 2-year old children did not follow instructions in a search task when they received them through a monitor, but when given the exact same information in face to face communication they succeeded to find the hidden item. However, if the children were first allowed to interact with the person on the video through a live broadcast they also started to pay attention to the instructions. Troseth and colleagues mention that while children younger than a year old seem to glean social information from people on video, this “ability” seem to disappear with maturity. I believe that this effect is due to the breaking down of a reality mode processing due to differentiating experience. Or rather, a third, televised world is created. With no recognition of a connection to the world outside of the screen children appear caught in-between reality mode and a full pictorial mode. But with help they can switch back to a reality mode view of video and solve certain tasks. Poss & Rochat (2003) mention the television-as-window illusion, but it is a mystery why they do not control it, since they call their study an assessment of seeing video images as referential.

The televised hiding-of-food method has potential for settling the issue of whether video images can be seen as pictures or not, but certain measures has to be taken to minimalise the illusion of a window. With this comes the problem of ruling out the "magical" aspect as well. It is impossible to say to what extent magical thinking can be stretched. If the familiar caregiver that does the hiding also is in view when the video is seen, does the informative function of a recording break down or is "being in two places at once" incorporated into the possibilities of the video world?

Can viewing oneself on video upset the belief that video is a form of window? Successful attempts have been made to allow apes (Menzel et al., 1985) to guide their hands through an opening and investigate the backside of an occluder by watching closed-circuit video. The two chimpanzees tested, Sherman and Austin, not only used the broadcasted video to direct their search in transfer trials, but showed a flexible adjustment to novel orientations of the screen, whether reversed, inverted, or both. They also readily made a difference between live and pre-recorded video, although the two chimpanzees did stick their hands through the openings to see if they appeared on the screen when there was, according to the video, already a hand there. They did not seem to expect to find another hand in the hole though. Testing the image like this might suggest that they were not entirely clear about the nature of the connection between the filmed events and reality. They might have believed that their actions were visible onscreen but somehow out of tune with their movements. Or they just wanted to test if the monitor would display their reaching after all, since the target was visible onscreen.

The development of Sherman and Austin's understanding of television is well documented in Savage-Rumbaugh and Rubert (1986). The initial interest of Sherman and Austin towards commercial television was very weak. They also paid little attention to closed-circuit video of events that took place in other places of their facilities. Alike attitudes have been found for other chimpanzees towards such imagery.

Sherman an Austin were then involved in social viewing of films of a better quality and in colour, mainly depicting other chimpanzees. They were encouraged to pay attention to the screen by the trainers who vocalised when interesting events occurred. In the beginning the chimpanzees' attention span was short, but it steadily increased with time until they could watch a whole movie of approximately 30 minutes. At this time they also began to show behavioural responses towards the content of the screen, such as "[…] display when they saw other males begin to display. They attempted to inspect female swellings and attempted to bite the screen when particular chimpanzees appeared" (Savage-Rumbaugh & Rubert, 1986, p. 305). By now they showed great enjoyment in viewing chimpanzee movies and could differentiate between old and new ones. They were also able to label objects with their lexigrams that were displayed on a small colour monitor.

When closed-circuit television was reintroduced Sherman and Austin this time spontaneously paid great interest in events in adjacent rooms. On the first occasion in which one of the chimpanzees was allowed to join the trainer in another room and the remaining chimpanzee followed this interaction on the monitor, the viewing chimpanzee anticipated the first one's re-entry in the room by its first signs of moving towards the door on the screen. The viewer had turned his attention to the door rather than the monitor. In subsequent testing the two chimpanzees clearly showed that they knew that they could move into the world seen onscreen, act upon it, and then return. They became very proficient in finding food hidden on closed-circuit video in this way. After the experiences with live video Sherman and Austin also became interested in several types of commercial television programmes. Not until now could television be used as enrichment when the apes were left on their own.

In a description of a typical afternoon session with Sherman and Austin when 7 and 6 years old respectively, they request to watch television. When watching movies of other chimpanzees they display towards these. In what words should such behaviours be described? Are they comments, social or private? Is it play and pretence involved? Is it a learned behaviour, a habit? Do they confuse video with reality? Given that it is a spontaneous reaction, "acted out involvement" is perhaps the best description. But why do they get involved with such stimuli? Do they act out in the same way towards a photograph of a chimpanzee? Towards a chimpanzee doll? Towards a lexigram that designates another chimpanzee? The element of the real is certainly a factor in such involvement. It is not a mental phenomenon working from the inside out, but a response on outside cues.

That said, excitement evoked by a stimulus does not have to pertain to the stimulus as an object but to the associations started by the stimulus. For example, when happening upon some photographs of the gorilla Koko, Sherman got very excited. Hair stood on end as he tapped one of the pictures with his finger. He continued to point to the photographs and ran excitedly around with them. Unusual for Sherman he then brought the photographs with him when he and his trainer went outdoors. He continued to repeatedly point at them. He steered the trainer towards the gorilla quarters which he had been allowed to visit on previous outings, "an event which both scared and thrilled him" (Savage-Rumbaugh & Rubert, 1986, p. 282). The trainer understood what Sherman wanted and said "no" and steered in another direction. When Sherman's request to go and see the gorillas was not granted he lost interest in the pictures, as well as losing his excitement, and even will to be outside. The photographs of Koko were dropped to the ground and ignored. It seems that although Sherman was very excited about gorillas, and his response to the photographs could be attributed to this fact, he had been more excited about the whole idea of going to see the gorillas than about the photographic expression as such.

Similarly, the anticipation of arriving food can be as exciting as the sight of food itself. Such behaviours can confound when trying to pinpoint true confusion behaviours. Nevertheless, properties of the real still play a crucial role. A lexigram would on its own hardly evoke excited anticipation if it was not coupled with a promise of fulfilment, either by e.g. being pasted on a commonly baited food box or being part of a discourse with a trustworthy caretaker. The evoked mental images in such contexts include properties of the real. The word "cake" does not get your mouth wet, but adding a few visualisations the concept "cake" might. A photograph of a cake, in a fully differentiated mode, would also not whet your appetite, but adding a few associations it might. In a less differentiated mode it might indeed make your mouth water, and in a full reality mode you might even bite into it. The whole concept of gorillas flooded Sherman when he found those photographs in a context where he was about to go for an outing. Happening upon a lexigram, or a stuffed toy gorilla, might not have had the same effect.

9.3.2 Self-recognition in video

Austin was the first to discover that he could view himself on a closed-circuit monitor. This came very suddenly during one particular filming session and ensued in 20 minutes of experimentation with body postures, facial expressions and ways of manipulating his food in front of the screen. A favourite activity that Austin later developed was to combine his pretence and self-recognition by studying himself onscreen when munching imagined food. An arduous work commenced trying to interest Sherman in his own image as well, but to no avail. However, several months later he suddenly recognised himself on the monitor and reacted in the same way as Austin had. They both became proficient in determining if they were viewing a live or taped video of themselves by testing the image by sticking out their tongues etc. Such testing never occurred when they saw other individuals onscreen. Later they also showed that they understood when a scene was a live broadcast or taped when it depicted someone else in another room. They attained this by comparing the monitor image to real world feedback, such as sounds from other parts of the building. If they detected such a correspondence extra interest was paid to the scene.

Sherman and Austin might also have grasped something about the role of the camera in the process as they often tried to film themselves when there was no monitor present by moving between the lens and the viewfinder. Also, to get a better view down his throat, Austin would adjust himself in relation to the camera rather than the monitor. Presumably this connection had been learned exactly by this type of exploration, or by being shown footage that just previously had been filmed. The gorilla Koko shows a similar interest in still and video cameras and is said to know when she is being filmed, which affects her behaviour (Schroeder, 1978/2006).

Even though Austin and Sherman received much experience with filmed events, and seemed to have understood a great deal about it, the screen never lost its magic. In experiments where the two chimpanzees had to pay attention to both the screen and events in the real world the latter was often compromised because their attention was transfixed on the monitor. When it came to reading the monitor, though, they excelled. For example when communicating about food in an adjacent room visible onscreen, they hardly ever gestured towards the screen but to their lexigram board, the projected lexigrams, or to the room that contained the food.

Even though Sherman and Austin excelled at using video to solve tasks in the real world, a "magical window" explanation is difficult to rule out. They furthermore seem to have found out the role of the camera in this magic. In addition, they found out that video seem to not only work as a “magical window,” but also as a “magical mirror.” One can just as well say that Sherman and Austin found out how video works, at least closed-circuit video. But how pictorial is closed-circuit video? There need not be anything “magical” about it at all since there is no need neither for differentiation nor reference when the feedback between the world, or body, and televised image is instant. Their comprehension of recorded video is unclear. It is only reported that they easily differentiated between live and recorded imagery, and that they were more interested in the former. Exciting work remains to be done in this field.

Hirata (2007) studied self-recognition on televised media in chimpanzees. Out of the ten Kyoto chimpanzees observed only two displayed self-exploratory behaviours while watching their own image on live video. One of these, Chloe again, also manipulated objects with a clear focus on viewing these actions in the monitor. It thus seems that more experience than the one provided by a single experiment is needed to foster an interest in video guided exploration in most subjects. This support the findings with Sherman and Austin who required quite some experience before they started to use video in this manner. Perhaps Hirata’s or Menzel et al.’s study would be even more interesting if live video would have been contrasted with delayed video. When the instantaneous feedback from the video is removed reality mode processing is potentially compromised.[109] In order to make sense of the video feedback and use the image for exploration one must consider the image to be separated from one’s immediate actions, but still depict them. However, this would not only be a pictorial challenge, it would also tax one’s attention to the line of events that one is currently involved in. One must both anticipate where one’s change in movement will take one’s limbs in the spatial layout of the screen, and then update this anticipation with the actual video feedback. Only attending to either part of the task will result in aimless movement.

The depicting nature of delayed video can for example be tested as in Povinelli and Simon (1998). When they are viewing a video recording of events that took place three minutes earlier, 4-year-old human children, but less than half of 3-year-olds, have been found to reach for a sticker that had been covertly placed on their heads during filming of the event in question.

Law and Lock (1994) tested the spontaneous reaction of four gorillas towards their live video, as well as video of themselves recorded at an earlier occasion, and video of unfamiliar gorillas. The gorillas did not seem to show social behaviours towards video images of other gorillas with the exception of one possible social gesture. One response noted during the delayed playback condition seemed to indicate that the recorded video was mistaken for live video. The viewing gorilla turned around and looked behind himself when he saw himself on screen approaching the spot where he was currently sitting. He could either have made sure that there was no gorilla approaching him from behind, or he might suddenly have recognised his enclosure in the video and looked around to check something. In either case it suggests

---

[109] It remains an empirical question.

that he perceived the monitor more as a mirror than a window of sorts, but the evidence is inconclusive.[110]

The response to live video feedback was markedly different from the previous two conditions. Besides avid attention several gorillas showed signs of image testing and face exploration. Also stretching for objects without taking the eyes of the video screen was evident. The image testing gave the impression of exploring the "strange" match between one's own and the video image's movements. Facial exploration consisted either of looking into one's mouth, a potential that was discovered seemingly by accident during yawning, or feeling along one's brow ridge which is an area that one normally do not have visual access to.[111]

When watching a chimpanzee movie, Premack's Sarah, suddenly after 30 minutes of calm, gets very excited by the capturing of a young chimpanzee in a net. Sarah hoots and throws paper towards the screen "seemingly aimed at the animal's captors" (Premack, 1976, p. 346). When the trainer who was present touched the captured animal onscreen "Sarah shuddered and turned a wildly startled face to the trainer." Rightfully, Premack evokes the problem of knowing whether the animal confuses reality and image at such occasions. But he concludes that rather than confusion, what is displayed is similar to when humans sob when watching a sad movie or caresses the photograph of a loved one. He furthermore suggests that one should view the depth of such participation as a measure of intelligence. Only an intelligent species can treat a scene as if it was real, and still not show all the behaviours that would accompany a real scene. Sarah did for example not flee the room although the scene seemed scary. Outward signs of participation are indications of internal participation, he maintains. This is of course true for confusion events as well and Premack adds that an individual must also show that it can use pictures referentially, e.g. in communication, before confusion can be ruled out.

Premack might be right, but I think that the specific case that he describes is still a case of seeing television as a *kind* of reality. Sarah for example throws objects at the screen, which contradicts Premack's suggestion that Sarah inhibits her reality based responses (all she is *not* doing is leaving the room). He also implies that Sarah knows that the television image cannot harm her, but still she reacts strongly when the caretaker touches the screen. The scene on television is thus seen to be somewhat differentiated from reality in that, while Sarah is very upset, she does not act towards it in an improper way, such as charging the screen. She rather takes into consideration that a television is a television and only throws paper at it. Perhaps what is lacking is the referential knowledge that the scene is far removed in time and space. However, Premack would probably maintain that Sarah would continue to respond in the manner she did because her reason for acting was never confusion but involvement. Furthermore, her inability, or lack of motivation, to inhibit her responses would be the same no matter what.

---

[110] That primates look behind themselves when they see something in a mirror is not uncommon (see section 9.4).

[111] A question for future research is whether apparent self-interest in front of mirrors is a result of image testing or vanity.

If Premack is right a prediction would be that Sarah could throw objects towards the screen and react with fear when someone gets too close to it also when a cartoon is played. Given, of course, that Sarah can comprehend cartoons. She would also be similarly excited when an upsetting story is read to her, given she comprehends language. I believe that Sarah's reaction was not motivated by a need of hers to express her attitude towards certain acts in *principle*, but by the perceived realness of the situation and what was directly *there*, in front of her eyes. The reason humans cry when watching sad movies is not because they allow themselves this luxury *despite* that it is not real, but because parts of them did never make the difference in the first place. "Involvement" is when this part takes over. Reality mode gets a revival in such circumstances. So does Sarah watch television as the typical human, only that she is expressing her involvement in a chimpanzee manner, or is her way different? I agree with Premack (1976) that free response television viewing alone cannot answer this question, but further tests are needed to map Sarah's pictorial competence.

9.4 Mirrors

Mirror self-recognition has been a favourite topic in modern and historic primate research alike (see Mitchell, 1999; Yerkes & Yerkes, 1929/1953). The modern interest, however, can be traced to the advent of the "mirror mark test" developed by Gordon Gallup Jr. (1970). An odourless mark of an unusual colour is placed somewhere on the subject's body which can only be discovered by using a mirror. In the original test the mark was placed on the eyebrow and on the ear. (White paint on these areas can be discerned in *fig. 13*) If the animal, following inspection in a mirror, touches its body in the places marked, it has demonstrated that it recognises its own body in the mirror image.

The Gallup mark test has been used to infer not only self-recognition but also self-conceptualisation in animals (e.g. Gallup, 1977; but see e.g. Mitchell, 1997). Great apes (see e.g. Swartz et al., 1999) as well as bottlenose dolphins (Reiss & Marino, 2001) and Asian elephants (Plotnik et al., 2006) have passed the mark test. The bonobo has not as yet been tested on the mark test as such but has shown mirror-guided self-investigatory behaviours (Hyatt & Hopkins, 1994; Inoue-Nakamura, 1997; Walraven et al., 1995). The gorilla, with the exception of Koko (Patterson & Cohn, 1994), has had a reputation of being notoriously bad at the mark test (e.g. Shillito et al., 1999; Suarez & Gallup, 1981), but has likewise shown self-directed behaviours in front of mirrors in targeted studies (e.g. Inoue-Nakamura, 1997; Parker, 1994), and is now also a confirmed self-recogniser as measured by the mark test (Swartz & Evans, 1994; Posada & Colell, 2007).[112] That the gorilla ever has been markedly different when it comes to understanding mirror images might in fact be an exaggeration. None of the great ape species really excel on the mark test. When looking at all the studies done until 1997, less than half of the subjects have

---

[112] The successful gorilla in Swartz and Evans (1994) is King, the same individual that referred to people, objects, and places by way of picture cards (see section 7.7).

been successful, regardless of species, and there is no significant species difference (Swartz et al., 1999).

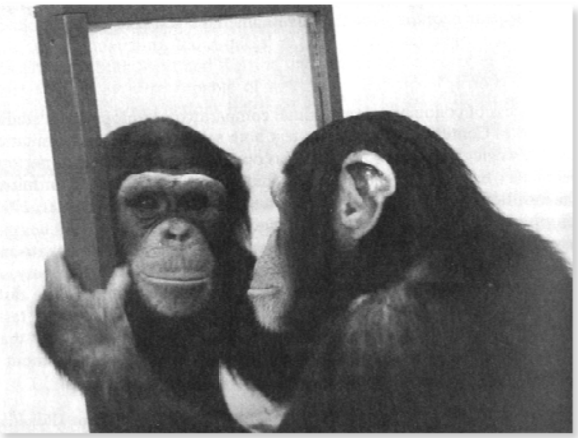


Figure 13. Is the chimpanzee Scott (4.5 years) viewing a photographer inside the mirror, the photographer in front of the mirror, or the reflection of the photographer?

From a pictorial perspective mirrors are interesting because the mirror image is similar to a picture in several ways. Importantly, Sonesson (e.g. in press a; 2003b) notes that when mirror recognition is demonstrated the mirror qualifies as a sign. Its expression is in direct contact with the visual senses, but the focus is on that which the mirror reflects, i.e. the body or scene in front of it. This furthermore implies that the things in front of the mirror are clearly differentiated from the image as such. That attention is on the body or objects in front of the mirror precludes that the mirror image is mistaken for reality, a transparent window, etc. Thus, the mirror image is about something other than itself and is clearly differentiated from this.[113]

Like the static picture the mirror image is an iconic sign, but differently from the picture the image in the mirror surface is continuously updated with the perspective of the viewer. New content can in this way be discovered by adjusting the mirror or the scene in front of it. Static pictures can also release new information as perspective is changed, but once optimal conditions are met there is a limit to what can be discovered. No matter how you angle the photograph of a person you can not necessarily look up her nose. In mirrors there is really nothing that corresponds to expression, as a close scrutiny does not reveal what the image is made up of, in terms of lines, pixels etc. (Sonesson, 2003b). The mirror image is behind the physical mirror surface, even when this surface is a poor reflector. In this sense the mirror is more like a window than anything. Unlike many pictures the mirror therefore fails to abstract in a selective and systematic manner. For example, one cannot create a reflective surface that blurs the internal properties of e.g. a face, while containing its sharp contours.

9.4.1 Development of mirror use

Although there is no systematic study of Viki's mirror use Hayes (1951) gives some information on its development. At about 4 months of age Viki first caught sight of herself in a mirror. Her first response was one of sudden fear. The fear was soon replaced by nervous curiosity and she made the common mirror response of investigating behind the mirror, perhaps looking for the chimpanzee in the mirror, or to

---

[113] Fagot et al. (2000) makes a similar claim regarding the different view of mirrors in "confusion mode" and "equivalence mode" respectively.

get to grasps with the properties of this new object. If Viki looked behind the mirror for clues about the apparent content in the mirror, it is noteworthy that she choose this method rather than to try and put her head through it. The common response of touching and sniffing mirrors, which also accompanies initial exposure, probably gives enough information to discourage such attempts. In addition, most apes that get access to mirrors also most certainly have experience with window glass. Interpreting mirrors as some kind of window is therefore not farfetched, and it would make sense to reach around it.[114] Although mirrors allow see-through, the laws of hard surfaces seem to apply. Viki soon started to make faces in front of the mirror, but it took another year before she stopped searching behind it "for the rest of the chimpanzee" (Hayes, 1951, p. 25).

Similarly, at 8 months of age the chimpanzee Gua showed interest in her own mirror image, but also showed that she did not yet know about the nature of mirrors. It "[…] caused her to reach behind this at first with caution and hesitancy, then in a continuous series of groping hand movements" (Kellogg & Kellogg, 1933/1967, p. 241).

Investigating mirrors by reaching or looking behind them has been reported in numerous studies on apes' initial responses to them (e.g. Inoue-Nakamura, 1997; Yerkes & Yerkes, 1929/1953). Remember for example the Köhler (1925/1957) chimpanzees in section 1.4. There is also similar data from e.g. Ladygina-Kohts (1935/2002) for her chimpanzee, and by Yerkes (1927a; 1927b) for a gorilla. Miles (1994) made the similar observations for the sign-language trained orangutan Chantek. He continued to reach behind mirrors right up until he passed the mark test when 2 years old. However, during the following year he failed the mark test repeatedly. Interestingly, his grabbing behind the mirror prevailed during this time, but stopped when he again passed the mark test at 3 years of age.

Chantek's ambivalent understanding of his reflection might not be unique. Two mirror-naïve subjects of about 4 years of age in Boysen et al. (1994) displayed a mix of self-directed and mirror-directed behaviours during their first ten minutes of exposure, among them reaching behind the mirror and sexually thrust against it, but also making contingent movements. A mark test was not performed, however. But after a mere 20 minutes of mirror exposure the utilisation of reflective surfaces transferred to using a small hand-held mirror to inspect the bottom of a foot.

Another case of ambivalent response is the home-raised gorilla Toto who showed mirror-guided self-directed behaviour at a young age when she studied the eruption of a new tooth in her mouth. However, she also acted on the mirror image. The account reads: "[…] she used to stand in front of the long looking-glass […] opening her mouth and gazing raptly at the little pearl-like protuberance, often preening herself before the glass, turning her head from side to side, delighted at what she saw. Then suddenly she would become angry at the figure in the mirror and attempt to attack and destroy it [...]" (Hoyt, 1941, p. 113).

---

[114] After the removal of a mirror, following an episode of social responses towards it, an orangutan in Yerkes and Yerkes (1929/1953) is described to "[…] gaze at the place on the wall where the new world had appeared to him and he approached it from all directions in turn as if to assure himself that nothing remained" (p.139).

Hess (1954) reports that the chimpanzee Christine, at an age of less than a year, looked behind mirrors, again the purpose being interpreted as "[...] to find the other little ape, but when she could not see it, she was content with making faces and kissing the image" (Hess, 1954, p. 12). At 2 years she had long since stopped looking behind the glass and instead used mirrors to inspect herself, especially after having draped her body in household objects. She smiled and "petted" the image.

Lin et al. (1992) found no evidence of mirror self-recognition in chimpanzee infants below the age of 2.5 years as measured by the Gallup mark test. But they did display some contingent movements, such as grimacing. The use of the mirror for self-directed behaviours was much more frequent for the older infants (4-5 years), and they were also the ones that passed the mark test. However, 2.5-year-olds, who failed the mark test, still showed some self-directed mirror-guided responses. These findings were replicated in Bard et al. (2006). In Lin et al. (1992) behaviours directed towards the mirror itself decreased as a function of age, but all ages, except the youngest ones (under 2 years), displayed social responses towards the mirror at initial exposure.

Inoue-Nakamura (2001) found that age of onset of displayed mirror-recognition is dependent on mirror experience. By exposing a chimpanzee infant daily for three months she could demonstrate self-investigatory behaviours at 1.5 years. However, although raised in a milieu rich in human cultural artefacts, like mirrors, the gorilla Koko did not display mirror-guided self-directed behaviours until being 3.5 years old (Patterson & Cohn, 1994).

The Gallup mark test was not developed in the late 1940s but Viki showed signs of understanding reflective surfaces when also she actively used mirrors to investigate inaccessible parts of herself, such as genitals and teeth. Such investigations are common behaviours and occur earlier than passing the Gallup mark test, which emerges between 4.5 and 8 years of age (Povinelli et al, 1993). In fact, typically only half of the subjects that show self-investigatory behaviours in front of mirrors pass the mark test, and there seem to be a marked decline with age, after having passed roughly the 16th year (Povinelli et al., 1993; de Veer et al., 2003).

Understanding mirrors as a looking-through object, but not a reflecting one, seem to be common in animals, as is confusing mirrors for a real scene and act directly towards its surface. Monkeys commonly attack mirrors or show fear responses towards their own mirror images (see e.g. Anderson, 2000). With prolonged exposure apes, on the other hand, usually get over this part. Appreciating that mirrors reflect things, but not be certain about what it reflects, seems to be another stage. It is common for apes to test the mirror image by e.g. performing unusual movements in front of it. Finally, to see mirrors as reflecting objects and understand what it reflects is the last stage. Investigation of specific parts of one's body probably takes place at this stage. This development is usually catered for by a familiarisation phase during which the subject has free access to mirrors. Without such a phase apes usually do not pass the Gallup mark test (Gallup, 1970). They thus need to learn the properties of mirror surfaces before it is possible to test self-recognition. From this observation follows that many negative results with mirrors can probably be explained by a failure to get to grips with reflective surfaces, rather than a lack of self-awareness (Loveland, 1986). That is, mirror-recognition is primarily a problem of signification, just like pictures. Bard et al. (2006) writes about mirror competence: “Prior to attaining this iconic symbolic capacity, mirror-guided mark touching is not possible and cannot be enhanced, even by fully informing about the mark” (p. 125).[115]

Sonesson (in press a) notes that when mirror recognition fails, it is not a case of failure to differentiate one’s own body or the surrounding scene from the image in the mirror, but understanding the different natures of the entities inside and outside of the mirror. After all, the image is mistaken for *another* individual. However, I would like to add that the mirror world might be perceived as different from the everyday world after all, but that this recognition has no bearing on understanding the real relationship and therefore prevails as a *kind* of reality. We might therefore see some reality based behaviours towards the mirror that are somewhat different from confusion behaviours, but are still not recognition behaviours. Such behaviours can take the form of testing the image. Initially it is tested as a social phenomenon, and then as a visual phenomenon. This is supported by data from e.g. Inoue-Nakamura (1997; 2001). This might also be the reason why contingent facial and bodily motions are displayed by chimpanzees not only in front of mirror, but also to some degree when looking at video tapes of conspecifics. Self-exploratory behaviours on the other hand are more restricted to mirror use (Eddy et al., 1996). Somewhere along the way, when “testing” the image, whether intentionally or not, the ape realises the correspondence between its own movements and the ones in the mirror. Monkeys, for example, display contingent behaviours but not self-exploration when confronted with mirrors (Inoue-Nakamura, 1997; de Waal et al., 2005).

Human children likewise follow the developmental sequence of first acting on the mirror image, including making social responses, then studying their own movements in the mirror, and lastly using the mirror for investigation and play “in the knowledge that it [is] not real” (Inoue-Nakamura, 2001, p. 297). When testing young subjects, human and chimpanzee alike, this development can span weeks or months, but with older subjects the whole sequence can be traversed in about one half-hour session (Inoue-Nakamura, 2001; Boysen et al., 1994). However, with a late exposure to mirrors only chimpanzees and children above 3.5 years old reach the stage of self-exploration. With early and extensive exposure, on the other hand, it can be pushed down to the second year in both species. Display of self-recognition in mirrors is thus truly a joint function of maturation and mirror experience.

Other that using mirrors to get visual access to hidden places or adore oneself, understanding mirrors as a reflective surface can manifest itself in a very different way. The chimpanzee Austin at the Language Research Center, for example, enjoyed using mirrors to redirect beams of light (Savage-Rumbaugh et al., 1998). He also

---

[115] This seems to be a recent conclusion, however. In Custance and Bard (1994), understanding of the causal relationship between one’s body and its reflection is the central explanation for the onset of mirror self-recognition.

did this when picture slides were projected into his enclosure, making the pictures bounce around the room.[116]

However, another common response to mirrors, less published but still striking, is to ignore mirror images altogether. Such habituation might occur when initial exploration fails to inform the individual about the interesting, useful, or entertaining properties of mirrors. But lack of actions towards mirrors does not need to imply that an animal has learned that a mirror is not a kind of reality. Menzel et al. (1985) describe how rhesus macaques that were habituated to mirrors and had seized to make social displays towards them, resumed this habit when food was introduced in the test they were working on. Especially when the reflections of their own hands came too close to food the monkeys responded by threatening their mirror image. Similar observations are made by Itakura (1987b).

9.4.2 The mirror picture

Mirror recognition can only be argued to be truly a pictorial mode competence if it can be shown that the perceiver is not dependent on continuous feedback from the image to understand it as a reflection of real events. Even if the image would magically stop moving the subject must show that it still recognises its own image. Or a short delay in the feedback can be induced and the ape should still be able to use the image to guide its actions.[117] Otherwise mirror use has very little in common with pictorial understanding and can at best be described as an isolated case; A case which is probably made possible by a striking correspondence between reflection and reality. But such correspondence need not be striking on all accounts. Movement correspondence has probably a very high level of salience in this process. Self-recognition in chimpanzees can namely occur also in distorting (concave and convex) mirrors, or mirrors that reflect multiple copies of an image (Kitchen et al., 1996).

The instantaneous visual feedback to one's own movements which allows self-recognition is what Mitchell (e.g. 1997) calls kinaesthetic-visual matching. It allows the mirror observer *some sort* of understanding of the reflection of the viewer's own body, without requiring a self-concept as a source for this perception of correspondence. This has a parallel to the stretchability of reality mode processing. A plausible principle in reality mode is that "what works, works." If an individual appears to be in two places at once in photographs, so be it. If one can attend to and use such information, there is no need to reflect on its improbability. Likewise it is a fact that a body is reflected in a mirror that perfectly matches one's movements. Bodies are very special and interesting objects, and it makes sense to investigate its unknown places when such a reflection allows it. There is no need to wonder how this is possible, or who the person in the mirror *really* is.

While Mitchell (e.g. 1997) seem to suggest that appreciation of reflective correspondence in mirror surfaces is already present when kinaesthetic-visual matching

---

[116] Such play arguably gave ample opportunities to learn that projected pictures have a very real source.

[117] Typically only a sensitivity to the difference between live and recorded video is measured in terms of occurrence of contingent movements. See Brooks-Gunn & Lewis (1984) for a review regarding children.

occurs, the argument can also be made that kinaesthetic-visual matching is in fact responsible for the discovery of these reflective properties in the first place. The data reviewed above strongly suggest that experience with mirrors is important for recognition to occur, but most accounts also report that contingent movements follow after eventual social responses, and always before self-directed behaviours. It is kinaesthetic-visual matching that is kicking in during the contingent phase. In fact it would be very difficult indeed to learn about the reflective properties of mirror images if one has never been able to view the contingent reflection of one's own body, if so only an arm sticking out of a hole in a search-task (i.e. Menzel et al., 1985). As an illustrative anecdote, remember that Köhler's (1925/1957) chimpanzees did not start to use reflective surfaces until after they had been subjected to self-recognition tests.

Without the possibility for kinaesthetic-visual matching two entities cannot be bridged outside and inside of mirrors. They would forever be tandem events. However, a correspondence between "tandem events" can also be learned, and be utilised in mirror-guided tasks. Understanding of *reflection* per se is then not necessary, only the appreciation on an *if-then* relationship. Such learning of reflected images has been demonstrated in e.g. pigeons (Thompson & Contie, 1994).

The mirror image, as well as the live video broadcast, is set apart from other pictorial media in that its nature is directly testable. With continuous visual and bodily feedback the subject can learn about the nature of the reflective surface and how it relates to the objects outside of it. The same manipulations cannot be done with static pictures, replicas, or video recordings. This means that a picture – referent relationship, including differentiation, can be discovered on its own by a subject, without social scaffolding.

However, while mirror-image use seems the most promising picture-like competence in apes, it might be confined to the media and not generalisable. While an ape probably learns a lot about reflective surfaces with mere exposure, it learns probably nothing about static pictures from those experiences. While one borders a pictorial mode with mirror reflections, one can therefore still view photographs in reality mode, and perhaps drawings in surface mode.

Even monkeys (*Macaca fuscata*), following extensive training on using mirrors to guide their hands in a search task (Itakura, 1987a), can use mirrors to recognise that an object is fastened to their own body (Itakura, 1987b), and presumably not to that of a monkey in the mirror, although this cannot be ruled out. Remember the concerns about "tandem events" above. However, when seeing objects, photographic slides of objects, or individuals in the mirror, these macaques typically turned around after observing the reflection to view the objects directly. Furthermore, illustrating my claim above regarding different modes being active under different circumstances, the macaques made social responses towards photographic slides, but not to the mirror reflections of these (Itakura, 1987b).

There seems to be only one primate study that directly compares recognition in mirrors and photographs. Tobach et al. (1997) could not convincingly demonstrate self-recognition of life-sized portraits in six orangutans, but the one subject that showed most preference for her own portrait was the one, among two, that showed self-directed behaviours in front of a mirror. In one session she gave the impression of possibly comparing the two representations of her own body by repeatedly moving between the mirror and her own portrait, which she singled out among several.

Like other animals, human children seem to learn to recognise themselves in mirrors before recognising themselves in photographs (see Brooks-Gunn & Lewis, 1984). One can argue that a more proper comparison is perhaps not recognition of *themselves* in the respective two media, but of static objects. However, remember the distinction between reality mode and pictorial mode processing. Even though self-recognition as such complicates matters, recognising oneself is a stronger case for a referential and differentiated view of the static picture.

Koko, the signing gorilla, gives us a concluding remark about the pictorial nature of mirrors, recorded in an interaction with a caretaker on a particularly boring afternoon. In response to the request to find an interesting picture in a catalogue Koko browses a few pages and points to a pot. (Koko is described to be in a bad mood.) When asked to show a pretty picture Koko points to a roasting pan. The caretaker, about to give up, asks Koko to find a scary picture, to which Koko replies by picking up a mirror and places it in front of the face of the caretaker (Patterson & Linden, 1981).

Chapter 10

Producing pictures

While this thesis focuses on receptive pictorial competence a chapter on picture production is in order, especially as there are claims of representational drawing and painting for several of the apes involved in language research (e.g. Fouts, 1997; Patterson & Gordon, 2001; Savage-Rumbaugh et al., 2001).

But let us first consider a more basic prerequisite for imbuing marks on a surface with meaning, which is that of attending to such marks in the first place. Singling out marks on a surface is indeed necessary for imbuing it with relevance.[118] Specks of dirt where it is not supposed to be, cracks in a wall, mist on windows, are all deviating details on something otherwise well known that can grab the attention of any curious ape. This is not trivial. The ability to be engrossed in one's experiences of the world, which manifests itself in behaviours such as playing with one's sensations, is frequently manifested in apes, but is perhaps not widespread in the animal kingdom. Common examples seen in apes are blindfolding oneself in play, toying with the feel of dripping water on one's body, or ask to be tickled.[119] Thus, apes spontaneously and actively pay attention to and act on features of the world to seemingly construct experiences for themselves. Using pen and paper is one such activity that apes readily adopt and enjoy when provided with the means.

There is perhaps only one account in the literature of spontaneous possible drawing behaviour in apes that was independent of human intervention. A gorilla, between one and two years old, at London zoo was observed in 1939 to trace his own well defined shadow on a cage wall with his index finger. This was made in an apparent air of focus and goal-directedness, and he directly repeated the action twice. Later attempts to replicate the event by supplying the gorilla with shadows cast through artificial lightening failed. He was never observed to repeat the activity again in the one year before his death. This brief report was published in the journal Nature and called *Origins of Human Graphic Art* (Huxley, 1942).

A less ambivalent example from a human milieu is the chimpanzee Austin, who actively used his shadow in what seems like play. By for example stepping between a projector and the projection on a wall he would impose his shadow on projected movies in order to "chase the chimpanzees in the movies" (Savage-Rumbaugh et al., 1998, p. 35). When noticing his shadow on a wall outdoors he would also move in

---

[118] For a more in-depth discussion of marks on surfaces, see e.g. Ittelson (1996)

[119] Experimenting with one's own sensations makes a case for subjective experience in animals. The forms and extent of such experiences is not the topic of this book, however.

unusual ways and note its effect on the shadow. While not being a case of drawing as in creating marks, it can still be argued to be an intentionally constructed image.

Shadow play in chimpanzees was studied by Boysen et al. (1994), but in an imitation context. Both a 3 year old and an 8 year old copied the making of manual shadow puppets on a brightly lit wall. They also showed that they understood the shadow cast by objects. Only the older animal, however, also displayed recognition of her own shadow.

10.1 Making marks

It is perhaps an exaggeration to say that apes spontaneously take to drawing when provided with the materials, because they also seem to benefit from seeing the acts modelled. Common for several findings is that the apes discover the joy of scribbling by observing with interest the activity of writing or drawing by their caretakers. They are then described as imitating the ability.[120] For example, the human child Donald in Kellogg and Kellogg (1933/1967) scribbled spontaneously when he got hands on a pencil in a test at 14.5 months of age, while the chimpanzee Gua, 12 months old at the time, had to see the activity demonstrated before she followed suit. At a testing session one month later she scribbled spontaneously when given pen and paper. Soon scribbling becomes a rewarding activity in itself, independent of social reinforcement. In fact, some subjects often choose pen and paper at the expense of food (Schiller, 1951).

The chimpanzee Viki in Hayes (1951) also started to scribble with pencils from an early age. Already at nine months she started to imitate C. Hayes' writing, superimposing her own scribble on her foster mother's. Her scribble later transferred to books, floors, walls, furniture and apparently any piece of paper she could get hold off, using any pencil, crayons, screwdrivers and the sharp corner of a wooden block. Every part of her body could be recruited in the action. She also scribbled in mist produced on windows, which has likewise been reported for the gorilla Koko, who enjoyed creating her own mist by breathing on windows (Patterson & Linden,1981).While the above accounts serves to demonstrate Viki's power of generalisation when it comes to mark-making activities, her behaviour also hints at the fact that leaving a graphic mark was not *always* the function of her scribble. Wooden blocks presumably leaves behind poor graphics.

Joni in the home of Ladygina-Kohts (1935/2002) similarly imitated scribbling after observing his foster mother. Like Viki's, Joni's scribbling had absolutely to occur in the very same places as Ladygina-Kohts' in the beginning, even when given paper of his own. Later he got more independent, and as soon as Ladygina-Kohts tried to interfere or help with Joni's drawing he immediately lost interest. His scribbling went through a change with time and he started to make longer, continuous lines, as well as small crosses, or acute angles. No age is given for this transition.

---

[120] Imitation is here used in the broadest sense of the term. For a review of diverse social learning abilities in apes, see e.g. Whiten et al. (2004).

Ladygina-Kohts found it difficult to judge if Joni enjoyed the visual aspect or the motor aspect of drawing the most.

Apes often do not want to be disrupted when painting, and they do not like to be urged to continue once they consider themselves finished. Authors (i.e. de Waal, 2001) have imagined this to be about the painting as a finished or unfinished product, when it rather could be about the activity. Apes probably do not like to be interrupted in any activity that they enjoy, or forced to continue something against their will. It does not have to do with considering the *product* finished or unfinished at all.

Other apes, like the adult chimpanzee female Alpha in Schiller's (1951) study, do not seem to care if they are interrupted, and they are not at all protective of their finished work. Schiller (1951) thus concluded that Alpha was not particular interested in the end product of her drawing, but rather enjoyed the activity. However, her interest was by no means solely in the motor activity since she did not work when given pointed sticks instead of pencils in control sessions. She also stopped working as soon she broke the pencil points, and instead made her wish for a new, sharpened, pencil known to the experimenter. The perceptual aspect was thus very much an integral part of her enjoyment. Similar conclusions can be drawn from a longitudinal study of chimpanzee infant scribbling on computer touchscreens by Tanaka et al. (2003). All subjects scribbled substantially less when there was no visible colour feedback. In addition, different colours elicited different amounts of scribble.

It is possible that apes, perhaps especially when painting, seldom focus on the end product, i.e. the sum of all the strokes, but enjoy each stroke on its own. A suggestive counter example can be found in Morris (1980, cited in Lenain, 1997) where the chimpanzee Congo seemingly continues where he left off in completing a fan-pattern, his hallmark motif. Otherwise claims for aesthetic composition seem to be weak (Lenain, 1997).

But in accordance with ape painting being more about the process than the product, still some claims of a compositional view of the *activity as such* can be made. For example, some apes seem to apply a relational view of the marks on the paper and not only see each mark as separate. When for example a shape is positioned by an experimenter to the right, the subject subsequently focuses its drawing equally far to the left. Or if given a paper with a pre-made shape on it, they tend to focus their scribble inside the borders of this space. Both effects were found in Schiller (1951) and Morris (1963) (*fig. 14,* p. 186).

However, in Schiller's (1951) study glued-on, cut-out, or dotted shapes, were only marked within if they were larger than an inch in diameter, unless there were several of them in which case each received a check mark. When pre-printed shapes are too small ape painters instead seem to focus on their borders, or crosses-over the whole shape. Morris (1963) believed the latter behaviour to be a form of manipulation of what was perceived as objects in the paper, thus a case of reality mode processing. Schiller made the same observation, noting that Alpha poked at and explored the shapes, as if using the pencil as a finger "in an attempt to scratch beneath the figure or to scrape it off" (Schiller, 1951).

If lines instead of shapes were pre-drawn they were either specifically targeted (at an angle) or used to divide the paper into separate drawing surfaces (Schiller, 1951; Morris, 1963). In conclusion, compositions made by apes seem to be quite affected by the type of visual surface that they can interact with.

Figure 14. Targeting of a pre-printed shape by a gorilla (left) and claimed balancing of a shape by a chimpanzee (right). From Morris (1963).

The balancing of skewed pre-printed shapes is the most promising indication of a relational view of the drawing. Both Morris (1963) and Schiller (1951), found that if single smaller shapes were skewed off the centre of the paper, they were balanced with marks on the opposite side. Schiller (1951) judged this effect to be a balancing of mass.

Balancing might only be true for pre-printed, or glued on, shapes. If balancing or filling-in would also be done in relation to marks previously drawn by the subject itself, the claims for composition would be much stronger. This does not seem to have been the case in any of the studies. The subjects did also not readily complete incomplete pre-printed geometrical figures when the shapes were continuous, but in Schiller (1951) Alpha targeted an empty space if the shape to be completed was made up of lower-level elements, as in filling in the space where the missing square would be in a circle of squares. Neither Schiller nor Morris came up with the idea of testing incomplete figural drawings, for example of a face missing an eye. That could be a promising experiment for the future.

With regards to balancing and completion it should be noted that neither Smith (1973) nor Boysen et al. (1987) managed to replicate Schiller (1951) and Morris (1963) when subjecting three young chimpanzees respectively with similar materials. All enjoyed scribbling, marking stimulus figures and filled in large empty spaces, but no claims for composition effects can be made from the results. Terrace (1980) on the other hand report that the young chimpanzee in his sign-language research, Nim, habitually joined circles that had been pre-drawn on paper by scribbling back and forth between them.

Another example of composition is the careful choice and application of colours. The juvenile chimpanzee Congo in Morris (1963), for example, became famous for his fan patterns of discretely separated colours. When the process rather than the composition is the focus paintings tend to become a grey mess.

When Viki started to paint with colours she smeared the paint together, clearly enjoying the physical properties of paint. With time, however, she started to separate colours and seemed to pay attention to their visual effect. She even washed the brushes on her own accord between colour changes (Hayes & Nissen, 1971). At the same time a transition could be seen in her pen scribblings. At 4 years of age she started to make a new type of "hen scratchings" with great concentration. This might mirror Joni's change from just lines to small acute angles in Ladygina-Kohts (1935/2002). As with her painting, Viki's first type of scribble had likewise been more about the process than the product. When human comparison subjects of Viki's age started to make representational drawings Viki and Joni held on to their hen scratchings.

When looking at ape paintings in Lenain (1997) and Morris (1963) the two styles represented in Viki's painting development can easily be distinguished. They could be called a smearing style and a compositional style. However, a feeling for composition, as in balancing figures, and insensitivity to colour use, as in smearing, are not mutually exclusive. Alpha did not change the location on where she was scribbling when handed a new colour, nor did she take into account the colour of the paper when choosing colours, which resulted in designs that could barely be seen. But still she was very apt at balancing and completing some geometric designs (Schiller, 1951).

10.2 Marks that mean something

A most surprising account is given in Hoyt (1941). The gorilla Toto, age at the time unknown, developed a ritual with Hoyt's mother which consisted of drawing a face in the air with the index finger. A circle for head, three dots for eyes and nose and a line for mouth. Toto seemed to have copied this action in a recognisable fashion and used it to greet the mother whenever she visited. "Later, Toto transferred this crude representation of a face to the sidewalk with soft white limestone, and while these pictures of hers would scarcely justify the hope that she might become a great portrait painter, they were definitely recognizable as representations of a human face" (Hoyt, 1941, pp. 150-51). This story gives rise to two interpretation problems. The first is exactly how recognisable these drawings were, or whether they had to be filtered through the eyes of a loving mother.[121] The second is if the drawings were in fact only a generalisation of the well familiar motor patterns of the gestures in the air, or made to *represent* the scribble in the air and/or a (human) face. Without any discriminative actions towards the drawings, like a kiss on the mouth (which in this case would perhaps qualify as pretence and not confusion), we can make little of this information.

A bonus with language trained individuals is that they can comment on their own drawings and paintings. Fouts (1997) tells us that the sign-language trained chimpanzees in his care paint representationally and name their paintings. The favourite motif of Moja, one of the chimpanzees, was birds. Washoe on the other hand, the most famous of the inhabitants of the Chimpanzee and Human Communication Institute, likes to paint in fiery colours, and name her paintings accordingly. Lenain (1997) notes in regards to one of the Institute's reports that when Moja names her productions, on only five out of the 26 occasions does she name them as if they were

---

[121] Filtering through the eyes of a hardcore sceptic might be equally unsuitable.

depicting an object. On the remaining occasions she names drawing materials or signs "scribble" or "colour," which is her word for the activity of producing paintings.

Koko the gorilla also name her creations, her favourite subjects apparently being birds and alligators (Patterson & Linden, 1981), but besides objects she has been reported to also depict emotions, such as anger or love (Patterson & Gordon, 2001). It seems that data for the spontaneity of the naming has not been given, or whether Koko remembers her names.

As with human children the best pedagogical stance is not to question the naming of the creator but to act as if the likeness is perfectly clear. Thus, if animals put paint on paper and (spontaneously) comment on their creations, Fouts', Patterson's, and others', roles as interspecies communicators is to suggest that the apes paint representationally. The controlled experiment for children in the same situation is to show them their own pictures at a later occasion and ask them what the pictures represent. It has turned out that children have excellent memories for what their drawings have been said to represent, even when likeness is minimal. 3-year olds making crude depictions can remember the designated content for up to 3 months, and 6-year olds up to 6 months (Gross & Hayne, 1999). In the latter case, however, there is notable iconicity to go by in the renditions. Such check-ups does not seem to have been published for named ape paintings.

A question of special interest is whether pictures produced by language-competent apes are purely conventional, or show elements of iconicity, and furthermore if iconicity in those cases stands as the major ground for a sign relation. The problem with ape art is that the iconic nature cannot be inferred until after naming of the piece.

The notable (published) exception would be some of the drawings of lexigrams by the bonobo Panbanisha (Savage-Rumbaugh et al., 2001). If human communicators cannot see what lexigram is pointed to at the portable lexigram boards because it is too far away, the bonobos can be asked to draw with chalk on the floor to make a larger lexigram. The bonobos then point to the chalk version instead of the smaller version on their boards (Segerdahl et al., 2005). However, it is not reported how successful such communications usually are. Besides the fact that some lexigrams are more easily copied than others, Panbanisha's spontaneous drawing is reported to be most clear when she is communicating a reason for going outdoors, which can refer to specific places or foods that can be found at those places. But her drawings are still crude, and if given a random sample of her work it is questionable if independent raters would make the same judgements of which lexigrams they represent. To Panbanisha's defence should be said that iconicity is relative and not independent of context. The audience she draws for often has background information that helps them in their interpretation. *What are we already conversing about? What do Panbanisha usually ask for? What do her previous drawings look like?* In those cases where the drawings make sense due to context, but resemblance still being a crucial element for interpretation to be possible, an iconic sign-relation can be said to have taken place. Panbanisha can be said to have successfully drawn a representational *picture* of a lexigram.

A further example: "After a talk with Penny about spiders, Koko takes a pencil and draws black squiggles on a piece of paper [...]. 'Spiders,' she explains. The orange scrawl, Koko says, portrays her drinking glass" (Patterson, 1978b, p. 448). These squiggles, out of context, just look like nonsensical scribble. But given context, and background knowledge of what other scribble of Koko's looks like, it would be possible to agree with her that the black marks and the yellow patch indeed appears to be spiders and a drinking cup. Koko's choice of colour, if nothing else, would be iconic. However, we are not informed about the colour of Koko's real cup at the time. We are also not informed about why this particular drawing was chosen for presentation in National Geographic Magazine. A qualified guess is that it was one of the best examples that could be found, and if that was the case Koko's ability to produce drawings or paintings that *pictorially* represent objects seems minimal.[122] It should be emphasised that this "lack" of ability can be one of gorilla motor control rather than aspiration. If this is the case, colour and other motor independent choices should show some iconic properties. Indeed this is what is reported for Koko. Patterson and Linden (1981, p. 137): "[Koko] can come up with some fair representations, especially if she is copying from a picture or model." They specify: "She uses appropriate colors and gets objects in their correct place." But when Koko copies circles, they turn out more like triangles, again suggesting that motor and/or perceptual ability rather than intent is to blame. The copies of lexigrams made by Panbanisha are likely crude for the same reasons.

10.3 Why drawing is difficult

The act of drawing is very different from the act of comprehending pictures. Motor control, understanding drawing materials, and the process of translating a mental image to a physical one, are all competences that grow on top of comprehending that a picture depicts.

Drawing, in terms of motor activity, requires a lot of practice. This is evident from studies of the development of human drawing (try and draw with your non-dominant hand), and then humans are a species with particularly useful predispositions for drawing. We sport a hand, wrist, and arm freed from the requirements to support a body in knuckle walking, and perhaps adapted for a material culture dependent on fine motor control. Apes are by no means devoid of using their hands with precision, but how this precision is learned and controlled is difficult to study, and unfortunately remains an unknown (Byrne, 2004). In order to feedback the motor system and start to draw with precision it is necessary to link one's intent of producing anticipated shapes with the control of one's drawing utensils. There must be a goal. A strive for a circle is seen as an important step in children's drawing, and the completion of one as a milestone (Golomb, 1992). An equally important transition is then to be made from seeing lines as "paths" to seeing them as describing shape boundaries (Spielman, 1976). If one is still struggling with the motor de-

---

[122] Her paintings, of course, can be quite beautiful as art nevertheless.

mands of completing a circle there is little hope that one will produce much else but ape art.

The role of social factors cannot be understated in relation to children's drawing development. The first circle is often attributed to the goal of drawing a face. This no doubt is an effect of drawing with older humans that help you name your scribble for you, and consequently you strive for shapes that are asked from you.[123] Apes have been reported to prefer to draw in private, and even stop their activity if someone persists in wanting to take a closer look (e.g. Schiller, 1951). This might be an important difference with implications for conventional drawing development.

Drawing is also a material activity. It entails knowing about surfaces, utensils, colours, inks, and so forth. For some animals some materials are confusing, difficult to work with, or plain scary. For example, a capuchin monkey, Pablo, learned to scribble on the walls on his cage, but he never managed to transfer this to paper (Morris, 1963). Two rhesus macaques learned to draw when reinforced with chocolate candy on plastic sheets and blackboards with crayons, but never on paper, which they preferred to tear to pieces (Brewster & Siegel, 1975).

Inability to transfer drawing might also indicate that it is a learned ability specific to certain situations and not a general and flexible one. The two above macaques developed independent stereotypic drawing styles, which can as much have been a result of the heavy reinforcement received as of individuality. Both monkeys looked at their drawings less than 50% of the working time and they refused to draw when the experiment moved to a different room. It resumed when back in the original one. Neither of the macaques drew spontaneously when given the opportunity outside of the experimental setup. Often their arm movements used in drawing continued after they had dropped the crayon. These findings speak for the interpretation that these macaques were not drawing, but learned to make motor responses in exchange for chocolate candy.

The patterns made by the macaques can also be found in the repertoire of the chief subject in Morris' (1963) experiment, but the chimpanzee was somewhat guided by visual feedback, while the macaques seemed to ignore that part of the process altogether (Brewster & Siegel, 1975). Furthermore, chimpanzees readily transfer their drawing from paper to other materials, like cardboard or dry leaves (Schiller, 1951), and the household at large (Hayes, 1951; Ladygina-Kohts, 1935/2002). But they can also have some taboos, for example plywood and other heavy materials in Schiller (1951). The chimpanzee Joni learned to dip his pencil in ink when scribbling, but his dipping of the pencil often had no relation to its need for more ink. However, he still generalised his ink knowledge to wetting sticks with his saliva before scribbling with them (Ladygina-Kohts, 1935/2002). A mix of causal understanding and blind routine seem to have been present in these individuals.

The clinical psychologist Lightner Witmer tested the stage chimpanzee Peter in 1908 on a battery of tests he used with children (Candland, 1993). Peter's age and background is unknown, but he seemed to easily learn new skills from only a few demonstrations or on his own accord. He is reported to have been good with tools,

---

[123] Parents of course want those first circles to be them.

including roller skates, and could thread a needle with ease. He responded adequately to several spoken commands.

On Witmer's Peter failed with imitating a simple pattern of three pegs in a board and placing blocks in the corresponding holes in a form board, and he tried adequately but failed in lightening a match. But Peter succeeded to use a cigarette, unlock padlocks, use a screwdriver, open a box, spit, string beads, and, which impressed Witmer, copy the writing of the letter 'W' on a blackboard (Candland, 1993) (*fig. 15*).[124] Before the writing trial started Peter seized the chalk and spontaneously scribbled in the corner of the blackboard. This production is not preserved. Witmer made a capital 'W' and asked Peter to do the same. Peter made another scribble. This shape is said to be similar to the spontaneous one. Witmer judged that Peter had not paid sufficient attention and therefore retraced his 'W', again asking Peter to imitate. This time the chimpanzee reproduced a 'W' as good as Witmer's. Then Peter scribbled a bit more. Witmer asked him verbally to do the 'W' again and Peter complied, making a near perfect 'W'. Unfortunately Witmer settled with this demonstration, even though he himself was critical and suggested that Peter could have copied the physical movements rather than the written shape itself: a "motor-minded" solution (Candland, 1993). Could Peter have copied the 'W' if he had not seen Witmer first make it? It is a pity Witmer did not test this, or other shapes than 'W' that would have required more attention to the drawn form as such.

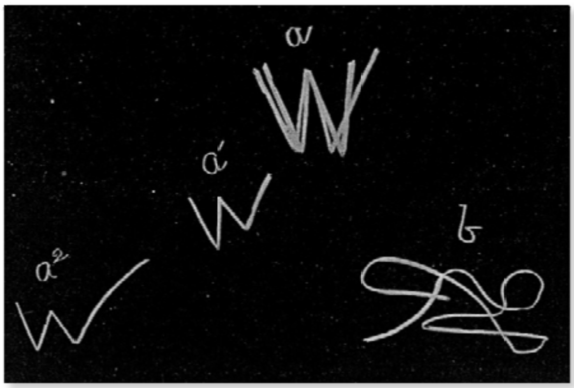


Figure 15. Peter's imitations of the letter W. A: Witmer's twice modelled 'W'. A1 and 2: Peter's reproductions. B: Spontaneous drawing by Peter. From Candland (1993).

Incredible as it may seem, Peter's performance is supported by the imitative abilities of Viki. At 3 years of age she could imitate the drawing of lines, circles and crosses; although she did not rise the pencil when doing crosses (Hayes & Nissen, 1971). When a model changed to scribble from a straight line, Viki did the same (Hayes, 1951). Neither in Viki's case did the experimenters seem to have tested the copying of finished figures, where motor cues were controlled for. Viki was an avid imitator of human actions (e.g. Hayes & Hayes, 1952). Of special interest here is that she could copy the stepwise building of a tower of colourful wooden blocks. Hayes (1951) report that they were anxious to see if she could also copy finished towers or pictures of towers, but they do not report the result of this endeavour other than that they mentioned that she imitated the finished result of a three-block tower at the age of 3 years. Viki could also copy the manipulatory actions on lumps of clay, but does not seem to have copied clay forms. On her own accord, she spontaneously constructed towers and structures of wooden blocks or Tinker Toys. Some of these structures she kissed and covered in

---

[124] We are not told why 'W' was specifically chosen for this experiment, but if e.g. Spielman (1976) is right, path-drawing should be easier than shape-drawing.

blankets, very much as she handled dolls. This was when Viki was 2.5 years old, in a period when her play in general is reported to have gotten more "imaginative" (Hayes, 1951).

The signing orangutan Chantek was tested on imitation of drawing of circles and lines within the "do as I do" paradigm. He tended to either retrace the form made by the experimenter, or draw partial copies. But a few of the straight lines seemed to be full imitations (Miles et al., 1996). Although his freestyle drawing is said to resemble the style of a 3 year old human he is not said to produce representational drawings (Miles, 1990).

Regardless of the fact that Witmer thought Peter to be motor-minded (remember that he failed on matching peg patterns and block shapes but succeeded on most tool tasks), Witmer's description of Peter is one of admiration. It is possible that Witmer attributed more to Peter than was actually going on, eager to prove his point that the "missing-link" is still among us in the form of chimpanzees (Candland, 1993). But subjecting Peter to the same type of tests and deductions that he applied on children, and seeing that there were similarities, at least he cannot be accused of double standards. Above all he understood that the similarities, as well as differences, had to be studied in more depth, and that it was to be the work of future child rather than animal psychologists (Hornaday, 1922). He reckoned that given a human upbringing, a chimpanzee might well learn language. This was to become an inspiration for the first cross-breeding experiment with the chimpanzee Gua 20 years later (Kellogg & Kellogg, 1933/1967).

An issue in child research has been whether young children who have reached a figurative stage predominantly draw what they see or what they know, or a mix of the two. This is usually called visual and intellectual realism. Several studies have shown ways to affect which style the child uses when copying a model item, such as an object. For example Lewis et al. (1993) have pointed out the importance of the instructions given by the experimenter. What these various affecting variables seem to have in common is that they guide the attention of the child back to the model item. Thus, rather than distinct stages in drawing development, intellectual and visual styles reflect different attentional strategies (Sutton and Rose, 1998). What makes it seem like a developmental issue is that younger children's attention tends to wander from the model item as soon as they start drawing, while older children remember to continuously check their progress against the view of the model.

If we are to generalise this finding to ape (representational) drawing it means that we cannot be certain to know which aspects the animal attends to and chooses to reproduce, unless we engage in communication while the drawing is done. If such interaction would take place it is likely that iconic relationships will become more apparent to the onlooker.

Human children take a long time to develop drawing copying skills. Rather than motor ability, the parsing of features and transferring this parsing into drawing seems to be the major problem. Young children predominantly approach a copying task bottom-up. They start out with local features and then build up to the complete form. Two lines forming an 'X' is for example typically copied as four shorter, disjoint lines at 4 years of age. Not until the age of 6 do children produce the 'X' copy by crossing two lines. For complex pictures children do not seem fully able to work top-down until the age of 9 years. When copying complex pictures from memory instead of direct copying, children of all ages (6-10 years) seem more likely to use continuous lines and global shapes. Older children fill in more detail, but also remember things that were never seen in the model (Akshoomoff & Stiles, 1995).

10.4 Alternative tests of productive skills

For the above reasons we should not be surprised if Koko's copies of models are restricted to colour, that Panbanisha's drawn lexigrams are crude, or that Viki makes her X:es without lifting the pencil. The most promising way to study drawing as a compositional ability might be to do away with motor demands altogether, as well as construct tasks where attention does not wander.

Kawai and Matsuzawa (2001; Fujita & Matsuzawa, 1990) report results from computerised tasks where the chimpanzee Ai, versed in symbol use, construct lexigrams by combining their elements in delayed matching-to-sample (DMTS). In a test of working memory she performed at 80% correct with a 32 second delay. Given the number of elements to combine chance performance would have been 1.2%. Human comparison subjects performed at 70%. However, only humans performed better with semantically meaningful lexigrams than with nonsense-designs made from the same elements. It seemed that mental rehearsal in Ai is not a case of "verbal" rehearsal. This finding is only tentative though, since Ai has shown the human trend in earlier research. It might just be the case that there was no demand for verbal rehearsal for the specific setup or delays used. A more serious limitation of the study, with regards to reproductive memory for composition, is that the nine different elements used to make up the lexigrams are always combined in a set of three, and each element has always the same position regardless of which lexigram it is included in. In terms of memory Ai only had to single out which elements were involved and remember three identities (e.g. *white circle, horizontal line* and *black circle*) never needing to retain the relation between these, as in a proper composition. However, in other tasks Ai has shown an ability to keep at least five discrete items, in terms of both identity and placement, in memory simultaneously (Kawai and Matsuzawa, 2001).

An approach that can be accused of the same limitations as above, but actually happened to demonstrate that also relational properties are included in memorisation of a pattern, is given in Vauclair et al. (1983). In the "memory for pattern reproduction" paradigm lights in a matrix are lit for some time, and following a delay the subjects are then required to light the same lights that they have just seen. Two experimentally naïve juvenile chimpanzees learned to reproduce horizontal, vertical or diagonal lines made out of three lights. Diagonals were more poorly reproduced and varied the most with length of delay (0-5 sec.). The same has been found for monkeys and for human children (Vauclair et al. 1983). When a fourth light was added to make more complex patterns, diagonal patterns again fell behind. If patterns were remembered as distinct lights there should be no such difference. Now they seemed to have been remembered as forms, and some were harder than others.

In fact only horizontal and vertical patterns seemed to have been remembered as forms. An analysis of the order in which the three lights were pressed in every trial revealed that diagonals were reproduced in a less systematic manner than the other patterns, supporting a conclusion that diagonal patters had indeed been memorised as independent elements. If apes have problems with reproducing certain simple forms (but not others) his is of course a finding that has implications for the possibility of figurative drawing as we would recognise it.

Drawing in chimpanzees has also been studied by the Kyoto laboratory using computer touchscreens. Besides free response finger painting that can be studied at an earlier age than the complex manipulation of brushes and pens allow (Tanaka et al., 2003), touchscreens allow for controlled manipulation of conditions and excellent recording possibilities. Iversen and Matsuzawa (1997) report that chimpanzees (one of them Ai) can learn to draw lines on a touchscreen, going from a defined starting point to an end point, and also that they can copy direction, and to a lower degree length, of a model line that is present on the screen during copying. However, during copying when the end point was lacking the subjects instead tended to draw towards the model line, or towards the corner that contained the area that the chimpanzees pressed for indicating that the trial was finished. These places thus took on the role of the end point. Successful performance was overall unstable and strongly suggests that the apes did not learn the task as a copying task as such. However, the experiment showed that in terms of perception and motor requirements, chimpanzees can draw lines between two spots on a touchscreen. In section 5.1 a similar competence was reported for Viki with multiple dots, using pencil and paper. The Kyoto chimpanzees can now also trace on top of lines in geometrical figures, but this required considerable stepwise training (Iversen & Matsuzawa, 2001).

Another way to study composition of pictures that demands even less motor ability is to work with a cut and paste method. This was attempted by Premack (1975), who studied the reconstruction of facial stimuli in his chimpanzees.[125] A template of a blanked-out chimpanzee face was used and two eyes, a nose and a mouth were to be placed within the blank area (*fig. 16*). The pieces were cut as to not form a jigsaw and hence provide shape cues. Importantly the chimpanzees were not shown a completed picture before they were tested. Had the Premack chimpanzees first seen a completed picture, or a model completing the task, it is possible that they would combine the elements as a matching task rather than as a puzzle task, bypassing the process of completing an actual meaningful motif.

While the free response procedure of Premack (1975) yields most valid results when it comes to the spontaneous picture production of apes, pictures completed in a non-ideal way are difficult to interpret. But at least one chimpanzee of four, Sarah, seem to have readily reconstructed from trial one the photographic chimpanzee face from the cut-out parts. However, the subjects were food reinforced irrespectively of performance and it is possible that Sarah's solution was one of chance. All four chimpanzees had different styles and it is possible that this idiosyncrasy was cemented already in the first trial due to reinforcement. Nevertheless, that they developed distinctive styles in arranging the cut-out elements entails that they handled

---

[125] A similar approach was used by Terrace (1980) but details are not given.

the stimuli in, if not compositional, a relational manner. They e.g. lined up the elements beside each other or on top of each other. It is reported to have been very clear at which moment they considered the product finished, and communicated this to the experimenter. This is a sign that they had a clear idea of the task, i.e. that they were to put out all the items within the boundaries of the large photograph (Premack, 1975). It is unclear how they perceived the actual stimuli, though, but one cannot rule out that they recognised the contents of the cut-out pieces but did not know how to, or indeed why, combine them. Premack and Premack (1983) mention that subjects who failed in this completion task could use cut-out noses, eyes and ears in conceptual MTS tasks.

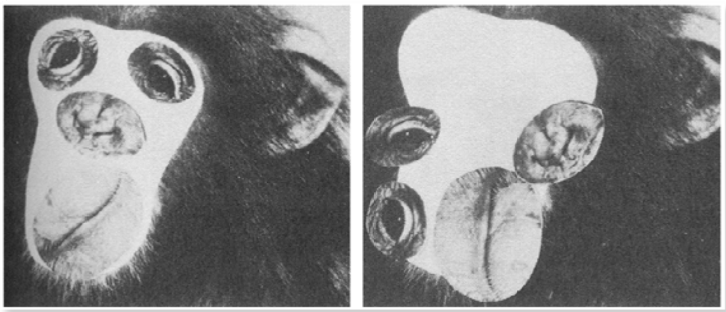


Figure 16. Two responses by chimpanzees to a jigsaw puzzle task. From Premack (1975).

Perhaps a reason to suspect that Sarah did indeed not only recognised the individual elements but also the potential of their combination is that she performed some seemingly meaningful transformations (Premack, 1975). These were meaningful because they could be related to external events that preceded the testing. Had she only expressed random improvisation or playfulness the data would be useless. But on three occasions she improvised for the completed chimpanzee face what Premack judged to be a hat using inverted cut-outs or banana peel. On all three occasions she had first completed the face before she made the alterations. Importantly, all three occasions had been preceded by a much enjoyed hat fitting in front of a mirror. Her atypical placement of the elements, in combination with turning them backside up, did never occur on other occasions. When it comes to the banana peel Premack (1975) makes the comment that it acted as a symbol used iconically. However, rather than standing for a "hat," the banana skin could have represented a banana skin on the head of a chimpanzee.

As suggested by Premack (1975), completion of motifs might be a promising way of testing productive pictorial competence in subjects who have not developed representational drawing themselves. Because the picture is constructed by the subject itself differentiation should in theory be facilitated and performance in reality mode be unlikely. Likewise, if learning of particular solutions is controlled for, performance in a surface mode would also seem implausible.

III

The bananas are there to make one think, to spur one to the limits of one's thinking. But what must one think? […]

Sultan drags the crates under the bananas, piles them one on top of the other, climbs the tower he has built, and pulls down the bananas. He thinks: Now will he stop punishing me?

The answer is: No. The next day the man hangs a fresh bunch of bananas from the wire but also fills the crates with stones so that they are too heavy to be dragged. One is not supposed to think: Why has he filled the crates with stones? One is supposed to think: How does one use the crates to get the bananas despite the fact that they are filled with stones?

John M. Coetzee (1997, p. 126)

Chapter 11

Methods and assumptions

Had my interest been exclusively to understand the ontogenetic processes behind pictorial competence, studying children would be my first choice. We can know that by some age children are typically fully competent, and we might track their route getting there. Cross-cultural research, in addition, can fill us in on some environmental factors. Studying nonhuman primates in order to ask the same questions seems unnecessarily cumbersome and irrelevant considering that they as species are not very pictorial. However, I do not attempt to write a developmental story for humans (nor any other species). My focus is on the factors involved in pictorial competence. One such factor is the necessary cognitive substrate for pictorial competence from an evolutionary perspective. That is, is a specifically (modern) human brain necessary to understand pictures as pictures?

Comparative cognitive study from an evolutionary perspective can take at least two forms. The first one can be called "cognitive cladistics," from Haun et al. (2006) (see Herrmann et al. (2007) for another recent example). Cladistics is the reconstruction of ancestral relations by grouping e.g. species according to shared features, which is contrasted to groups that does not share the same features. In such a way ancestral trees can be built, where unique features for a given group are said to be derived, while features shared with ancestors are primitive.

In order to find out whether a certain cognitive profile is derived or primitive cognitive cladistics implies finding out which species that can, and which species that cannot perform a certain task in a certain way. In this framework we not only need to pinpoint species differences (or whatever taxonomic level we are focusing on), using good comparative methods, but in order to be more than descriptive on a most general level we need to map out why these differences exist. Little could be said without extensive research in such a framework. Many subjects would be needed too.

The practical consideration of subject availability has a most central role in primate research. All great ape species are endangered animals with complex needs. Working with established zoological institutions is a convenient solution in this regard and can be of mutual advantage. Academia and modern zoos share the goal of advancing the knowledge about animals through research, and to rely such knowledge to the general public. However, relying on zoo collaboration does regulate subject availability and also the type of studies that are possible. This in turn affects the quality of comparison. Finding subjects with comparable histories that respond similar in an experiment, is close to impossible. Focusing on individuals rather than populations is one way of dealing with these constraints.

The equation is not simplified by acknowledging that variation, for example in terms of culture (see e.g. Whiten et al., 2001), is a central part of ape "typicality." Using cladistic reasoning, the same was certainly true for the most recent common ancestral population. Common ancestors were not born common ancestors, but developed into sexually mature common ancestors in complex interactions with physical and mental environments. Even when being flawlessly conducted, cognitive cladistics when used on its own to make evolutionary inferences, is therefore severely limited. By only contrasting averaged groups it can never really exclude that an ability was present, and furthermore relevant, in the life of common ancestors.

The second way of doing comparative cognitive research in an explicitly evolutionary perspective is therefore to focus not on the typical, but on the potential of the ape mind. This can be studied on the level of individuals rather than species, and implies violating assumptions about natural states and ecological validity. Commonly it is about exploring the limits of individuals (without assuming that there is a limit) while keeping a close eye on what makes them learn.

For example, if a handful of scientists had not suspected that apes indeed can learn language, the initial failures with verbal language training, replicated several times between the 1890s and 1950s, might have settled the question prematurely. But after new discoveries we are instead faced with fundamentally different positions, such as: "[i]t seems clear that language is a general system of which human language is a particular, albeit remarkably refined, form" (Premack & Premack, 1972, p. 92). A focus on typicality would probably not foster such conclusions. Many abilities are "discovered" just by looking for them in a better way. But without such ambition in the investigator, animals will always be underestimated (see Thomas, 1996). Performance is for example dependent on feeling well, and sometimes a competence will remain hidden (i.e. a potential) until the individual is fully content with his or her social and physical living conditions (Boysen, 1994). Suddenly they can do new things.

Research on potentiality can answer evolutionary[126] questions on two levels. First of all it paints a picture regarding the possible variation in a trait. Evolution is dependent on variation, but by focusing on typicality one downplays this fact. As a result one can miss the very factors that set off a certain population in a new direction. A second set of questions can be answered by describing those factors in more detail. If one can isolate the factors involved in realising a potential, one can say more about a possible evolutionary scenario. Thus, the more we can discover about animal potential, the more we can ground human abilities in plausible history, rather than just ascribe decontextualised "human uniqueness" to a trait.

Thus, if we focus on potential and study the individual ape rather than species, our questions can dwell more on processes, and explanations, than on descriptives. For sure, the typicality of chimpanzees or the typicality of orangutans is interesting for many reasons, not the least in combination with in-depth study of individual variation, but a single language-using bonobo can potentially teach us much more about language and language evolution than can a hundred "typical" bonobos. A

---

[126] Assuming that the investigated ability implies adaptive consequences.

language-using bonobo can tell us that language is possible for a nonhuman, it provides us with a valuable tool for understanding why this is so, and it can help us guess whether these factors were active also in human prehistory.

For the above reasons comparative cognitive research on pictorial competence in nonhumans is not misdirected, even though true pictorality seems to be a learned, cultural, ability. On the contrary, it should not matter who learns it. Apes are rather prime subjects, because there seem to be individuals that can, and individuals that cannot, decode pictures in a pictorial mode.

The questions that constitute the starting point for my empirical work were evolutionary in the sense that I wanted to know whether nonhumans can learn to understand pictures as pictures, and explorative and developmental in that I wanted to see which forms this competence can take, and why.

Two lines of original empirical investigation will be presented in this part of the thesis. The first one was an attempt to teach experimentally and picture-naïve zoo gorillas a method for showing their baseline pictorial ability, with the intent of possibly taking this ability a step further. The next study was a series of tests on symbol competent bonobos, who had extensive experience in using pictures in testing situations and in everyday communication, but who had never been tested on non-photographic stimuli before. The gorilla study took place at Givskud Zoo in Denmark across 11 weeks between November 2003 and March 2006, and the Bonobo study at the Great Ape Trust of Iowa, USA, in the summer of 2006.

11.1 Categorisation

The main method I have been working with is matching-to-sample (MTS). I have already written at length about MTS in the context of the review of studies in Part II, but a further look at the assumptions underlying the method and the factors involved in the emergence of a matching competence is needed. We will also find that some versions of MTS ties in with pictorial competence in significant ways in that they are concerned with categorisation based on more or less explicit similarity judgements.

Recognition always implies recognition *as* something. The scope of this "as" is what concerns us here. Expecting animals to recognise pictorial instances of objects, especially when these diverge significantly from the standard exemplars, presumes that animals categorise their world in a stable but flexible way, where e.g. an apple remains an apple even when a bit different.

Two apples are both apples not only because they look similar, but because they share many other properties as well. As a result, two apples often have an interchangeable status in the world. However, this is only one tentative definition of an "apple" concept. Defining what a category or a concept is is not straightforward. It is not simplified by the fact that the terms "concept," "category" and "stimulus equivalence" are used interchangeably in the literature (Thompson & Oden, 2000). To this list can be added "class" and "classification."

To take an example of a definition of "concept," Hayes and Nissen (1971) suggest: "consistent response to a constant aspect of a variety of stimuli regardless of the specific context in which this aspect occurs" (p. 78). They continue to posit that "concept formation [is] a process of discrimination in which the animal abstracts the 'constant aspect'." This quote was chosen because it puts emphasis on an aspect of concepts that are relevant for pictorial competence. Regardless of what concepts really are, *stable and retained* principles for grouping are necessary for the interpretation of pictures.[127] Introspectively, my "apple" concept remains unchanged although many other things in my inner and outer world change between my uses of it. It is necessary to have access (this time I do not mean introspectively) to such independent concepts in order to have something to relate a pictorial display to.

Since a pictorial display is static it is not possible to gain more information from a picture than is actually there, without filling in. What is filled in naturally stems from one's knowledge about the objects that one thinks that one perceives in the picture. All that I can fill in regarding apples stems from my "apple" category. This is the reason cross-modal discrimination tasks have been used to ascribe mental representations to animals (e.g. Davenport & Rogers, 1970; see section 5.2). In a cross-modal matching one must add something from one's e.g. "apple" concept to bridge two experiences. The exclusive haptic feel of an apple can only be matched to a visual display of an apple, rather than to that of an apple-like orange, if there is access to a mental hypothesis of some sort based on one's "apple experience."

Being able to retain concepts across contexts is not the only requirement for pictorial competence. It is equally crucial that processes of categorisation continue even though a concept can be said to have been formed.[128] Having extendable categories, i.e. being able to include more and more variations on a theme, including abstracted instances, is a necessity for an open-ended pictorial competence that can move beyond the most life-like photographs. A significant factor for expanding one's recognition of pictorial exemplars is in other words that one can learn from pictorial experience. Such learning is arguably tied to in concepts that primarily relate to the real world. My network of picture-apple experience needs to tie closely to my network of apple experience if we are to argue that pictures of apples are about apples at all.

Do animals have stable but open-ended concepts? Most animal seem to parse their world in entities that appear stable over contexts. Both natural and artificial stimuli, well-known and novel, can be recognised by most animals in a variety of two-dimensional orientations and three-dimensional rotations (see Jitsumori & Delius, 2001). This type of consistency seems to be built into object perception.

Few would argue against the fact that there is much to gain from grouping such stable entities according to shared properties. This allows for effective learning and handling of a dynamic world, rather than endlessly learning new instances. It seems that most animals, and certainly primates, recognise certain groups of things as

---

[127] I see no need here to define a possible source and nature of these proposed "principles."

[128] Unfortunately one cannot experimentally entangle concept formation from "concept possession," since concepts can always be argued to have formed as a response to the task that was designed to test it (Hayes & Nissen, 1971).

members of the same category. Recognition of e.g. predators, edible foods, nests, conspecifics etc. would otherwise not be possible (e.g. Spinozzi, 1996; Jitsumori & Delius, 2001).

But it might be premature to apply the term categorisation to all forms of grouping. Herrnstein (1990) lists five levels of categorisation and concludes, following a modest review of the research, that nonhuman animals reach at least the fourth level, concepts, and in some circumstances the fifth, abstract relations. The full list reads: discrimination, categorisation by rote, open-ended categories, concepts, and abstract relations. Abstract relations are seen as the hallmark of human intelligence.

Premack (1976) makes the important distinction between *discrimination*, which more or less all animals with sense organs should be capable of, and *classification*, which he reserves for an open-ended ability to group all types of stimuli. A dog might for example be able to discriminate between classes of dogs (e.g. young and old dogs) but not be able to attend to any basis for sorting cutlery, although there are a multitude of possible dimensions to sort by.

But categorisation is not without constraints. It is grounded in attentional predispositions. The dog example can be seen as just a strong case of this. Some dimensions for categorisation are salient and transparent for given individuals, others less so. While I do not pay much attention to colour in my categorisation of apples, a connoisseur might. This does not reflect a general and open-ended ability to categorise, but an ability to categorise based on my life with apples.

"Open-ended ability" might in turn be limited to different types of relationships between entities. For example, further distinction is needed between grouping according to properties that can be experienced directly in objects, grouping according to learned relationships, and grouping based on inferred relationships. One can be "open-ended" when it comes to visual properties, but severely limited in grouping based on e.g. function.

Most studies have indeed focused on visual properties of objects, such as shape and colour, but Tanaka (e.g. 2006) has tackled also the associative case by showing that chimpanzees can group objects according to "thematic relations." These are learned relations between objects that usually go to together, such as a jar and its lid, hammer and nails, etc. However, the nature of such relationships do not need to go beyond statistic co-occurence. Even plain familiarity of two items fosters spontaneous grouping at the expense of a novel object (Tanaka, 1995). Tanaka (2006) draws the line for chimpanzee categorisation at "thematic categories" rather than abstract, e.g. functional, ones. However, symbol training can greatly improve the latter type (Tanaka, 1997). Grouping according to function took for example also place in Savage-Rumbaugh et al. (1980).

Vonk and Povinelli (2006), in their "unobservability hypothesis" for human uniqueness, suggest that even more abstract categorisations, such as "things that make good gifts," [129] entails reasoning that depends on factors that cannot be perceived in the stimuli at all, and is therefore beyond ape minds altogether. On a pic-

---

[129] This reasoning pertains to grouping. If needing to give someone a gift the ape might very well select the most appropriate object for the occasion. (If the ape then proceeds to giving another gift, or two at a time, has it not employed the categorisation "things that make good gifts"?)

torial note, the ability to read dynamic content into static pictures would potentially be a promising test of the unobservability hypothesis.

Then there is the question of how well categories are formed in response to contrasting categories. This can take place on different "levels of abstraction" (see e.g. Vonk & Povinelli, 2006). This notion was introduced in section 7.3 with the now classic "kingfisher-experiment" of Roberts and Mazmanian (1988), where it was found that pigeons and squirrel monkeys found it easier to pinpoint the kingfisher from other species of birds, than birds from other animals. The latter, the "intermediate level," was found to be the most difficult also for a gorilla in Vonk and MacDonald (2002).

Most of the time pictures depict objects in the same way as they appear in the world, i.e. open to visual categorisation based on physical similarity. Any further ascription to more "abstract" categories does not pertain to the picture as such. In other words, in order to make sense of pictorial displays a primate does not need to learn categorisation in a different way than it already applies to the everyday world. That said, in manually constructed pictures a kingfisher might be drawn more like a bird than a kingfisher. An gorilla must therefore also be able to recognise birds in general, rather than discrete bird species, if it is to recognise the kingfisher in the picture.

Categorisation in primates has been tested in three main paradigms: spontaneous sorting tasks, similarity – difference judgements, and matching tasks.

11.2 Sorting

Viki the chimpanzee showed a spontaneous inclination to sort objects according to shape and colour at the age of 3.5 years. She was later, at 5.5 years, subjected to a formal test containing 13 groups of objects. Viki was given a mixed collection of two types of objects, usually 10 or 5 items of each type, and without any probing or experimenter-given sample spontaneously sorted these into their respective categories. With these objects she made only 20 inconsistent placements among 600 (Hayes & Nissen, 1971). She was by no means limited to two categories. In an informal session Viki sorted six types of objects simultaneously, without any error.

If Viki simultaneously sorted into more than one category (e.g. blue and red objects into two piles, rather than singling out only the red ones) already when 3.5 years old, her sorting behaviour seems to have maturated quickly. The testing and findings at 5.5 years, however, is in accordance with the findings of Spinozzi (1993). Among the five laboratory chimpanzees tested by her, only one spontaneously formed two consistent classes across a range of conditions. This behaviour emerged in the fifth year of life. This is an important distinction because only when there are more than one group can we say that the dimension colour and not e.g. the colour red was the basis for sorting. Only in the former case can a comparison be said to have taken place; A comparison with respect to colour, instead of just picking out the red ones.

The items in the above studies could all be sorted on surface features, such as colour, shape or material. Hayes and Nissen (1971) therefore gave Viki a second set of tests. This time objects could either be sorted according to function, or by surface features. They wanted to see what classes she preferred to use. It was found that she did indeed prefer to sort by material and properties, such as buttons and coins with holes in them versus buttons and coins without holes in them, rather than according to the use of the objects. She knew perfectly well how to use them in appropriate contexts, but it is fully possible that the function of the objects lost its salience when the context was a sorting task. Or it could be that she just *preferred* to sort by looks.

A most relevant finding was that when given groups of objects that could be sorted according to different surface dimensions Viki could switch between these when given the objects a second time. This suggests that she sorted in a mindful way, according to explicit and volitional criteria.[130] By being able to select specific features of the items as basis for her categorisation, such as colour, Viki showed evidence of "abstraction" (Hayes & Nissen, 1971). Also Premack (1976), when subjecting two of his chimpanzees to sorting tasks, found that they sorted spontaneously according to single properties when several were available, and that one of the chimpanzees, Peony, often switched between criteria on her own accord. Abstracting up to three properties simultaneously was also easily done by Sarah in the same laboratory, who could identify a specific object in an array of objects by symbol-mediated conjunctions such as "red," "large," "round" (Premack, 1976).

Premack's (1976) chimpanzees Peony and Elizabeth seemed to be able to sort not only according to discrete properties of objects but according to their association to a superordinate category as well. When given plant material versus animal material they sorted these in the respective two classes at a level above 80% correct. Given the vast difference in object types this sorting could not have taken place based on surface properties. When contrasted to office material they switched to sorting animate versus inanimate materials. However, it is not reported whether this switch occurred spontaneously or whether samples were provided by the experimenters.

That e.g. Viki and Peony could themselves[131] choose the basis for similarity between items is highly relevant for pictorial competence, especially when it comes to interpreting secondary iconic signs (see e.g. chapter 4). In one way or the other all things are similar to all other things, so the process of pinpointing specific similarities needs to be guided. To have some voluntary input into this process arguably broadens the type of similarities that can be appreciated. That said, it is different to choose a basis for similarity between things that are present, and between things where one counterpart is in the mind alone, or even not yet invoked in mind. Perhaps that is why secondary iconic signs need a "key" (e.g. Sonesson, 1989), i.e. a label, to be appreciated (*fig. 17,* p. 206). A label is a proposition given externally to the icon, against which the pictorial display can be tested. What the label does is to bring forth the comparison object and make it explicit so that similarity can be estimated. When a pictorial display is immediately recognised through primary iconicity, on the other hand, the comparison object has been evoked by the picture itself.

---

[130] Or she had a flimsy mind that only happened to blindly lock into different criteria each time.

[131] This is an assumption.

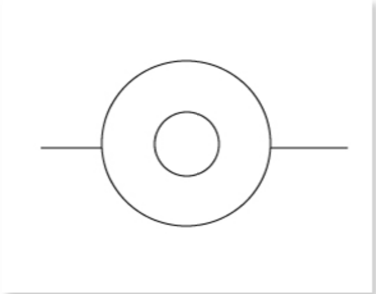


Figure 17. Try to interpret this picture without reading the following text: *A Mexican on a bicycle (seen from above).*

As a curiosity, both with regards to Viki and in relation to this thesis, Hayes and Nissen (1971) gave Viki a task where she was required to sort photographs of animals versus humans. A picture of Viki herself was included as the last item in the deck.[132] The experimenter defined the categories by placing one animal and one human photograph as samples in advance. Viki followed by sorting effortlessly and correctly. Her only reluctance regarded a fish, which she settled to place among the animals. With the same certainty that she sorted her chimpanzee biological father as an animal, Viki then proceeded to place her own photograph with the humans. This was the only "misplaced" photograph out of 40.

This episode is included without any specific claims in Hayes and Nissen (1971), but Viki is perhaps not alone in recognising herself in photographs. The gorilla Koko seems to readily comment photographs of herself in sign language, as in an example given in Patterson and Linden (1981): Koko signs "Me love happy Koko there" in response to photographs of her birthday party. The reason I say "readily" is that no comment is made in the text about this being an unusual event. However, the reason for why this statement of Koko's has been chosen for presentation is not given either, it is only said to be a typical statement for that age (7 years). Nim in Terrace (1980) is also said to have learned the sign for his own name by way of mirrors and photographs. The question in all these cases is whether the correspondence between the individual they learn to name in the picture and themselves is perceived.

Savage-Rumbaugh et al. (1980; Savage-Rumbaugh, 1986) showed that the chimpanzees Sherman and Austin could sort objects according to the experimenter defined categories "tool" and "food." However, extensive training was needed for them to sort according to these criteria. Importantly, the tool-function of tools had to be continuously pointed out to the subjects by letting them use the tools before naming/sorting them as either tools or food. To abstract this particular aspect of the objects they thus needed guidance. However, with time they could sort without first using the objects, including tools that had not been named previously. Thus, the "toolness" of tools seemed to have been available in the conceptualisation of the objects from the start, but had not held a salient position for categorisation *in this particular task*. Attention to this shared aspect of the tools had to be brought out by actually using them.

A different pattern was shown for the third chimpanzee in the test, Lana, who could sort tools and foods from the start in respective sorting bin, but was unable to ascribe the lexigrams "tool" or "food" to them. The subjects' differences probably stemmed from their different lexigram training. Sherman and Austin learned the

---

[132] Self-recognition would not surprise since Viki readily used mirrors to inspect herself (Hayes, 1951).

names of specific tools in a paradigm where they requested tools, or were requested to hand over tools, in order to solve a problem. However, they could not as easily name tools with lexigrams when these were just held up to them. Lana, on the other hand, had predominantly learned lexigram names of tools by naming them without using them. She could in turn not request tools with lexigrams when she needed them in a task, although she was able to make requests by pointing to specific objects that she needed.

It can be argued that the chimpanzees did not apply a proper "tool" concept in this study. They can for example have been sorting according to a food versus non-food distinction. But it is not transparent why Sherman and Austin were then helped by using the tools as tools. The concept seems to have been close enough to the intended one for allowing sorting of all kinds of novel tools and foods. Only a single systematic misplacement took place. Sherman consistently sorted a sponge as food which he seemed to have based on his fondness for eating pieces of sponge. From this simple example we can learn that failure to categorise according to criteria does not imply absence of the proper categories. Ambiguous choices must therefore be avoided in an experimental designs. This is important also when testing pictorial competence. The ape mind interacts with a pictorial display according to its conceptual system, which is often bound to differ from the experimenter's.

After the above study, and some further training, Sherman extended his sorting to four categories: "food," "drink," "locations" and "tools." In a test of 11 novel items (presumably lexigrams or photographs were used for "locations") he only made one error, sorting milk as "food," which he continued to do despite correction (Savage-Rumbaugh, 1986). Sherman and Austin later also sorted photographs in groups of "cars", "animals," and "people" without specific training. Following training they also applied their sorting ability to sort objects according to simple features like size, shape, colour, texture, and form (Savage-Rumbaugh, 1986). However, no data is given on whether they could spontaneously switch sorting criteria like e.g. Viki could.

Miles (1990) relates that the sign-language trained orangutan Chantek "[...] completed a number of complex sorting tasks that required an understanding of classification, classes and categories" (p. 536), but no further details are given. Surely, these tests did not require an *understanding* of classification, classes and categories. A desire to group things would be enough.

As will be evident below, language training as such does not seem to be an important factor for "complex" categorisation, but some very specific drills seem to be necessary.

11.3 Same – different judgements

For some authors the ability to judge categorical identity between new exemplars in a MTS setup suffices to prove an abstract notion of "sameness" (Spinozzi, 1996). Others grant similarity concepts, of some sort, to nonhumans but reserve the notion of an *abstract* similarity concept to when relations between relations are being judged, i.e. analogical reasoning (e.g. Premack 1983). Still others grant neither similarity concepts nor true matching competence to animals, but instead ascribe their performance to conditional discrimination (Sidman et al., 1982).

Given that pictures reserve visual information to some extent, an analysis of a pictorial display can come a long way with categorisation based on just physical resemblance. I will argue, however, that seeing pictures as pictures requires the more abstract ability as well, i.e. a concept of similarity. But rather than similarity I would like to call it a concept of comparison, of which similarity is a natural part. The point being that it is not an appreciation of similarity which is the crucial step but the comparison. The ability to abstract properties for comparison was mentioned, and furthermore the ability to control, to some extent, the choice of which properties to compare (e.g. sorting with respect to the colour dimension). This would constitute a concept of comparison in action. Comparison is the key component, both for understanding a referential function of pictures, and for perceiving a similarity which is not directly given, as is the case for non-naturalistic stimuli (per definition). I will also suggest that some form of analogical reasoning, defined as judging relations between relations by e.g. Premack (1983), is synonymous with a central part of pictorial competence, which is that of decoding complex scenes containing multiple objects that cannot be fully appreciated in a reality mode of picture processing. We will come back to this point below.

The easiest method to assess whether animals can respond to classes is the standard discrimination task where the animal is reinforced for choosing one category but not another. Positive transfer to novel items means that some form of categorical response is given. However, as have been pointed out repeatedly in Part II when it comes to pictorial recognition, the basis for a category membership is not controlled for in most discrimination tasks. That is, one does not know what information the subject uses for its discrimination: local properties or object recognition. The standard case in point is D'Amato & van Sant (1988) where capuchin monkeys in fact categorised colour rather than the concepts intended by the experimenters. Using the criteria of e.g. Premack (1973) such performance would not even count as categorisation. It does not matter if an animal is able to judge sameness or difference between stimuli on a *purely* visual basis, i.e. in a surface mode. Here we are concerned with conceptual similarity, e.g. two versions of an "apple" (or whatever level of abstraction is meaningful for the animal in question).

The basis for comparison is a concern also for MTS procedures, but MTS has an advantage in being open-ended through its use of a sample and comparison stimuli which can be combined in various constellations to rule out alternative sources for matching. In the standard discrimination task each new target category has to be drilled to criteria before it can be tested in transfer trials.

In a relational version of the standard discrimination task, what can be called the same – different discrimination task, the subject is taught to respond in one way when a display of two or more items are the same (either identical or according to category), and to respond in a second way when the items in the display differ. Usually the "same" or "different" response is made by indicating one of two simultaneously presented choice items, such as differently coloured cards. The subject has to learn the meaning of respective choice item (same/different with respect to X) by performing a number of problems (usually in the hundreds), as well as demonstrate generalisation by making judgements of completely novel comparisons. This circular way of testing "same" and "different" concepts is thus open to the critique of possibly forming the concepts that one is set to test. However, it seems that such judgements may be intuitive at least to chimpanzees.

Oden et al. (1990) demonstrated that 9 – 11 months old chimpanzees, measured in a preferential looking task, looked at novel objects more than at familiar objects. They also looked more at novel displays of object pairs that were either two identical objects, or two different objects, if this relation differed from one from that were displayed in a familiarisation phase (Thompson & Oden, 1995). It was thus judged that the infants perceived similarity or difference also in relations *between* objects. However, they could not apply this ability in a MTS task when shown a pair of similar or dissimilar objects as sample and then asked to match to the equivalent relation among the choice items (Oden et al., 1990). However, the same subjects could perform well in ordinary identity MTS (Oden et al., 1998).

As predicted and shown by Premack (1983) the judgement of relations between relations could only be made explicit by way of mediating labels for the concepts "same" and "different." His prime subject, Sarah, could solve complex analogical problems by bridging them with her token for "same." Not only could she match objects that shared relations with other objects, she could also compare quantities of liquids and masses across objects (Premack & Premack, 1983). Furthermore, Sarah was able not only to *respond* to object displays in an analogical fashion, but also spontaneously and accurately arrange objects in configurations based on analogies (Oden et al., 2001).

That experience of token-use was essential to appreciate abstract relations (relations between relations) between entities was believed to be confirmed when Thompson et al (1997) subjected three of the subjects in e.g. Brown and Boysen (2000) to a test of judging relations between relations. Boysen's chimpanzees used printed symbols for "same" and "different" to indicate the relationships between displays of animal photographs. These subjects were not specifically language-trained, which Premack's analogical chimpanzee Sarah is argued to be. However, the specific labels "same" and "different" were not available as tokens in Thompson et al. (1997), but the subjects were instead required to match photographs of object pairs to real object pairs that displayed equivalent visual relationships. But label training had occurred in other domains, such as using numbers to judge quantities. Such training proved to be all that was needed for the three chimpanzees to spontaneously and without differential reinforcement apply the concepts "same" and "different" in their appreciation of relations between relations (Thompson et al., 1997). It was suggested that such token-use made the subjects pay attention to abstract relationships between entities. That is, the chimpanzees were drilled in appreciating a *stand-for* relation, at least when it came to numbers. It seemed to be the common denominator between the chimpanzees in the Boysen and Premack laboratories. Sarah, who was included as a fourth subject in the above study, performed as well as ever, while a fifth subject, who served as control and had not received any numerical token training, performed at chance level.

However it appears, after all, that not even a history of label training is necessary to match on the level of relations between relations. Smith et al. (1975) trained two experimentally naïve adult chimpanzees to successfully match coloured geometrical plywood shapes according to sameness or difference. That is, two items were given as a sample and based on the relation between these the subjects had to choose one of two sets of items that displayed the same relation. Like in Thompson et al. (1997), the chimpanzees succeeded both when correct comparison sets were identical with the sample set, and when only the relation was identical between sets. That said, it is clear that they did not perform well spontaneously, like the chimpanzees in Thompson et al. (1997) did, but required extensive training. Unfortunately, data on training is incomplete.

Using delayed MTS (DMTS) Vonk (2003) has a more detailed account of the same type of relations-between-relations-matching by token naïve subjects. This time four orangutans and one gorilla were required to match geometric shapes that shared relations with other geometric shapes on touchscreen. Items could be similar on shape and/or colour. The subjects had not matched abstract shapes before but were trained in DMTS with photographs. Matching based on relations between relations had not been tested previously. After training to proficient matching by all subjects on simple DMTS for the novel stimuli, four of the five subjects learned to match on a relational level. That is, choosing the two shapes that displayed the same relation as the two shapes in the sample. Transfer seemed to have occurred relatively effortlessly (24-60 trials). One of the subjects received the conditions in the reversed order but did not show slower learning of relational DMTS, which means that physical and abstract matching can be equally easy.[133]

It has been proposed that there is a difference between monkey and great ape categorisation in that monkeys predominantly use physical resemblance while apes are capable of both concrete and abstract relations between stimuli (Spinozzi, 1996; Thompson & Oden, 2000). But this view has been challenged by Bovet and Vauclair (2001) who have shown that olive baboons can make same – different judgements for both physical and conceptual identity, using real objects and foods. However, analogical reasoning in the sense of being able to match half an apple to half a glass of water, and a whole apple to a full glass of water (Premack & Premack, 1983) remains to be reported.

There are arguments that posit that language, or symbol training, significantly restructures the concepts in animals (e.g. Premack, 1973; 1983). One such example is the possible new use of similarity judgements when a label is provided to bridge two items in comparison. However, the real question is whether the label only adds an interface between the intentions of the experimenter and the subject who is to understand the task, or whether labels really reshape connections between items and their concepts.

---

[133] The gorilla in Vonk (2003) was 4 years old at the time of testing, which lends credibility to the claim that the chimpanzee Viki (Hayes & Nissen, 1971) could abstract properties and sort objects accordingly, well before the age of 5 when she was formally tested.

Premack (1983) proposes that there are two codes in the human mind: an “imaginal” one and an “abstract” one. Ape minds can only solve e.g. MTS tasks based on physical similarity because that is all they can visualise in their imaginal code. By acquiring abstract code on the other hand, ape minds can approximate human minds when it comes to certain stimulus relations. (Whether Sarah’s abstract code was in her mind or in her plastic tokens is a further question.)

Premack did not intend representation of discrete objects when he suggested a difference between symbol-trained and symbol-naïve subjects, but in their ability to represent abstract relations. However, pictorial displays, when viewed as pictures and not a case of reality, are abstract relations in the sense that their correspondence to reality is a propositional “idea.”

A tentative suggestion is then that comparing the understanding of pictures can be a way of comparing the concepts in language and non language-competent apes. Without a specific apple concept for example, matching all types of drawings of apples to real apples or photographs of apples would not be possible (if surface mode is controlled for). With an apple concept, however, and the necessary pictorial competence, recognising also “non-apples” as apples would be possible. Note that I do not imply that apes do not have apple concepts, only that without “abstract code” the concept possibly cannot be applied, i.e. be made explicit, in a matching task with ambiguous stimuli. Having access to the label “apple,” however it is realised, has a significant advantage in such a situation.

Furthermore, relations between relations have a real counterpart in pictorial competence. As also hinted at above, analogical reasoning would be necessary for resemanticisation in a scene where objects need other objects in order to be identified. Take a drawing for example, where perhaps a box, itself bears no likeness to a television. But the same box on a table in front of a sofa where someone is sitting and looks at the box is likely to be a television because the same configuration in real life is likely to include a television. In this sense a relation between relations is involved in that a pattern, or rule, gained from one instance is applied to a second instance, but not in a helter-skelter way. There is a relation, i.e. a sign relation, between the two instances, which implies an act of comparison.

11.3.1 “Similar”

Many MTS setups implicitly makes use of a perceptual “similarity” judgement rather than a perceptual “sameness,” or identity judgement. (On a conceptual level identity can of course prevail despite perceptual differences.) Matching e.g. one apple to another apple always implies *some* difference in the two instances of apple. Sometimes this difference is large, sometimes it is hard to even notice. To be aware of this difference, but match despite it, is often part of matching on a conceptual level. To experimentally assess whether such an appreciation can be explicit, however, requires that “similar” and “same” has to be pitted against each other. Premack et al. (1978; see also Premack & Premack, 1983) did just that in a “same/different/similar” test with Sarah the chimpanzee. Comparison stimuli were produced by tracing toys on paper and then colour these in different colours.[134] Sarah's task was then to indicate with her tokens whether two shapes were the "same," "different," or "similar." However, only two of the three choices were made available on each individual trial. The similarity of choice items was varied by changing either size, colour or shape. To instead make the items different from each other all three properties were dissimilar. "Similar" was new in Sarah's vocabulary and the test was therefore also a test of acquisition. As expected, Sarah used "similar" randomly at first, but learned quickly. In later part of the experiment she used "similar" equally successful as "same." But proper "different" judgements dropped below chance on some of the variations of the test.

Remember that only two of the three choices were available on each trial. Variation in performance depended on contrasted choice item. Pinpointing "same" was the easiest when "different" was the contrast, and "different" when "same" was the contrast. That is, both "same" and "different" judgements were made more complicated when "similar" was the contrast (74-78% correct irrespective of contrast). Sarah had thus learned a new token, "similar," and used it in a unique and accurate way.

When it comes to perceiving pictures, in a strictly pictorial mode, a picture is never the "same" to a referent, but always "similar" by virtue of being differentiated from reality. In reality mode, on the other hand, a picture and an eventual "referent" are both same and similar to each other in the same manner as two real-world items can be same or similar.

11.4 Matching

If apes have concepts and are able to apply comparison and perceive similarity, when it comes to realising MTS the next problem is to apply these abilities in a matching task. Ever since the days of the study of chimpanzee colour discrimination by Ladygina-Kohts (Yerkes & Petrunkevitch, 1925) Match-to-sample, or "choice from sample" as it was then called, or "sample-matching" (Yerkes, 1943/1945), has been a favourite empirical method in both comparative and developmental psychology. This is a bit surprising since matching procedures requires a lot of preparatory drilling of the subjects.

Generally, learning MTS takes time and sometimes it fails altogether. However, surprisingly, few examples are necessary for training it. Nissen et al. (1948, in Oden et al., 1988), as well as Oden et al. (1988) only needed two training stimuli to teach matching to chimpanzees. The seven chimpanzees in Nissen et al. (1948, in Harlow, 1951) needed between 177 and 547 trials in a physical MTS where the sample was presented in-between the two choice objects. To reach criteria for stable performance the infant chimpanzees in Oden et al. (1988) needed between 642 and 1 002 trials. After this initial drilling, matching transferred effortlessly to new stimuli of very different natures. That few training stimuli suffice, and more do not facilitate

---

[134] This test cannot count as a test of pictorial competence since identification of the toy shapes was not necessary.

learning, implies that it is just a question of "getting the rules" of the game, i.e. specifically apply matching, or the comparison concept mentioned above. This is in stark contrast with monkeys (Katz et al., 2002; Thompson & Oden, 2000) and pigeons (Katz & Wright, 2006), who benefits from a large pool of stimuli, and seem to instead narrow down a general rule from many exemplars. Something very different seems to be going on in the ape case. However, it is possible that apes are sometimes steered towards the "pigeon approach" to the problem, which results in laborious training, sometimes with limited success. Many routes can lead to a similar ability. One notable reason why Oden et al. (1988) succeeded with ease might for example be that they used three dimensional objects for training, and not pictures or the like.[135]

All stimuli are not equally easy to match. Daehler et al. (1979) found striking differences in 20 to 32 month olds' ability to match different media. Object – object matching was significantly easier than photograph – photograph, object – photograph, or photograph – object matching. There was no difference between the three latter conditions. Thompson and Oden (1988) had trained their subjects with objects. Premack and Premack (1983) note that "[c]himpanzees have extreme difficulty with photograph - object matching [… and] may be five to six years old before successfully matching photographs and objects (such as a real shoe with its exact photographic replica, etc.), whereas an eighteen-month old child can pass this test even though it has never previously been exposed to pictures." (p. 101). One would presume that Premack and Premack include all chimpanzees in this statement. They then mention that on initial trials the chimpanzee will often place photographs together and objects together rather than to match across media. This is an interesting observation since it hints to the cause of the matching difficulties. It is not matching per se that is difficult but to ascribe content to the paper surface, either in reality or a pictorial mode. Both modes would suffice in such a task. The chimpanzee has probably too much experience of photographs as pieces of paper, and too little experience of occasions where its depicted contents have been informative and useful. Attention seems to be tuned to the wrong properties. Had the task been done for example on a computer touchscreen the confounding paper properties would possibly have been eliminated. Premack and Premack (1983) then claim that human children would not group pictures and objects in this way, but their only reference to a picture experiment with children is Hochberg et al. (1962), which is not a test of object - picture matching but a naming experiment. From this Premack and Premack seem to conclude that a child that is 1.5 years old would perform well on a range of pictorial tasks.

A notable difference from previous variations on the MTS theme is that language-trained apes need very little, or none at all, training in matching procedures. Matching competence furthermore seems to be a general ability that transfers easily across modalities and contexts (Savage-Rumbaugh et al., 1988). This ability should not be confounded with symbolic competence per se, however. Savage-Rumbaugh et al (1988) suggest that non-symbolic aspects of training, such as learning to orient one's attention, look for similarities between objects etc., can facilitate comprehension of

---

[135] I believe the situation in Nissen (1948) was similar.

matching tasks. This training is heavily intertwined with the use of symbols. This conclusion is supported by the ease by which Sarah in Premack (1976; Premack & Premack, 1983) seems to have learned basic MTS *before* she had acquired her plastic tokens for "same" and "different." Furthermore, she trained on only two sets of objects (cups and spoons) but generalised with apparent ease to novel ones. That said, the other chimpanzees in the same research needed much more training before they could match (Premack, 1976).

In addition, there seem to be limits in the matching abilities of also symbol-using individuals. Dugdale and Lowe (2000) failed to display symmetry in Sherman and Lana's matching abilities. (Austin had to be dropped because he did not learn to match at all.) The others needed a fading program to learn to arbitrarily match colours to designated shapes. The shapes had to be placed on a coloured background of the same type as the sample colour and then successively be reduced to a small square behind the shape, before it could be removed altogether. The chimpanzees required 1104 to 1440 trials to reach criterion, which is similar to 2-year old humans using the same fading procedure. Then, when required to instead match the shapes to colours rather than colours to shapes, neither of the two subjects transferred successfully. A display of symmetry would have been indicative of an "equivalence relation" which in e.g. Sidman et al.'s (1982) view would be evidence for true matching.[136] Instead of matching the chimpanzees had displayed conditional discrimination of an "if-then" sort.

That Sidman and colleagues reserve matching competence for equivalence relations is problematic when studied with stimuli that is not arbitrarily matched but based on similarity. As argued previously the connection between pictures and entities in the real world is not symmetric, because similarity between two things is not symmetric (e.g. Rosch, 1975; Tversky, 1977). This can be argued to be especially true for pictures, where real entities are more prominent than their flat pictorial counterparts (Sonesson, 1989). What is less prominent more easily stands for that which is more prominent, and it would not be surprising if a difference is found between object-picture and picture-object matching in some subjects. Given, of course, that matching does not take place in a reality mode, where pictures can gain prominence by being perceived as real.

Sherman and Lana need not be worried about their performance. Applying MTS in new situations can be tricky even for experienced subjects. The chimpanzee Popo in Tanaka (1996) was used to cognitive testing using pictures, and was proficient in matching objects to objects. But when required to match objects to their photographic counterparts she not only failed to reach criterion in training, she did not improve at all in the 600 trials she endured before being dropped from the experiment. Popo's inability to perform well also seemed to affect her motivation to participate. Her comparison subject, Pan, had learnt a similar type of matching in less than 200 trials.

The reasons for why MTS learnability varies can be numerous, of which experimenter pedagogy and personal inclinations of the student are probably major fac-

---

[136] Equivalence relations are commonly characterised by reflexivity, symmetry and transitivity. All three requirements must be fulfilled.

tors. This is evident in the different successes in different laboratories. Furthermore, there are many factors that the experimenter cannot affect stemming from the simple fact that the animal in question has its own mind. That trials are not unique and that animals integrate information gained across problems has been known at least since learning sets were discovered by Harlow (1949). The learning context for MTS is exceedingly complex, with many factors outside of the experimenter's control. This probably accounts for the vast number of trials usually needed for subject to discover the application of a matching ability in a specific experiment.

Once matching has been acquired it can take place on many levels, of which feature, identity and conceptual matching comes to mind. Feature matching is matching grounded in shared properties, e.g. matching two displays that contain the colour red. Identity matching is the matching of two identical displays, regardless of whether one ascribes a category membership to the displayed entities or not. These two cases are visual matching of a local and global type. "True" conceptual matching on the other hand, i.e. matching two different apples or two different fruits (as opposed to e.g. furniture), should be based on knowledge that transcends surface similarity. As argued above, to investigate pictorial competence in a matching paradigm it is required that the subject is able to match on a conceptual level.

That primates can match on a conceptual level has been proven in several of the experiments reviewed in Part II. But we have also seen examples where primate concepts have clashed with human intentions. One example is where the chimpanzee Sherman in Savage-Rumbaugh et al. (1980) insisted on sorting sponges under the lexigram category "food" rather than "tool." Another lesson can be added from Premack (1976) where Sarah could only match pieces of fruit to whole fruits after she had experienced the preparation of fruit. In the beginning of her training she apparently had never seen apples or bananas being chopped up into the neat wedges she was used to. The meaning of her plastic tokens spontaneously followed this development in her conceptualisation to include both whole fruits and parts of fruits.

Choosing stimuli and their configurations into sample, matches and non-matches therefore requires great care. Also, without the right control matches, a whole experiment might turn out useless because the experimenter cannot be sure what the subject really matched.

Now that we know that MTS is both a promising and a challenging method, let us turn to the empirical work conducted at Givskud Zoo.

Chapter 12

The Givskud studies

In order to test what type of pictures picture-naïve apes can match to objects, and what they can learn in this area, an explorative study with zoo gorillas commenced in November 2003 and totalled 10 weeks until April 2005. An additional series of opportunistic sessions took place during a week devoted to other research in March 2006. The studies were conducted at Givskud Zoo in Denmark. The idea was to train an open-ended method for testing pictorial stimuli. Three types of experiments were conducted: Matching-to-sample tasks, simple discrimination tasks, and object-choice tasks.

All testing took place during winter months when the zoo was closed to the public and the gorillas were held in their indoor facilities. Hopefully they found some diversion in the research activities. Boredom is a major concern in captive environments. Providing environmental enrichment decreases stereotypical behaviours and is argued to increase well-being in zoo animals (Shyne, 2006; Swaisgood & Shepherdson, 2005).[137]

Animal research is ideally made as non-invasive as possible. The gorillas were for example not used to being isolated for experimental purposes. All testing therefore took place in the social group without separating the individuals. Participation was voluntary and subjects could come and go freely to the area of testing (*fig. 18,* p. 218). They were not coaxed to participate. Testing took place between 10 a.m. and 15 p.m., five days a week, one week at a time. Short breaks were taken by the experimenter (i.e. the author) when necessary, or the gorillas' interest in being tested was low. An extended break was also taken for lunch each day. Since a second subject almost always became interested when a previous one became bored, testing was more or less continuous throughout the day. Subjects were not food or water deprived at any time during the studies.

---

[137] The potentially mutual benefit of zoo – academia collaboration from the animals' perspective was recognised already by Yerkes (1951): "[Such studies] may even be to the great advantage of the animals because of the diversion, exercise, and novel situations provided by the scientist. Thus, incidentally, the investigator may become an important asset instead of a nuisance in the zoo or circus set-up." (p.434). A targeted study would be needed to settle whether research activities indeed qualify as environmental enrichment.

I have been fortunate to work with a zoo that gave me free hands to explore researcher - gorilla interaction. I learned many things that are not easily observed across a water moat, and in terms of getting to know one's study animals the endeavour was an educational success. I was also free to make my own methodological mistakes, which will be evident in the results and discussion in this chapter. Because of the scant availability of detailed training data in the literature, above all of training failures, I happened to chose methods that proved difficult to implement. In this regard the Givskud studies were less successful.

However, the findings are not in vain. As suggested in the preceding chapter, the attentional and relational demands confronting the subject in e.g. matching tasks, are demands that are also posed by picture – referent relations. Studying other stimulus relations than the specifically pictorial one can potentially inform us about constraints on the ability to apply comparisons.

Figure 18. The door through which testing took place, seen from the subjects' side.

Name	Sex	Birth year, origin	Rearing	Experiments
Amani	F	1985, Stuttgart Zoo	Mother reared/ hand reared	MTS/ discr./ obj.-choice
Kidogo	M	2000, Givskud Zoo Mother: *Minnie*	Mother reared	*Results not included for presentation*
Kipenzi	M	2002, Givskud Zoo Mother: *Sally*	Mother reared	*Not included in testing*
Minnie	F	1992, La Palmyre Zoo	Mother reared	Discr./ obj.-choice
Nille	F	ca. 1964, Cameroon (Wild caught)	Mother reared/ Hand reared	MTS
Sally (Salome)	F	1969, Frankfurt Zoo Deceased in 2005	Hand reared	MTS/ discr./ obj.-choice
Samson	M	ca. 1972, Cameroon (Wild caught)	Mother reared/ Hand reared	MTS/ discr./ obj.-choice

Table 3. The gorillas living at Givskud Zoo during the main study years, 2003-2005. The hand reared gorillas had extensive contact with peers during infancy but not with adult gorillas.

12.1 Subjects

The adult male Samson and the adult females Sally, Amani and Nille were the main subjects in the Givskud experiments (*table 3*). But also the adult female Minnie occasionally participated in testing. Samson and Nille had arrived at Givskud Zoo with the completion of the new gorilla facilities in 1998. Sally joined from Dublin Zoo in 1999, Minnie from La Palmyre Zoo in 1999, and Amani from Münster Zoo in 2001. Both Kidogo and Kipenzi, mothered by Minnie and Sally respectively, were born at Givskud Zoo.

At the beginning of the study all subjects were experimentally naïve. But in addition to my own experiments, in 2004 Amani, Nille and Sally were subjected to a test of their ability to use experimenter-given communicative cues in an object-choice task, i.e. gaze direction, pointing and tapping (Byrnit, 2005). Nille received the most testing in these experiments, amounting to 21 sessions (24 trials per session). Sally was given 15 sessions, and Amani 10 sessions. Nille performed well with pointing and tapping, and marginally well with gaze direction, while Sally only performed above chance with cues that consisted of tapping the baited container. Amani was only tested on tapping, but failed to perform above chance

Because Nille was low ranking and could not receive food by the experimenter when in the social group, she was predominantly tested in the first two weeks of the study when she was housed in isolation due to an injury.[138]

In the early stages of testing in the social group one female, Sally, was the primary subject. But due to apparent changes in dominance structures Sally successively gave way to a second female, Amani. Then, with increasing confidence mid-project, the adult male Samson more or less monopolised the testing area from all three adult females. Later conditions therefore mainly include Samson as the subject. Due to this opportunistic style of testing the amount of training each subject received in each condition is highly variable. Being both a training study and a search for a promising method not all subjects received specific conditions in the same order, but in terms of type of experiment they all worked on MTS before discrimination, and discrimination before the object-choice task.

12.2 MTS and discrimination

Before experiments involving different types of pictures were possible, it was judged that the subjects would need a robust matching competence. Matching, which is a case of conditional discrimination, is a more flexible experimental platform than simple discrimination. The latter depends on learning the correct target stimulus for every new set of contrasted categories, while the former "only" hinges on learning a matching principle.

---

[138] This was in the MTS phase of the study. In hindsight, given that she was proficient in following social cues in Byrnit (2005), it is a pity that she never received the baited-container version of my picture tasks.

A difference from most contemporary studies was that MTS training did not take place on a touchscreen, but with physical stimuli handled by the experimenter. This method is one step back in time, and is a reason for why the discussion of my results relies heavily on research from pre-touchscreen days. The pros and cons of touchscreens in picture studies are commented in section 13.3.

12.2.1 Materials and procedures, MTS and discrimination tasks

Laminated cards the size of A5 were used as stimuli. These comprised:

1. Homogenous coloured cards (red, green, blue, yellow, white).
2. Single geometrical shapes of the colours above on a white background (triangles, circles, crosses) (see *fig. 19, top right*).
3. Colour photographs (taken with a 5 megapixel digital camera at 1024x768 resolution) of familiar fruits, vegetables, nuts and chow. Photographs depicted single items, multiple items of the same kind, or multiple items of different kinds (see *fig. 19*). Objects were reproduced in near life-size.

In initial MTS training laminated cards that only differed in colour were used to keep attentional demands at a low level. But to depend solely on colour might be risky since there are reports of at least one chimpanzee with colour vision deficiency (Saito et al., 2003). Therefore motifs that in addition to colour also differed in shape were introduced, even though matching competence had not yet been acquired with colour cards. The shapes chosen (circles, triangles and crosses) have been used successfully with gorillas in previous matching tasks (e.g. Vonk, 2003).

In early testing lamination was glossy but after getting hold of a variety with a matt finish, stimuli were remade with this instead. In Savage-Rumbaugh et al. (1980) it had been found that the chimpanzee Austin's attention to pictorial motifs was compromised by reflections in the glossy materials used. The reflections seemed to have obscured the fact that there was a second surface behind the glossy encasing. Teasing apart the information pertaining to surfaces, e.g. reflections, from that pertaining to motifs independently of these surfaces, is arguably not always straightforward. In conclusion, learning about surfaces and their relations to markings upon, in, or behind them, is central for imbuing such marks with meaning (Ittelson, 1996). This is an extra task demand for naïve subjects.[139]

During training another change in the stimuli was introduced. After prolonged failure with using photographs of single items against a clean grey background (*fig. 19*), pictures with more complex motifs were produced. The reason for this change was that "simple" and "clean" pictures are not by default easily recognisable. On the contrary, these often show objects devoid of their common context, and can therefore be *less* recognisable. New photographs were taken of items against a wooden background, photographs of multiple exemplars of the same target item, and photo-

---

[139] But surfaces must not be eliminated altogether. "[…] the phenomenology of seeing a picture as a picture is entirely dependent on seeing the surface. If we remove the appearance of the surface, the picture is transformed phenomenologically into the real world" Ittelson, 1996, p.178).

graphs of a mixed array of items (see *fig. 19*). It was reasoned that if e.g. a grape was not recognised as a grape when on its own, maybe it would be recognised together with other grapes, or with e.g. a banana.

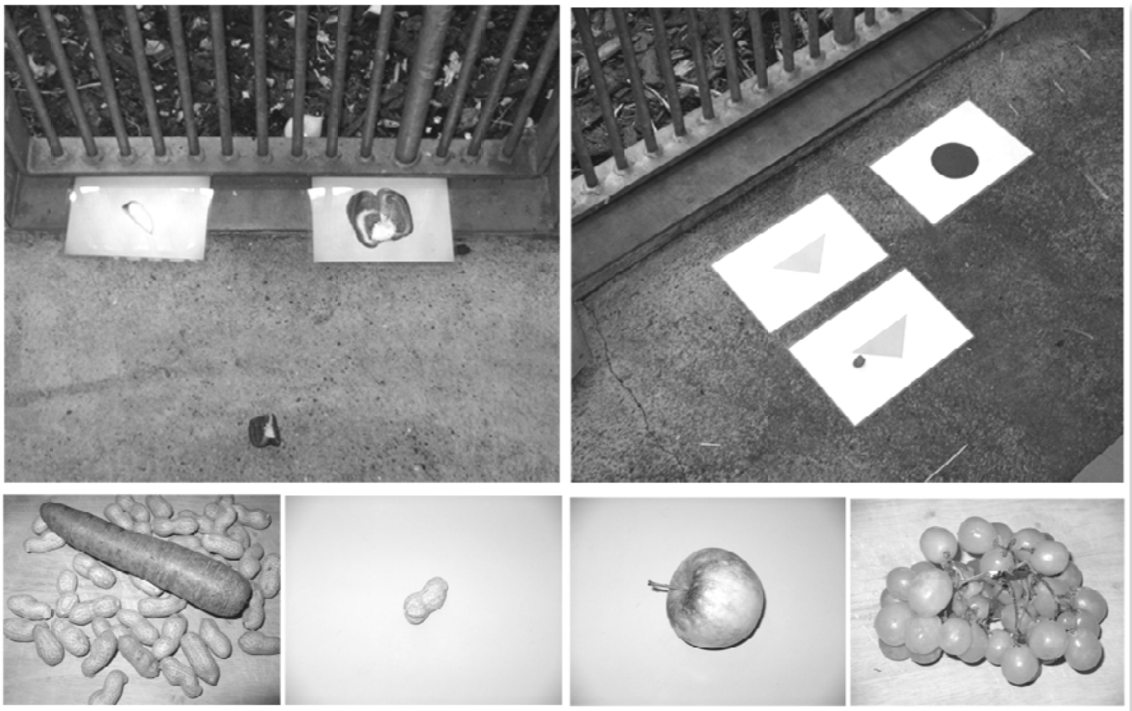


Figure 19. Setup (top) and examples of photograph stimuli (bottom) used in the MTS experiments. Note that t op right shows an *indexical* version of MTS, using cards with geometrical shapes.

In the coloured cards and geometrical shape conditions choice items were matched against identical cards that served as samples, while in the photograph conditions pictures were matched against real food items (see *fig. 19, top left*). All choice stimuli were circulated as matches and non-matches.

Harlow (1951) rely that in physical matching tasks performance is facilitated by presenting the match adjacent to the sample. Arranging these close together presumably facilitates attention to both of them, and comparison between them, although it does not automatically solve the problem of having to attend to their relevant properties. In simultaneous MTS the optimal display is thus to present the sample in the middle of the match and non-match. This has turned into the typical triangular display commonly used. In the present study this format was used whenever the stimuli setup involved three cards. When a piece of fruit served as sample it was instead placed between the two comparison photographs. When presenting the choice stimuli to the bars the sample naturally stayed behind.

To test whether nearness (indexicality) was a salient cue in this type of MTS setup, a condition was included where the sample and the match were closer together than the sample and the non-match (see *fig. 19, top right*).

A trial commenced when a sample was placed between the experimenter and the subject. In the case of laminated glossy cards the sample was angled towards the subject before it was places on the ground in order to give many views of it and counteract reflections. After making sure the subject had paid attention to the sample, the match and non-match were held in front of the subject at ground level until attention had been paid, and were then simultaneously presented within pointing distance (see *fig. 20*, p. 228). Left – right positioning was pseudo-randomised with the target picture never occupying the same position more than three times in a row. Presentation order was decided during testing. The subject then indicated one of the pictures through the bars with his or her hand. If the choice was incorrect both pictures were instantaneously withdrawn and a time-out commenced of approximately 20 seconds before next trial. Correction trials were not given in the standard setup.[140] If the subject pointed to both pictures it was judged as an incorrect reply. This happened seldom. If the choice was correct, the match was left under the finger of the subject while the non-match was withdrawn, and the subject received a food reward. In the photograph conditions the sample item itself was given as reward. In the "colour card" and "geometric shape" conditions the reward had been placed on the sample card at presentation and was now given to the subject. Food items were stored in a container behind the experimenters back. To avoid overfeeding, rewards were always small pieces of food, e.g. one shelled peanut, half a wedge of apple, etc.

Conditions that utilised simple discrimination, as opposed to MTS, are included in this section. They took the same form as the MTS task with the difference that the sample was replaced by a pre-set rule for which choice item would continuously yield a reward and which one would not. Proper discrimination was preceded by two cued conditions. In the first one, subjects were intentionally cued by the placement of a food item *on top* of the correct choice item. This food item subsequently served as a reward. The second cued condition was to place the food *underneath* the correct stimulus. Both types of baiting occurred in full view of the subject.

In the condition where food was placed on top of the stimuli it was removed and placed between the experimenter and the target picture before presentation, thus retaining an indexical relationship to the correct choice picture but not close proximity per se. In the "under" condition, however, the food was first placed on the ground in front of the subject. Then the pictures were simultaneously placed close to the bars, thereby obscuring the food with the target picture. At presentation the food reward was picked up between the fingers of the experimenter and held behind the correct choice picture. This created a finding game of sorts. Colour and shape pictures were used for discrimination trials. In addition, two new shapes were produced that served as constant positive (a green heart) and negative (a white square with a red diagonal) stimuli. These also differed from other shapes in that they were cut-outs, placed in A5 lamination, thus having transparent backgrounds instead of white.

---

[140] Immediate correction trials were occasionally given when choices were ambiguous. Such trials are excluded from the present analysis.

In all testing the experimenter looked at the subjects' bodies during response, but avoided direct eye contact. It was thus not possible to use gaze cues in order to pinpoint the correct choice objects.[141]

All sessions were filmed using a portable video camera mounted on a tripod. Trials were scored and noted with pen and paper between each trial. They were checked and corrected against the video uptake before analysis. Because of this notation process there was a time gap also between successful trials, but during this time the subject was busy ingesting the reward. The gap was significantly longer after incorrect trials because of the added time-out.

Nille, Sally, Amani and Samson received MTS training and Sally, Amani, Samson and Minnie participated in discrimination training.

12.2.2 Conditions

The order of the below conditions differed between subjects, but Nille, Sally and Amani received MTS before simple discrimination training. The inclusion of discrimination training was judged a necessity when it became clear that the gorillas did not seem to pay attention to, or use information, that stemmed from the colours, shapes or motifs of the cards. In hindsight, discrimination training should naturally have preceded MTS. However, Samson received extensive discrimination training before MTS. All subjects received training in indicating laminated cards before given the below conditions. Because of the explorative nature of the training and the opportunistic mode of testing, not all subjects received all conditions, or a substantial amount of trials in each condition.

Proper MTS

- "Colour" – Coloured cards as sample and choice stimuli.
- "Shape" – Cards with geometrical shapes as sample and choice stimuli.
- "Photo" – Real food items as sample and photographs as choice stimuli (see *fig. 19, top left,* p.221).
- "S-tactile" – The subject is presented with the sample (coloured card or card with geometrical shape) and is required to manually indicate this before being presented with the choice items. Thus the rule can be said to have changed from "touch the card that is similar to the one over there" to "touch the similar cards twice." Focusing attention by having the subject touch the sample image before the onset of choice stimuli is a common procedure in computerised MTS.

---

[141] Subjects never looked at the experimenter's face when making their selection. They thus gave no impression of actively searching for social cues.

“MTS” with indexical cueing

- “Cued, near” - The matching picture is presented close to sample picture (see *fig. 19, top right,* p. 221).
- “Cued, on” - Reward placed on top of match in view of the subject. It is removed and placed on the sample picture before choice is made.

“MTS” with stimulus cueing

- “Mixed” – Real food items as sample, photographs of food as matches, and coloured cards, cards with geometrical shapes, and photographs of gorillas as non-matches. Purpose of condition was to test whether food pictures were preferred above non-food pictures. These trials can be solved, in principle, by ignoring the sample.

Simple discrimination task

- Proper discrimination.
- Cued discrimination with reward *on* target.
- Cued discrimination with reward *under* target.

12.2.3 Results, MTS

All subjects instantaneously learned to point to a card to receive a reward, but it did not seem that they developed a principle for choice based on matching at any time during the study (see *table 4*).

Since robust MTS performance could not be obtained as hoped with “simple” stimuli, subsequent training took the form of searching for a procedure that worked. Thus, after approximately 500 trials of MTS training with Nille using colour cards and geometrical shapes, with performance still at chance level, a switch was made to the photographic condition. When starting work with the next subject, Sally, photographs were used from the start to see if this would facilitate MTS training. The “mixed” condition was then introduced to test whether the gorillas at all could develop a preference for photographs. It is here presented both as a type of stimuli, and as a condition.[142]

---

[142] The alternative would be to lump it with “Standard MTS” together with the clean colour, shape, and photograph trials, but the mixed trials do not require attention to the sample.

Subject:	*Nille*	*Sally*	*Amani*	*Samson*
	% corr. (total no. trials)			
Total:	50.7 (1054)	53.1 (2567)*	53.3 (1228)	44.1 (263)
Distributed over type of stimuli				
Colour	52.1 (236)	45.3 (364)	49.5 (214)	-
Shape	52.6 (519)	58.4 (1201)**	55.7 (687)*	43.6 (234)
Photo	49.0 (204)	50.4 (918)	43.6 (225)	-
Mixed	40.0 (95)	41.7 (84)	65.7 (102)*	48.3 (29)
Distributed over type of MTS setup				
Standard MTS	46.5 (417)	48.8 (1814)	47.3 (562)	53.8 (26)
S-tactile	-	35.8 (137)*	56.7 (67)	-
Cued, near	57.4 (411)*	74.5 (525)**	57.8 (464)*	-
Cued, on	50.4 (131)	57.1 (7)	45.5 (33)	42.3 (208)
Mixed	40.0 (95)	41.7 (84)	65.7 (102)*	48.3 (29)

Table 4. Overall results (percentage correct) on MTS, and distributed across different types of stimuli and setups. Note that number of trials varies substantially between subjects and between conditions. Note also that the "mixed"-trials appear both as type of stimuli and as a setup. * $p<0.05$, ** $p<0.001$, binomial test.

Nille performed marginally above chance (57.4%, X^2 (1, 411) = 4.55, $p<0.05$) in the MTS condition that utilised a nearness cue. This task can be solved by attending to the relative weight between the choice pictures resulting from the sample lying close to the match.

The same effect could be induced in Amani, who was 57.8% (X^2 (1, 464) = 5.62, $p<0.05$) correct. These trials also account for Amani's performance with shape stimuli at a level above chance (55.7%, X^2 (1, 687) = 4.56, $p<0.05$). In standard MTS Amani did not reach significant performance, other than with the mixed stimuli (65.7%, X^2 (1,102) = 5.15, $p<0.05$). In this condition the match was always a photograph and the non-match a non-photograph. Thus Amani indicated that she sometimes recognised a difference between photographic stimuli and other stimuli. However, without a conditional form of matching one cannot know whether it was the content of the pictures that she responded to, or any other invariant feature between photographs. Amani was exclusive in finding a strategy for solving the mixed trials.

Sally was the only subject that performed above chance overall, but marginally, with 53.1% (X^2 (1, 2567) = 4.93, $p<0.05$), which she did with shape stimuli (58.4%, X^2 (1, 1201) = 16.94, $p< 0.001$). However, all her performance can be attributed to indexical cueing in the "nearness" condition where she excelled with a performance of 74.5% correct (X^2 (1, 525) = 66.91, $p<0.001$). She also performed significantly *below* chance in the "S-tactile" trials (35.8%, X^2 (1, 137) = 5.67, $p<0.05$). This is due to the fact that during these trials she got stuck in a positioning preference. As a consequence of me trying to break her fixation by presenting more correct choices in the opposite position, the left – right distribution of matches is not 50%.

12.2.4 Results, discrimination

Table 5 displays the result from the discrimination training. Colour and shape pictures were used for discrimination trials and are lumped for all subjects.

Subject:	*Sally*	*Amani*	*Samson*	*Minnie*
	% corr. (total no. trials)			
Cued, on	62.0 (213)*	75.2 (125)**	52.1 (564)	84.5 (71)**
Cued, under	69.4 (160)**	79.4 (155)**	61.3 (775)**	83.9 (62)**
Non-cued	80.0 (25)*	46.0 (176)	39.4 (180)*	51.5 (103)

Table 5. Results for cued and non-cued simple discrimination. Cue was reward placed on top or underneath target stimulus. Note that number of trials varies substantially between subjects and between conditions. * $p<0.05$, ** $p<0.001$, binomial test.

All subjects performed well in cued trials, especially where the food reward was placed under the correct picture and held there during the choice. It is important to note that most subjects also performed well when the food was removed before choice, as in the cued condition where food was placed on top of the correct choice. This means that it was not just a matter of reaching for the food and accidentally indicate the picture in the process in the "under condition". Also, the gorillas did not give the impression of trying to remove the card from my hand when they did their choice in those trials. But the food evidently had a strong effect on attention to the correct choice item.

Sally was the only subject who performed well in the non-cued condition at 80% (X^2 = (1, 25) = 4.95, $p<0.05$) correct. Unfortunately this testing took place during a period when the adult male (Samson) started to monopolise the testing area and her number of trials are low. She also performed above chance when the reward was placed either on (62%, X^2 (1, 213) = 6.19, $p<0.05$) or under (69.4%, X^2 (1, 160) = 12.48, $p<0.001$) the target picture.

Amani performed even better than Sally in the cued condition, both when reward was placed on (75.2%, X^2 (1, 125) = 16.95, $p<0.001$) and underneath (69.4%, X^2 (1, 155) = 29.23, $p<0.001$) the target picture.

Minnie, although she received the least number of trials, performed best of all subjects in the cued conditions, but by chance level in the non-cued, just like Amani. Minnie 84.5% (X^2 (1, 71) = 19.19, $p<0.001$) correct when reward was placed on the target, and 83.9% (X^2 (1, 62) = 16.07, $p<0.001$) when it was placed under.

Samson, in turn, only performed above chance when the reward was placed under the target picture in cued trials (61.3%, X^2 (1, 775) = 20.01, $p<0.001$). Interestingly, in Samson's case, the non-cued trials were exclusively conducted with two novel pictures that were created specifically for the discrimination task. The positive stimulus (a green heart) and the negative stimulus (a white square with a diagonal red line) where held constant. These were cut-out shapes encased in non-glossy lamination the size of A5. Thus, Samson had never received a food reward for indicating the white square, but continued to point to this card. Even *more* so than to the green heart. His correct response was thus significantly below chance (39.4%, X^2 (1, 180) = 4.06, $p<0.05$).

That Samson (as well as Amani and Minnie) failed to respond to a simple discrimination might seem a mystery. In e.g. Rosenfeld and Van Hoesen (1979) experimentally naïve rhesus macaques needed on average 100 trials in order to reach a criterion of 90% correct in a similar discrimination task with abstract stimuli. They learned colour discrimination significantly faster than shape discriminations though. Since the gorillas could simultaneously use colour, shape, and in several conditions food cues, their task ought to be even easier. However, a notable difference was that the monkeys in Rosenfeld and Van Hoesen (1979) were required to manipulate their stimuli directly to acquire their food rewards. It is possible that at Givskud the relationship between the act of indicating a laminated card, followed by a possible external reward, had a far from transparent connection to the specific visual properties of the cards. That this was indeed the case will be argued further in the discussion below.

12.2.5 Free-response

Were the gorillas at all sensitive to picture motifs? Results for the mixed condition, where photographs of food was pitted against non-depicting stimuli (or photographs of gorillas), suggests that the subjects did not have a preference for food pictures, at least not coupled to their understanding of the role of the choice process. There is tentative evidence captured on film, however, that the gorillas did target food items in photographs with investigatory behaviour. The adult gorillas' hands were too large to reach the depictions through the bars, but the young ones could, as can be seen in *fig. 20* (p. 228).

To one particular photograph of a pepper Sally responded vividly, repeatedly touching the part of the picture that she could reach, sniffing and tasting her fingers.

After this initial interest she stopped paying attention to the pepper photograph. One cannot presume that they stop recognising the pictures when investigation subsides, but it seemed as if the pictures lost all meaning rather than acquired new meaning in the form of strange instances of fruit, or representations of fruit.

Figure 20. A gorilla (1.5 years) at Givskud Zoo pushes away his mother's hand and scratches and grasps at pictures of a carrot and a peanut.

12.2.6 Discussion, MTS and discrimination tasks

From chapter 11 it seems clear that there is a large variation in how quickly individuals learn to match in experimental conditions, ranging from almost effortlessly to not at all. It seemed as if the methods used for training are mainly to blame for this variation. Using physical objects instead of displays, for example, seemed to be a shortcut to performance. I will save the underlying processes for the general Givskud discussion and here only look at some surface differences between different experiments and relate these to the Givskud training studies.

Reported in Nissen et al. (1938), an attempt was made to teach conditional discrimination to chimpanzees by presenting a black and white box against a black or white background. The colour of the background indicated which box that was the correct one in that trial. The subjects needed 690 to 1217 trials to reach a criterion of 70% correct. This seems to be a simple and straight-forward task. Yet it required long training. Giving the complexity of the Givskud situation with up to three disjoint laminated pictures, a half-stranger as experimenter, the full dynamics of the social group in the background, and not really knowing what it is that makes the half-stranger give you food, it seems hardly strange that e.g. Sally's 2 500 trials would not teach her matching. She might even lack motivation. She had after all been given 1 250 pieces of food just for pointing at plastic sheets.

The reinforcement schedule used is a given candidate for why pictorial content was disregarded in the task. Most pictures need attention and active interpretation.

A superficial glance is not enough. That food rewards can have detrimental effects on apes' spontaneous use of pictures has been pointed out by Gardner and Gardner (1998). Referring to Morris (1963) (see chapter 10) they claim that the chimpanzee Congo exchanged his elaborate painting to inattentive scribble as soon as he figured out that he could paint for food. The interest in painting for painting's sake was forfeit. Similarly, Gardner and Gardner suggest that Viki in Hayes and Nissen (1971) (see section 5.1) for the very same reasons was unable to transfer her spontaneous use of pictures when requesting e.g. drinks, to a formalised deck of picture cards. What the food did was to take spontaneous interest and attention away from pictorial displays as such, and refocus attention on getting hold of food.

However, food incentives are not all bad. In a delayed object-choice task with baiting in full view, two of three juvenile chimpanzees in Cowles and Nissen (1937) performed better when expecting a preferred (i.e. large) reward, than when the containers were baited with a small reward. But, importantly, this effect was only evident in sessions where trials with large and small rewards were interspersed. Constant large rewards did not boost performance. In the Givskud studies the quality and quantity of food rewards were often highly variable.

Whatever the impact of food rewards might be in this case, the reasoning of Savage-Rumbaugh (1986) is even more enlightening. She maintains that a successful setup *must* allow for a cause and effect analysis of the workings of the task. Introducing a human experimenter, that are not even well known to the subjects, who seemingly decides in an unpredictable way when a reward shall be given, and when it should be withheld, is a confusing matter that hardly facilitates learning. There should be no need for a subject to interpret the effects of its own actions on the behaviour of the experimenter. If attention is on trying to figure out cause and effect relations pertaining to other things than the pictorial task, naturally it will fail.

It is thus not surprising that, following several difficult training studies, Yerkes and Nissen (1939) writes: "In our findings we consider most significant the evidence that delayed response, in the absence of spatial cues or with misleading cues, is either extremely difficult or impossible for most chimpanzees" (p. 587). They thus corroborate the Givskud results, i.e. that spatial relations, such as indexical cues, are far more salient to apes than are purely visual properties.

To discriminate colours in Riesen (1940, in Yerkes, 1943/1945), chimpanzees needed 60 to 180 trials to discover which of two choice items was the correct one. Differential reinforcement (i.e. food reward for correct choice) was immediate. However, when delayed more than a few seconds (i.e. 4), visual discrimination learning by naïve subjects did not occur at all. However, one can hardly say that it is the *delay* as such that is the problem. Experienced visual discriminators had no problems with delays in the above study. It seems to be the initial discovery of a relation that is hindered by delays. In other words, the initial stages in learning to match are very sensitive.

For a chimpanzee to discover the salience of an (arbitrary) visual cue in a discrimination task, the cue must be in an indexical relation to the reward (Riesen, 1940, in Yerkes, 1943/1945). The focus in learning theory has been on direct temporal proximity, but direct spatial proximity, as we will see in the discussions to follow, is also an influential factor. It might be that what causality does is to bridge the relations between a stimulus, the reward, and the response to the stimulus needed to bring about the reward. That is, a transparent reason for why there exists a relation, and a reason for why the reward is delayed, is a necessity (see Savage-Rumbaugh, 1986).

That the choice areas in the Givskud studies differ from the position of the relevant cues is a major flaw. In the MTS-tasks the actual cue was in the centre of the cards, which were also out of reach for the subjects. This might sound like an irrelevant factor, because surely the subjects see the card and the pattern in its centre as one and the same object? But spatial cue - response dissociation has been found to be a problematic issue in several studies (e.g. Iwai et al., 1986; McClearn & Harlow, 1954; see Meyer et al., 1965). That said, there was no significant difference in performance with coloured cards and cards with shapes on them, since both types failed miserably. In the former case the relevant cue, i.e. the colour, was within physical reach by the subjects. However, without a matching competence such a difference would not manifest.[143] But that colour is included in the response does not in any way guarantee that it is attended to as a relevant cue in the task. Colour, specifically, must be discovered to co-vary with reward. Colour must therefore be abstracted from the overall task. It was my belief that coloured geometric symbols would enhance the differences in colour in relation to homogenous cards, but this might not have been the case.

Smith et al. (1975) performed an error analysis on the results from training two experimentally naïve adult chimpanzees on sameness-difference matching (relations between relations). Unfortunately data is not given for the number of trials needed to reach criterion, but both subjects became proficient in selecting among two sets of items one that displayed the same relationship as two items that served as sample. As stimuli coloured geometric plywood shapes were used, and as rewards food was hidden in wells beneath the correct match. Initially the chimpanzees displayed, as expected, random responses, but also position preferences, which did not seem to have depended on specific reinforcement history. (In a randomised test with potential reward on each trial a position preference always yields 50% reward.) During the same period, however, the subjects did not display stimulus preference/aversion, which suggest that they paid little attention to the looks of the stimuli, at least as pertaining to the resulting rewards. As they became better at matching, random responses as well as position preference decreased. When switching to novel conditions random responses initially increased, while position preferences stayed low. In the end only random responses accounted for the subjects' mistakes.

If this is a general development it seems clear that the gorillas were in early stages of learning. Position preferences were common, they disregarded the appearance of the stimuli, and they responded randomly. Consequently there were no stimuli preferences/aversions. The notable exception would be Samson's avoidance of the picture that always yielded a reward, the green heart.

---

[143] This comparison would be more relevant in the discrimination task, but unfortunately only a few trials were run with homogenous colour cards.

So how can learning of MTS be facilitated? Using real objects has already been mentioned. The seemingly effortless MTS training of the chimpanzee Sarah took place in this manner, with physical objects and with shaping procedures, i.e. making Sarah touch first the sample and then the correct match (Premack, 1976; Premack & Premack, 1983). In addition, the indication procedure was to physically place the match with the sample, not only point at it.[144] Peony, another of Premack's subjects, performed poorly in a structured same/different task, but when freely allowed to spatially arrange the comparison objects as well as the labels "same" or "different," she showed that she understood both the task and the meaning of the plastic tokens used as labels (Premack, 1976; Premack & Premack, 1983). As mentioned, Oden et al. (1988) also used physical objects that seem to have allowed physical manipulation in their highly successful MTS training with infant chimpanzees.

Another method which allows instantaneous feedback, in both space and time, is the computerised touchscreen. It has also the potential to make the experimenter obsolete. It seems to be used successfully in most modern MTS setups.

Vonk (2003) does not specify how DMTS was trained in her orangutan and gorilla subjects, but likely on touchscreen. They had been proficient with photographic stimuli before the tasks in Vonk (2003), which were based on coloured geometric shapes, like the ones in my study. When subjected to these novel stimuli the gorilla juvenile in Vonk (2003) needed 72 trials to reach criterion. There was no significant difference in matches that depended on complete identity and those where colour or shape needed to be abstracted and then matched. One of the orangutans reached criterion within her first 12 trials on the novel task. The worst orangutan needed 168 trials. With very similar stimuli as my shape stimuli Vonk's worst subject performs better than the fastest gorilla in less modern circumstances.

However, as in all matching, and as found in Vonk (2003), individual variation can be expected to be large even with touchscreen. Boysen (1994), for example, found striking individual differences in four chimpanzee subjects that learned to match colour photographs on a touchscreen to samples in the form of physical colour photographs. All pictures depicted known human individuals. The chimpanzees received a session consisting of 20 trials each day. Sarah and Kermit only required 20 sessions to reach a criterion of 85% correct on two successive sessions. Darrell did not reach the same level until the 52nd day and Sheba on the 72nd, which equals 1440 trials. None of the subjects transferred their matching competence to cross-modal matching of vocalisations to individuals without further training. However, they differed markedly also in this training, with Sheba needing only 6 sessions, Darrell 28, and the remaining two 60 daily sessions. Only one subject, Darrell, matched also novel stimuli spontaneously and accurately on initial trials. In second place came Sheba. It is reasonable to attribute this varied pattern of performance across the study to very different ways of learning the task. It might be that the subjects that were initially successful hit upon the strategy of learning specific pairings in the first task, which is the reason they fell back to rote learning in the subsequent tasks. The subjects that arguably did not go for this option, but seemed to learn

---

[144] Note the indexical relations again.

matching instead, needed many more sessions initially, but fared better in the new condition and with novel stimuli.

Touchscreens are not immune to wandering attention either. The correct use of attention in a task that requires comparison is of course critical. Washburn et al. (1989; Washburn, 1993) facilitated the performance on a range of visual tasks that required comparison between stimuli by having the display move around on the computer screen. The rhesus macaques tested performed significantly better than when the stimuli were stationary. However, this effect of attention allocation is believed not to be due to the salience of moving stimuli per se, but to the added difficulty of the task (Washburn & Putney, 2001). For humans and animals alike it has been found that increased motor demands sharpens attention and success rate on a task at large. But regardless of the underlying mechanism, the fact that there is a marked difference in performance between static and moving pictures tells us that there was a lack of attention, or effort, to begin with.

How the task is structured on the touchscreen is also important. In a same/different judgement task by Katz et al. (2002) rhesus macaques were dramatically better at learning a same/different response if they were first required to touch one of the comparison stimuli before the other came on-screen. The "same" response was then to indicate the novel image while the "different" response was to avoid this second picture and instead touch a grey rectangle by the side of it. This setup not only focuses attention on the sample, it also reduces the number of comparisons on the display to one.

12.3 Object-choice task

The object-choice task employed at Givskud had its source in a tentative observation made during the discrimination training further above. As described, following extended unsuccessful training on MTS it was suspected that the gorillas did not attend to the visual information on the laminated cards. In order to investigate this, and perhaps evoke closer attention to motifs, trials on simple discrimination were given (see *table 5,* p. 226).

Two conditions were given where the subject was indexically cued by placement of the food reward either on top of the target picture, or by covering the food reward with the target picture before presentation. The observation referred to was the apparent ease by which subjects discriminated when food was hidden underneath the choice stimuli in the second condition. This task seemed to have taken the form of a hiding event, which indeed seemed an intuitive procedure. This was developed into the "box setup" presented below.

It was believed that a finding game would be more intuitive because it was a clearer case of problem solving, and the act of indication is instrumental rather than communicative. The subject really wants to get into the box it is indicating.[145] This

---

[145] That said, indication in the box-task quickly took the form of pointing rather than attempting to open the box. This type of response prevailed regardless of whether the baiting was hidden or had taken place in full view.

mode of response is arguably more causal than pointing to cards in a MTS setup. As will be shown, the box setup failed for very different reasons.

12.3.1 Materials and procedure, object-choice task

The object-choice task utilised two dark brown wooden boxes (*fig. 21*) approximately 8 x 15 x 7 centimetres in dimension. Food could be placed in the boxes, which could be sealed with lids. Pictures could then be fastened into two blocks on the lids. The slits in the blocks were angled, so that the pictures would slant somewhat and potentially allow easier inspection.

Two types of pictures were used. Photographs of foods, of the types described above for the MTS training, and the green heart and white square described above for the simple discrimination trials. Photographs were a mix of reused and novel ones.

Figure 21. Differentially baited choice objects in the object-choice task.

Experimenter and subject were sitting facing each other. The boxes were placed close to the experimenter and hidden behind an occluder that was placed between the subject and the experimenter. A food item was held up until attention was paid to it and was then brought behind the occluder. During baiting the boxes were placed adjacent to each other and the experimenter's hand moved between both boxes several times to preclude cueing. The food reward was placed in one of the boxes during these movements. The occluder was small enough to allow the subject to follow the manipulation of the boxes and their lids by attending to the experimenter's movements and the sounds of the material. The structure of the baiting process was thus potentially transparent, but the actual placement of the food could not be observed. Pictures were then placed on top of the boxes, the boxes were placed apart, and the occluder was removed. After making sure that the subject attended to the boxes, and hopefully their pictures, they were then pushed forward and held close to the bars so that the subject could indicate his or her choice.

Typically only one box was indicated on any given trial. If choice instead was ambiguous the boxes were withdrawn, a time-out took place, and the boxes were again presented. Following a proper choice the box that was not chosen was withdrawn while the indicated one remained close to the subject. The indicated box was opened and the content, if any, was shown and given to the subject. The box that was not indicated was then opened and shown to be either empty or containing the missed food. Both boxes then returned close to the experimenter and next trial commenced by occluding them again. If the choice had been correct a new baiting took place as described above. If incorrect the boxes were re-baited with the same food as on the previous trial. A food item that was re-baited was always shown to the subject before baiting to inform his or her that the item could have changed place in the coming trial.[146]

By now Samson was the main subject in testing. Besides testing the Givskud gorillas, brief sessions were replicated with Igelchen, an adult female Sumatran orangutan (*Pongo abelii*) at Furuvik Zoo, Sweden, and Linda, an adult female chimpanzee at the same zoo.[147] This testing took place in February 2007. The bonobo Matata, at Great Ape Trust of Iowa (GATI), USA,[148] served as the third subject in the replication. She was tested in August 2006. A retest of Samson is included in the same comparison, made in March 2006.

12.3.2 Conditions, object-choice task

- Object choice, photographs – One of the boxes is baited with food and both boxes are labelled with photographs. The photograph on the target box depicts a food similar to the one inside the box. The non-target photograph depicts food that is not included in the trial.

- Object choice, shapes – Same setup as above, but pictures do not depict content of boxes. They where a green heart for the target box and a white square crossed by a diagonal red line for the non-target box.

- Cued object choice, indexical cue – photograph or shape picture was only placed on baited box. Unbaited box remained unmarked.

- Cued object choice, transparent baiting – baiting of the boxes took place in full view of the subject. Following baiting the boxes could, or could not, be labelled with pictures. (Trials are lumped in analysis.)

---

[146] Trials without occluder confirmed that the gorillas could choose by exclusion if they saw that the boxes were *not* re-baited.

[147] Ing-Marie Persson served as experimenter in the testing of Linda.

[148] Takashi Yoshida served as experimenter in the testing of Matata.

12.3.3 Results, Object-choice task

Trials where the food shifted inside the boxes and gave away the location through noise cues are excluded from analysis. Ambiguous answers (trying to choose both boxes simultaneously) are counted as incorrect. In the object-choice task left and right presentation of the boxes was sometimes randomised, and sometimes held constant. In the latter case the problems could be responded to as a discrimination rather than as a conditional discrimination. Even responding to position would suffice to perform adequately, until the stimuli switched places, that is. In *table 6* these two conditions are lumped in order to yield a higher number of trials for each subject.

Subject:	*Sally*	*Amani*	*Samson*	*Minnie*
	% corr. (total no. trials)			
Photographs	40.9 (22)	66.0 (100)*	50.9 (53)	59.7 (72)
Shapes	100 (4)	57.1 (3)	47.1 (70)	100 (12)
Indexical cue	50.0 (54)	61.9 (42)	70.5 (217)**	45.5 (55)
Full view baiting	71.7 (53)*	75.9 (79)**	81.4 (177)**	77.8 (18)

Table 6. Results for cued and non-cued object-choice tasks with differentially baited boxes labelled with pictures. “Indexical cue” was placement of picture solely on baited box. Note that number of trials varies substantially between subjects and between conditions. * $p<0.05$, ** $p<0.001$, binomial test.

As can be seen in *table 6* above, all gorilla subjects, except Minnie who received too few trials, performed significantly above chance when baiting was made in full view. Sally 71.7% (X^2 (1, 53) = 5.24, $p<0.05$) correct, Amani, 75.9% (X^2 (1, 79) = 11.40, $p<0.001$), and Samson 81.4% (X^2 (1, 177) = 38.60, $p<0.001$).

In addition, Amani performed above chance (66%, X^2 (1, 100) = 5.25, $p<0.05$) without cueing in the photograph condition. This lends credit to the suspicion that she attended to photographic displays also in the MTS “mixed” trials (see *table 4*). However, when looking at her responses in detail one can see that she at best learned the task as a simple discrimination, and not a conditional one. Furthermore, series of trials where left and right targets were held constant accounts for most of this performance, which means that she could have relied solely on positioning and not at all on the picture displays.

Samson performed well in the indexical condition, at 70.5% (X^2 (1, 217) = 19.05, $p<0.001$). In these trials only one of the boxes, the baited one, was labelled with a picture, and in Samson’s case with the green heart on a semi-transparent background. However, when contrasted with the white square in the shape trials he performed at chance level. Thus the indexical cue was indeed indexical and nothing more.

12.3.4 Furuvik and GATI replications

It became increasingly clear that the object-choice task, or finding game, was not as "intuitive" as hoped. It was reasoned that perhaps the gorillas at Givskud had habituated to the stimuli after extensive exposure to it in unsuccessful matching tasks (see e.g. Wasserman & Miller, 1997). The attitude that some information is not useful, when in fact it is, is of course detrimental when learning MTS. Therefore the same task was given to subjects that had not been involved in previous testing with the material, nor the experimenter. The replication sessions at Furuvik Zoo and GATI can be seen in *table 7*. The trials are few but the trends in these subjects' responses were instantly recognisable from Givskud.

Matata's non-cued object-choice trials differed from the other subjects'. Instead of baiting only one box, both were baited, and the mounted photographs depicted respective content. Matata thus received 41 trials where whichever box she chose contained the food depicted in its photograph. In the session that followed, both boxes were again differentially baited, but this time only one of them contained food. The other one was baited with an undesirable object, i.e. a plastic spoon.[149] Thus the positive box was labelled with a food photograph and the negative box with a photograph of a spoon. A stable differential outcome rather than the presence or absence of food rewards, seemingly arbitrary placed by the experimenter, was judged to maybe help Matata utilise the photographic information (see e.g. Savage-Rumbaugh, 1986). As can be seen in *table 7*, it did not.

Subject:	*Samson*	*Igelchen*	*Linda*	*Matata*
		% corr. (total no. trials)		
Photographs	45.1 (82)	56 (50)	66.7 (48)	61.4 (57)$^{+}$
Indexical cue	78.7 (155)**	-	-	-
Full view baiting	89.7 (29)**	92.9 (14)*	85.7 (28)*	90.2 (41)**

Table 7. Results for cued and non-cued object-choice tasks with differentially baited boxes labelled with pictures, for Samson (*Gorilla gorilla*) (retested), Igelchen (*Pongo abelii*), Linda (*Pan troglodytes*), and Matata (*Pan paniscus*). "Indexical cue" means picture is on baited box only. Note that number of trials varies substantially between subjects and between conditions. * $p<0.05$, ** $p<0.001$, binomial test. $^{+}$Matata's trials differed from the others', see above.

---

[149] This is a similar setup to the original Köhler (1922; 1925/1957) experiment, where photographs of bananas and a rock were pasted on boxes.

In the replication one year after his last session, with no training, Samson again performed well in indexical trials with 78.7% (X^2 (1, 155) = 27.85, $p<0.001$) correct choices, and in trials where baiting took place in full view (87.7%, X^2 (1, 82) = 10.82, $p=0.001$).

Igelchen likewise understood the task when it was given with baiting in full view (92.8%, X^2 (1, 14) = 6.30, $p<0.05$), but like the others did she not utilise the information in pictures placed on the boxes. The same was true for Linda with 85.7% (X^2 (1, 28) = 8.19, $p<0.05$), and Matata, 90.2% (X^2 (1, 41) = 15.85, $p<0.001$) correct.

12.3.5 Discussion, object-choice task

Hayes et al. (1953a) suggests an important difference between two lines of development in the subject in the context of a discrimination task. The first one is to form hypotheses about the rules of the task, such as "only one box is always baited" and "the baited box can be in either position." Unless it is a conditional discrimination a central hypothesis would also be "the same box is always baited." The second line of development is the improvement of the skills necessary for successfully act according to the hypotheses. These are skills such as looking carefully before making a choice, notice how two objects differ, remembering what was learned on a previous trial, etc. These two facets of the task cannot work independently. From the data it is not possible to pinpoint where e.g. Samson (or the experimenter) fell short, but it can be in any of the skills just mentioned.

The subjects seemed to appreciate the fact that the task was indeed a finding game. Both boxes were shown empty before baiting and only one food item was shown to disappear behind the occluder during the baiting process. For example Call (2001; 2006a) has demonstrated the understanding of such relations in similar contexts. None of the subjects showed visible surprise or annoyance when an empty box was selected, suggesting that they expected only one box to be baited. In addition, the subjects were able to select the correct box when it was baited in view (also when the boxes switched places in full view). Such tasks are usually successfully solved by apes (e.g. Call, 2003; Barth & Call, 2006), and suggests that they do not expect baited food to magically move in and out of boxes but remain where it was placed.

However, all subjects seemed to pay more attention to the baiting process of the boxes than to the subsequently presented choice. They typically made their selection instantaneously. However, it is difficult to see where a gorilla is looking, and what has been noted. Much can be gained from a quick glance. I made sure that the subjects had looked in the direction of the boxes or the pictures before I presented them close to the bar where the choice was made. If I was not sure whether they had looked or not I called their attention by tapping or moved the stimuli conspicuously.

The problem was more likely due to not perceiving the function of the pictures in the task. For example, in a task similar to the one used here, 2-year-old children in failed to use picture cues placed on baited boxes while 3-year-olds succeeded

(Loughlin & Daehler1973). DeLoache (1986; DeLoache & Brown, 1983) found similar effects in a memory task for 2-year-olds. Only visual cues that were integral aspects of the hiding places were helpful when recalling the position of a hidden item. Picture cues placed on top of containers were not helpful. DeLoache attributed these troubles to a general difficulty in young children to integrate unrelated information. However, by reducing the usefulness of attending to spatial cues, or by providing labels for the photographs, Horn and Angrist Myers (1978) increased the performance somewhat for 2-year-olds. It was further boosted when photographs that closely matched the content of the respective boxes were used (Horn Ratner & Angrist Myers, 1980). In all of the above experiments the children could recognise the content of the photographs. It was rather the pictures' connection to the content of the boxes, and the usefulness of this connection, that did not dawn on the children until in the third year of life.

But also when the containers themselves are visually distinct does the ape attention seems to work with other variables. While the Yerkes chimpanzees, when searching for an object that had been hidden, used spatial cues without hesitation, they were unable to use visual cues after a delay during which the choice objects had been shifted from their positions (Yerkes, 1943/1945; Nissen et al., 1938). "[W]e made thousands of observations with different forms of box test, and later with a turn table apparatus, in an effort to discover conditions under which correct responses from memory would be possible. […] We finally were forced to admit that our subjects either failed to perceive the essential visual cue, or were unable to hold it in mind because they lacked a symbol or representative process comparable with our word 'green'" (Yerkes, 1943/1945, p. 179). A gorilla studied by Yerkes (1928) occasionally used visual appearance to remember a hiding place, but was much more prone to use positional cues. Likewise, Haun et al. (2006) recently confirmed that position and not visual looks governed which container all of the great ape species tended to select in a finding task. This was true also for 1-year-old humans. By the age of 3 years, however, human children had switched to using visual cues.[150] This switch was cautiously attributed to language, which has been shown to help with feature abstraction in other studies (Haun et al., 2006). This conclusion was also suggested by Nissen et al. (1938, p. 383): "It is only when we reach man with his highly developed symbolic system (language), however, that non-spatial stimuli are as efficient, and as readily used, as are spatial cues."

In view of the above ape and child data it seems hardly surprising that attention to visual cues such as the placement of a marker is not always informative for a naïve subject when it comes to finding food in an object-choice task (e.g. Tomasello et al., 1997; Call, 2003). The question, then, is whether it is a *stand-for* relation in the use of the designating objects that is opaque, or if this relation is obscured by the task as such. As will be argued in the overall discussion further below, the standard object-

---

[150] The ape results are tentatively explained by foraging strategies. Places were e.g. fallen fruit lurks are in general stable, and if anything it is the food that moves between places. That attention to visual cues as salient discriminative information in an object finding task can be outcompeted by spatial search strategies is in other words not surprising.

choice task has some classic limitations in its setup. The spatial discontinuity between locus of response, relevant cue, and reward, seems to complicate the task in the extreme (see e.g. Jarvik, 1953; 1956; section 12.4). This "contiguity critique" is the major flaw also in the Givskud setup.

In all studies a striking individual difference is evident in the use of markers (Tomasello et al., 1997; Call, 2003; Herrmann et al., 2005), or the use of container looks (Haun et al., 2006; Yerkes & Nissen 1939). A specific history of using the respective cues seems to be a necessity for good performance (Tomasello et al., 1997). For example Chantek, the sign-language trained orangutan, excelled in both Tomasello et al. (1997) and Call (2003), which was attributed to his training history. In addition, three chimpanzees with symbol training in Beran et al. (2005) spontaneously used a market in an object-choice task, while a fourth non-symbol trained chimpanzee did not. It remains to be clarified what it is in symbol-training that seems to have this effect. It is probably not language itself, since several of the successive subjects cannot be said to have language, at least not in the sense of being able to communicate with humans in a shared arbitrary code. This is especially true for Herrmann et al. (2005), where ordinary zoo subjects, albeit some with experience from object-choice studies using markers and pointing, successfully used indexical as well as iconic cues to locate hidden food (see section 5.4).

While all species of great apes are able to use functionally relevant experimenter-given cues in an object-choice task, such as sniffing, biting, or trying to take apart a container in an adequate way, they are less prone to use functional irrelevant (but intentional) cues (Buttelmann et al., 2008). It is also known that competitive cues, such as reaching, are more informative to chimpanzees in an object-choice task than are "communicative," cues such as pointing (Hare & Tomasello, 2004). More so, in the experiments of Buttelmann et al. (2008) the subjects tended to disregard previously successful cues when they knew that the containers were now empty, despite the fact that they were rewarded for continuing to choose cued containers. When the behavioural cue stopped to make sense, so did its communicative potential. A very similar effect can be argued to render the photographs in my box trials uninformative and useless, despite the fact that they too are steadily reinforced in relation to the location of the hidden food.

Causality can render even more static views than pictures informative. When cues are transparently causal, as when the slant of e.g. a piece of wood gives away that something is hidden underneath it, apes are successful in object-choice tasks (Bräuer et al., 2006). Call (2006b) compared results on the first 24 trials of three types of object-choice tasks given to ape subjects in a range of studies at the Wolfgang Köhler Primate Research Center. Available cues were simple colour or shape cues, arbitrary noise or shape cues, or causal noise or shape cues. Within these 24 trials, only the causal cues resulted in performance significantly above chance for the subjects as a group. For arbitrary noise and shape cues, as well as for just simple colour and shape, the groups did not reach above chance level.

Apes typically require 54 to 704 trials to recognise the usefulness of visual cues (Rumbaugh & Washburn, 2003). But 4 of the 41 individuals in Call (2006b) did perform above chance within 24 trials in the arbitrary condition, and 1 of 18 in the simple discrimination task. It is thus within the ape potential to realise the relationship between such cues and rewards, the question is only *why*. A further question is why pictures, that seem to be such intuitive stimuli, seem to group with the arbitrary or simple cues rather than with the causal ones.

It remains to be investigated whether visual properties of containers themselves can become meaningful if they can be made to look causally induced. Is for example a container distorted by teeth marks chosen before one that is not, although actual attempts to open the container has not been observed by the subject? Will such marks gain more salience if the ape sees that an experimenter struggles with opening one of the containers, presumably the baited one, with his or her mouth? Can a picture be similarly connected to a specific container?

Perceived causality stems from experience. The deduction that a hidden item is probably located under the flat object that slants, for example, derives from experience of the physical properties of objects: they take up space. Pictures, in turn, can perhaps obtain a causal relationship to objects in boxes through the recruitment of similar pre-existing knowledge, such as the experience of windows or reflective surfaces. One way could be to familiarise the subject to boxes with windows in them, so that the content could be directly seen. The next step would be to create an illusion of content by pasting photographs, with good depth qualities, of the real content of the boxes behind the box windows. Would the subjects continue to peer into the windows and gain information from the photographs, even though their correspondence to the real content would only be on a categorical level? If this is possible, the next question would be whether performance would forever be tied to a reality mode, or if also successive abstraction of the pictures would be allowed within the competence. By expecting the content behind the glass to be informative, an interpretative stance will be taken and a pictorial concept of sorts might be created. This remains to be tested empirically.

12.4 Overall discussion, Givskud

Sherman and Austin in Savage-Rumbaugh (1986) could use photographs to communicate the content of baited containers. This was content which they in turn had been informed about through the means of lexigrams. However, when asked to directly match photographs to lexigrams, without a hidden food-problem to solve, they had great difficulties in selecting the correct photographs. It seems that the translation between lexigrams and photographs had been dependent on solving instrumental tasks. Their lexigram and picture use, seemed to have been grounded in action. What objects were *for* seemed to have been a most salient part in their conceptualisation. However, this dependence was not beyond repair Both chimpanzees were able to make context independent naming of objects with further exposure to such tasks. Sherman made the switch spontaneously and jumped from 43% correct to 96% in the middle of a session. This most probably indicated that he had been working according to one hypothesis that he suddenly switched for another.

Often hypotheses can be translated to the purpose of a task (Savage-Rumbaugh, 1986). The purpose of using photographs to access food that is depicted in those photographs is very different from the purpose of matching photographs to e.g. lexigrams that are arbitrarily selected by an experimenter. The latter case makes very little sense if one is not drilled in matching for matching's own sake. That apes make active hypotheses regarding experimental tasks is thus a source for rapid learning as well as major frustration. All those factors that render a fruitful hypothesis opaque can be called poor experimental design. "[…] the experimenter should recognize that every intelligence test is a test, not only of the creature examined, but also of the experimenter himself" (Köhler, 1925/1957, p. 226).

Harlow (1951) mentions several interfering factors that affect most learning experiments. For the results from Givskud "stimulus preferences" and "response preferences" are especially relevant. There is little support for the possibility that the gorillas developed preference for specific photographs, neither spontaneously nor as a result of reinforcement history. This is true for both the MTS and the object-choice task. This strongly suggest that they did indeed pay little attention to the motifs of the pictures. They seemed to have functioned as identical pieces of plastic.

Response preferences are "natural inclinations" such as pulling (food) rather than pushing it away from oneself. Such inclinations can contradict the requirements of a (poorly designed) experimental task. Although problems with inhibiting certain reaching responses were found in Barth and Call (2006), problems with perseveration reaching seem not evident in the Givskud data, unless one counts position preferences. When baiting was made in full view the number of correct responses was high. The subjects updated their view of the baited status of the containers in each trial, or most of them. And in the occluded trials the boxes were always shown to be empty before baiting took place.

However, the reaching errors in Barth and Call (2006) can possibly also be explained by a tendency to reach for the container that was closest to the subject's hand (see also Call, 2001). The Givskud setup is definitely sensitive to such tendencies. Especially since each box had a rather high chance (50%) of containing food, the urge to go for the closest box could have been difficult to resist indeed. The subjects typically responded immediately upon presentation of the picture cards or the boxes. Such swiftness might have increased the frequency of careless choices. When presentation was slowed down in order to facilitate attention to both choice stimuli, signs of impatience were common. It was evident that the subjects were set on retrieving their rewards (at least at a 50% level) as quickly as possible. This might have been a drawback from testing in the social group. Impatience during baiting seemed to increase when other individuals were close by or on the move. This was shown especially in the females through glances over their shoulders, knocking on the metal bars, and general bodily unrest when the experimenter was not swift enough.

But response preferences are not only part of a species' standard behavioural repertoire, but can also be created in the moment. As mentioned, there is for example evidence for positioning effects in the Givskud data. This is typical for the earlier parts of discrimination training (Smith et al., 1975).

However, *learned* response and stimulus preferences might be better sorted under the term "ambiguity of reward." That is, when a subject is reinforced, many things are rewarded at once. For example, in a choice task the object *and* its position are both rewarded. The manners in which the subject indicates a choice can likewise be rewarded. Amani, for example, took some time before she adopted the convenient use of her index finger. Instead she used her whole hand, which was really too large to fit smoothly between the bars. During periods she also twisted her whole arm so that her palm faced outwards and her thumb down when making her selection. It seemed as if she was trying out a new hypothesis about why she was sometimes rewarded and sometimes not. She had developed a redundant response preference. Amani also insisted on using sticks, sometimes grossly oversized ones, to make her choice. This was an even more inconvenient style of indication.

Why would the location of a picture be more salient for a subject than e.g. which hand the experimenter holds up the food with before the baiting, where the experimenter looks during baiting, the box that is presented a little bit sooner than the other, or any other potentially "important" variable?

12.4.1 Connectedness

Indexical effects were found both in the MTS study and in the object-choice task. The "more" indexical, the stronger the effects seem to have been. When nearness was maximised, as when sample and match were placed closest together, or a food item was held underneath a photograph rather than placed in a box under it, performance was boosted.

Maximised nearness borders on a part-whole relationship. In semiotics, indexical relationships are said to be based on contiguities (nearness) and factoralities (part of a whole) (e.g. Sonesson, 1989). A history of co-variation between items can form indexical relationships. But to attend to such a relationship poses different demands on the perceiver. For factorality one only needs to have experienced the whole and attended to the relevant parts. For example, we can associate the leaves on the ground to a tree because we usually see the leaves on the branches of trees. To appreciate a relationship based on contiguity, on the other hand, one needs to have experienced the specific circumstances that bridges two entities. The traces that an animal leaves in the snow, for example, are not part of the animal in the same way as is the leaves of a tree It is reasonable that relations between entities based on factorality are more easily appreciated than those based on contiguity. A reason for this is that things that are physically connected usually form one and the same entity, while things that end up together do not necessarily belong together.

In fact, Jarvik (1953) demonstrated one-trial learning in a discrimination task with chimpanzees, macaques, and a spider monkey (*Ateles sp.*), by using maximised contiguity, i.e. factorality of sorts. He simply dyed pieces of good and bad-tasting bread green and red. The animals learned the discrimination instantly. They also performed impeccably when discriminating cues, i.e. coloured squares, were pasted on top of, or covered, white pieces of bread. However, subjects fell back to chance performance when the colour cues instead were placed a mere millimetre in front of the bread rewards. The colour discriminations did also not transfer to a conventional setup with a Klüver board containing baited food wells, covered with coloured cards. According to Jarvik (1953) this striking effect of physical contiguity in the form of *connectedness* had not previously been mentioned in the learning literature.

Jarvik (1956) corroborated his findings by subjecting chimpanzees to the standard procedure of covered food wells, and to a condition where the reward, a peanut kernel, was placed in a depression and taped in place on the backside of the coloured plates that were used as well covers/choice stimuli. Importantly, the subjects were allowed to extract the food reward themselves from the plaques. The striking effect repeated itself. Subjects that learned the proper discrimination on a single trial in the "connected" condition, needed 89 to 197 trials to reach criterion when the reward was disconnected from the stimuli, i.e. was laying in the wells a few centimetres under the covers. Jarvik (1956) reasoned that in the disconnected condition attention wandered from the choice stimuli to the reward. Jarvik's findings can explain several confusing results in discrimination studies. For example did Harlow (1945a; 1945b, both in Meyer et al., 1965) find that macaque monkeys learned a discrimination easier when shape stimuli were laying directly over a well baited with food, than when stimuli were mounted on a piece of wood (i.e. a wedge) that in turn was placed over the well. Although attributed to differences in discrimination of objects (free shapes) versus patterns (shapes on backgrounds) at the time, the visual nature of the stimulus patterns did not really affect performance differentially (Meyer et al., 1965). Perhaps a more plausible explanation is that when the shapes, which were still three-dimensional and object-like, were mounted on wooden wedges, the stimuli were further disconnected from the rewards inside the wells.

That a reward is perceptually "part" of a choice object or picture in virtue of being glued to it might enhance the salience of a correct choice because it constitutes a case of factorality rather than contiguity. Diamond et al. (1999) found for human infants what Jarvik found for nonhumans. Nine to 12-month-olds failed a standard delayed non-match-to-sample (DNMS) test, in which they were rewarded for choosing the novel of two items. However, they succeeded in the test if the reward had been fixed with glue underneath the correct choice item. Neither temporal nor spatial proximity could compete with the effect of actual connectedness (Diamond et al., 2003). In fact, children of this age typically fail on DNMS regardless of amount of training, if there is no physical connectedness between choice item and reward. In Diamond et al. (2003), connectedness is mediated by a box from which the reward pops up. The choice item is positioned against the box, not against the reward. In this sense it is very similar to the Givskud box setup. But important differences are that at Givskud the pictures were not the choice items, merely the cues. What were indicated by the subjects were the boxes. Furthermore, the reward in Diamond et al.'s (2003) research is not the object that pops up, but the popping up itself. The child is never handed the little teddy penguin that hides in the box.

Similarly to the Givskud findings Gellermann (1933, in McClearn & Harlow, 1954), gave up discrimination training with two chimpanzees who failed to learn a simple discrimination after 500 trials. Stimuli had been mounted on the front of two boxes. Although the subjects were required to open the boxes themselves and retrieve the reward, the visual difference between the boxes apparently failed to be recruited in the task. However, when Gellermann changed the design and moved the intended stimuli from the front of the boxes to their lids, the chimpanzees learned the adequate discriminations in 50 trials. McClearn and Harlow (1954) attribute this difference to the fact that the relevant cues now occupied the same location as the response, i.e. the box lids. Note however, that Gellermann's subjects still learned significantly slower than the animals in Jarvik's "connected" conditions. This might be due to the fact that the relevant cues on the lids and the loose rewards inside the boxes were still disconnected, as in the standard procedures with stimuli covered food wells on a Klüver board. Less successful was Bierens de Haan (1927, interpreted in Meyer et al., 1965) who gave up his training of macaques on a simple size discrimination after 1 250 trials. In his case the stimuli had been mounted on top of two boxes, but the subjects could not touch these. Instead they reached into the boxes through holes in the front of the boxes.

In Diamond et al. (2003) it seemed crucial that the subjects perceived the reward (jack-in-box penguin), the actual box, and choice item in front of it, as one single object. The three parts were therefore fixed to each other and the child could remove neither the choice stimulus nor the reward. If the children had been allowed to manipulate the parts independently from each other it was believed that attention would shift from the whole to any salient part, e.g. the choice objects, and learning of the non-matching rule would break down.

At Givskud the test material was intentionally handled in front of the gorillas in order to make the structure of the task as transparent as possible. The subjects could clearly see that the box, the food placed inside it, the box lid, and the picture placed on top of it, were all separate objects. The illusion of a coherent device, and the causal links that this would entail, might unfortunately have been precluded by the very pedagogy that served to simplify the task. Just stacking on top of each other a hollow piece of wood with a food item in it, a flat piece of wood, and a plastic sheet, might not at all have communicated "this is a box with a picture on it, and they form an integrated whole."

In the box-task used at Givskud, contiguity between action and reward was further compromised by the middle hand of the experimenter, who had to displace the lid for the subject, collect any potential food item from the box, and hand this over to the subject. Although a chain of events that allows cause and effect analysis from the experimenter's perspective, this might not have been the case for the subjects. If a slight discontinuity between the lid of a food-well and its rewarding content is difficult to bridge, it is probable that a situation where even more steps are introduced between choice and reward is also detrimental.

The main reason one cannot say that the Givskud gorillas generally *fail* pictorial tasks is that one cannot know gorillas well enough, through the means that are presented here, to make such claims. Longitudinal work with apes, as in ape-language projects, has taught us that it can take considerable time to find out which manipulations to make, but that striking results can follow once they are made.

My experiences at Givskud will define my future attitude towards ape learning in experimental tasks. Needless to say, taking the subject's perspective is paramount for success. But "taking the subject's perspective" must be something more flexible than just an updated set of assumptions. I have learned not to stubbornly prevail in a setup that does not work.

Chapter 13

Bonobos and drawings

13.1 Subjects

The bonobos Kanzi (male, born 1980) and Panbanisha (female, born 1985) were reared at the Language Research Center at Georgia University, USA, and are currently residing at the Great Ape Trust of Iowa (GATI). Both individuals are competent in using lexigrams receptively and productively (e.g. Brakke & Savage-Rumbaugh, 1996; Savage-Rumbaugh et al., 1986), as well as comprehending spoken English (e.g. Brakke & Savage-Rumbaugh, 1995; Savage-Rumbaugh et al., 1993). Kanzi and Panbanisha have a vocabulary size of 150 and 179 symbols respectively (Savage-Rumbaugh et al., 2006), although e.g. Kanzi is believed to understand at least 1000 words (Segerdahl et al., 2006). Since they understand words flexibly, based on their functions in syntactic and semantic context, putting a definite number on their language competence is as impossible as putting a number on our own. Their lexigram keyboards, however, consist of roughly 380 symbols (*fig. 22*). Panbanisha and Kanzi are living in a social group of bonobos but have extensive contact with humans around the clock. They are engrossed in a joint "*Pan/Homo* culture" (e.g. Segerdahl et al., 2006). To put it simply, this way of life is the reason they speak at all. Kanzi, and to a lesser extent Panbanisha, are perhaps best known throughout the scientific community for their spontaneous acquisition of English and lexigram use.

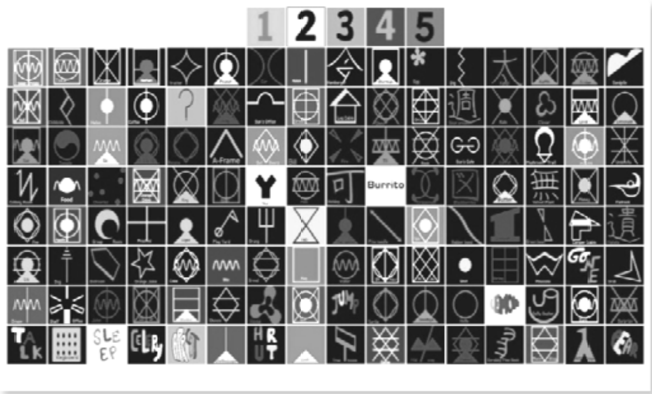


Figure 22. About a third of the lexigrams used with the bonobos at Great Ape Trust of Iowa. Note their arbitrary character.

Kanzi and Panbanisha have a long history of using photographs in daily life as well as in test situations. (Their lexigram board even includes the lexigram "picture.") For example the caption to a photograph of an infant Panbanisha in Savage-Rumbaugh and Lewin (1994) reads: "When the apes were too young to understand symbols, we often showed them photographs of the places we planned to go […].” This practice continued also when the bonobos and chimpanzees grew lexigram competent. For example, during the daily excursions into the forest surrounding the research facilities, Kanzi used photographs as well as lexigrams when queried about where they were heading in the forest, or what food was awaiting them there (Savage-Rumbaugh et al., 1998). Photographs are not only used to refer to places and objects, but also to people (regardless of species) and events.

Photographs are also frequently used in testing vocabulary. A case in point where Kanzi, at the tender age of about 3 years old, displays his competence with photographic stimuli, is Savage-Rumbaugh et al. (1988). In a cross-modal task inspired by Davenport and Rogers (1971), described in section 5.2, Kanzi picks out the correct photograph in an array of three in 100% of 54 trials when he is verbally asked for a picture of a specific object. Only well known object names were used and the procedure was of course blind to preclude cueing.

Much has been written about these individuals but tests specifically targeted at pictorial competence have never been given. During the summer of 2006 we conducted a set of picture experiments with the two most well-known language competent bonobos at GATI: Kanzi and Panbanisha.

Two types of matching experiments were conducted at GATI: A “receptive” condition, and a “productive” one. Both were sequential, i.e. the sample and the comparison stimuli (the match and non-matches) were not onscreen simultaneously. They were thus delayed MTS, i.e. DMTS. The receptive setup entailed first seeing a lexigram and hearing a voiced word and then be required to select the picture that corresponded to that category from an array of pictures. The productive setup was the other way around, i.e. first being shown a picture and then required to “name” that picture by selecting the matching lexigram. The receptive setup is the one commonly used in testing or training at GATI. This was therefore a natural choice. However, from the point of view of the questions addressed in this thesis there are limitations in such a setup. The main one is the possibly scaffolding role of asking for a particular motif. The referent is thereby placed in the mind of the ape without requiring that the pictures are spontaneously decoded. The subject only needs to find the most alike choice item from an array. It is difficult to address on what level likeness is appreciated in such a procedure. Does the subject rely on local cues or fully fledged object recognition? Some red and round objects are perhaps more likely matched to the word “strawberry” than are yellow and oblong shapes. In addition, in receptive mode the subject is potentially greatly helped by the combination of choice pictures and not only chose the “most alike,” but taking into consideration and excluding also the “least alike.” Red and round objects stand out as more “strawberry like” when surrounded by only non-round and non-red shapes. This requires, of course, that one is competent enough in a matching procedure to look for a match even when the choice items are ambiguous. In order to guess, or rule out an alternative, one needs to be aware of the fact that one is making a choice.

The use of a sequential “productive” setup deals with these issues, but has some disadvantages in terms of attentional demands and memory load. Because the picture is viewed before the “names” are given, and does not appear simultaneously, looking for the "most alike" in an array is not possible. The picture has to be recognised as an object before it can be named. There is an assumption involved here, however, in that a "reversed receptive" strategy will not be used. A reversed receptive strategy would be to store the picture in memory, without recognition, and try it against the choice names. "Is the image in my mind most like an apple? Banana? Strawberry?" The assumption is thus that what is stored in memory during the delay between sample and choice items is the *recognised* object, rather than a so far meaningless design. The extra attentional demands posed by a productive setup are to attend to the sample until recognition has occurred, and to keep this recognition in memory during the delay.

Neither the receptive nor the productive DMTS setup is limited to language competent subjects. In theory the use of lexigrams can be exchanged to photographs, given that the photographs have been proven to be recognisable by the particular subjects in independent tests.

13.2 Receptive pre-test

13.2.1 Materials and procedure:

Kanzi and Panbanisha had not been tested with drawings before the present study. It was therefore necessary to make a pre-test where novel drawings were interspersed with familiar pictures, in this case photographs with lexigrams pasted on them. Such photographs are used in everyday conversation with the bonobos. The choice of the photographs was made by experimenter Savage-Rumbaugh while the drawings were selected by me. Drawings were chosen that, from my subjective viewpoint, was particularly good examples from a collection of non-copyrighted clip-art (see *fig. 23*). All pictures were laminated in glossy plastic.

Figure 23. Drawings used in receptive pretesting. Bananas, celery, cherries and onions were in greyscale.

A request paradigm was used where the experimenter (Savage-Rumbaugh) placed three pictures in front of the subject and asked, verbally, him or her to indicate a specific picture (see *fig. 24,* p. 248). Cueing was precluded by not looking at the pictures but fixing the gaze at the face of the subject. If correct the subject got verbal praise and next set of pictures was presented. If incorrect the subject was again asked for the target picture until it had been selected. Not all trials included drawings.

Figure 24. The setup used for receptive testing of line drawings in physical pictures with bonobos Panbanisha and Kanzi.

13.2.2 Results, Receptive pre-test

Kanzi: Among the drawings in *fig. 23*, carrot, pineapple, bananas, cherries, apples, lemons, and tomatoes (no. 1) occurred as targets in Kanzi's session. The remaining drawings served only as non-matches. Kanzi received 25 trials and indicated the incorrect picture on four occasions. The target picture had been a drawing (lemon) on only one of these. Kanzi had chosen tea instead. (The bonobos often have lemon in their tea.) It is questionable whether the other mistakes were really mistakes since Kanzi took the opportunity to select pictures of food items during the testing that he wanted for his evening meal. He was correct on the first correction trial on all occasions. The main finding was thus that drawings of fruits and vegetables posed no problem for Kanzi.

Although lexigrams were pasted on the photographs, it is not certain that the bonobos used these to interpret pictures they did not recognise. On one question the experimenter misnamed the target photograph, mistaking sliced sweet potato for melon. The lexigram clearly said melon, but Kanzi choose, without hesitation, the sweet potato and was scored as correct. He either mistook the sweet potato for melon himself, or reasoned that the experimenter *must have meant* sweet potato. In either case he chose the picture that looked most like a melon, although it was not a melon. That is an expected performance by someone who can recognise line drawings.

Panbanisha: The drawings of pineapple, bananas, apple, strawberry, celery, tomatoes (no. 1), onions, lemons and kiwis (see *fig. 23*) occurred as targets in Panbanisha's session. Panbanisha received 23 trials and made no mistakes. Like Kanzi, Panbanisha also indicated what foods she wanted for supper (including the drawing of onions), but she did not let that interfere with testing. However, when the photographs depicted a specific brand of foods Panbanisha refused to make a choice until the experimenter had requested the specific brand in the picture (i.e. "Welch's grape juice" instead of just "juice"). The drawings used in the receptive pre-test did not pose any difficulties to Panbanisha.

13.3 Computerised “receptive” DMTS

A computerised touchscreen based MTS procedure developed at GATI was used, which both subjects were familiar with in the domain of receptive testing. Using the new touchscreen and software was not completely smooth, as the subjects had not fully calibrated according to the sensitivity of the screen which at times therefore seemed unresponsive. Also, sometimes the computer updated the display too slowly, which resulted in extra touches and error responses. Especially during training and testing in the unfamiliar productive setup was continuous verbal praise and small food rewards given to encourage Kanzi not to lose his confidence with the apparatus.

Receptive testing, on the other hand, went more smoothly and required less praise. The subjects typically signalled themselves when it was time for a food reward. That is, the motivation to perform correctly on single trials, or on a set number of trials, was not contingent on a food rewards, but was primarily due to the subjects’ willingness to perform well, and perhaps earn a reward eventually.

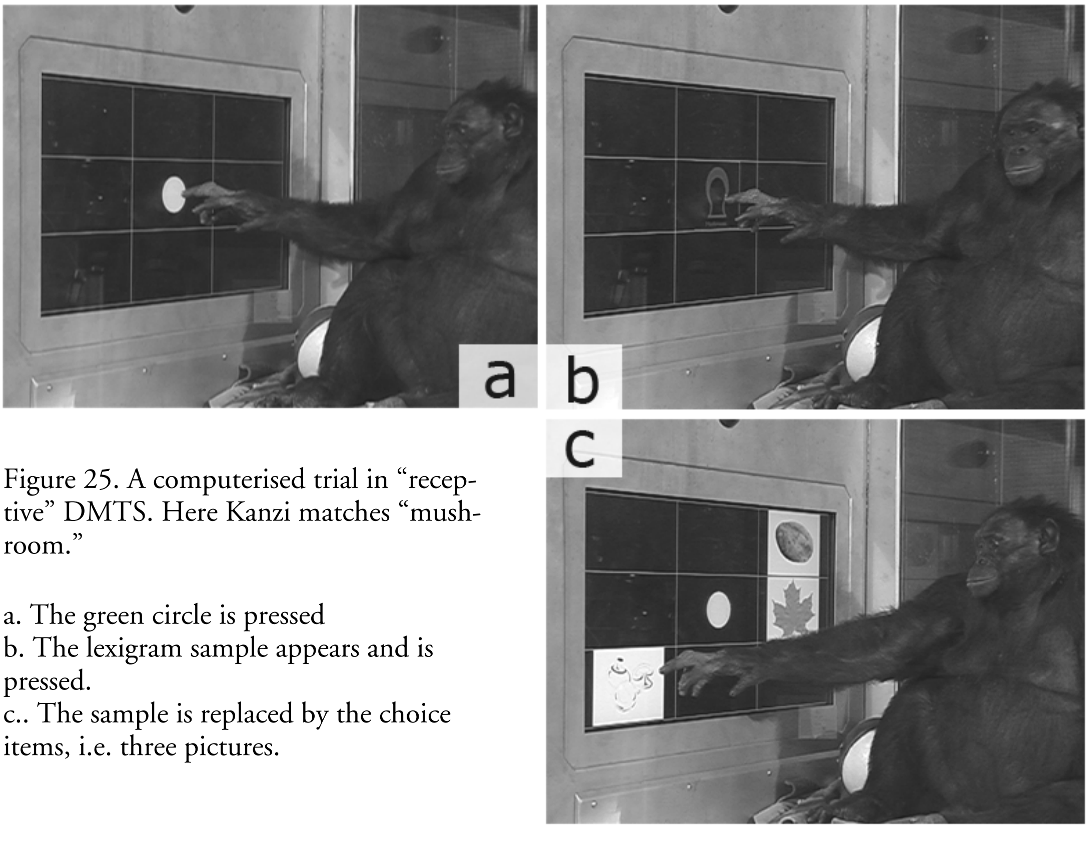


Figure 25. A computerised trial in “receptive” DMTS. Here Kanzi matches “mushroom.”

a. The green circle is pressed
b. The lexigram sample appears and is pressed.
c.. The sample is replaced by the choice items, i.e. three pictures.

A trial took the following form: A large green filled circle was presented in the middle square of a 3x3 array on the computerised touchscreen. After pressing the circle (see *fig. 25a*) it was replaced by a lexigram as well as a computerised verbal reading of that lexigram. After pressing the lexigram in turn the screen blacked out and three pictures were randomly distributed in the 3x3 array (see *fig. 25b*), of which one matched the meaning of the lexigram. Matches and non-matches could be drawings, colour photographs or black-and-white photographs of objects. They were drawn randomly by the computer from a pool of pictures, but naturally programmed to correspond to three different lexigram categories so that there was only one match.[151] Matches could thus also occur as non-matches on different trials. Together with the choice items the green circle reappeared in place of the lexigram. Pressing it restarted the trial by displaying the lexigram again. This was possible for a total of four trials before the sample lexigram and matching and non-matching pictures were randomly replaced with new stimuli. If the correct match was selected a computerised voice named the picture according to category and a musical tune was played. The screen blacked out and next trial commenced. In the receptive task a vending machine was implemented to yield a reward after a set number of trials, but since unreliable at the time food rewards, as well as verbal and physical encouragement, was given *ad libitum* by Savage-Rumbaugh. Cueing, however, was avoided at all cost. The experimenter was typically situated behind the subject, who was facing the touchscreen. If a non-match was selected the screen went back in silence to displaying the lexigram and a correction-trial commenced, to a possible total of four trials.

The subjects' use of the green circle during the choice phase was unclear (see section 13.4.1). Selection of the circle is therefore counted as a failure to match, and it will have to be included in calculations of chance levels although, strictly speaking, the reason to indicate a non-match or the green circle might be very different from the point of view of the subject. Chance performance was thus set to 25% correct (three pictures plus the green circle).

Photographs that were used in the receptive test had all been successfully identified by both subjects in a pre-test on physical laminated prints. The experimenter (Savage-Rumbaugh) placed an array of photographs in front of the subject, stood behind his or her back and asked for the different pictures. Kanzi and Panbanisha received this test somewhat differently in terms of procedure, but the result was the same. For a total of 72 novel photographs Panbanisha and Kanzi made two errors each. Drawings were chosen from non-copyrighted clip-art collections and were intended to represent a wide variety of styles. However, an active choice of not including particularly bad exemplars were made.

13.3.1 Results, Kanzi receptive DMTS

The following categories (lexigram samples) occurred as drawing and photograph targets in the test: apple, apricot, banana, bird, blueberries, carrot, celery, cherries, coconut, dog, egg, fire, grapes, hammer, honeysuckle, hose, key, kiwi, knife, leaf tree, lemon, lighter, marshmallow, melon, mushroom, onion, orangutan, peas, peanut, pineapple, pinecone, pine needle, potato, rock, rubber band, shoe, snake, spoon, strawberries, tomato, toothbrush, umbrella, water and watermelon.

Receptive testing was not preceded by training. In total, 60 pictures (28 drawings and 32 photographs) served as matches in Kanzi's receptive test. An additional 34

---

[151] Due to this completely random process, possibly helped by a bug in the software, fewer pictures than were intended for testing actually turned up as stimuli on the screen.

pictures served as non-matches that never occurred as matches. Kanzi was incorrect on 6 out of 112 trials (94.6% correct, X2 (1,112) = 103.4, p<0.001). Kanzi's incorrect matches occurred in the presence of five target pictures (three drawings, two photographs, see *tables 8*, p. 252, and *9*, p. 255). There was no significant difference between trials where a drawing was the match (93.5% correct, n = 46) and where a photograph was the match (95.5% correct, n = 66). Indicating the green circle accounted for two of the error responses. Only one error occurred in correction trials. Given the high success rate, looking separately at just 1st trial data is not meaningful.

13.3.2 Results, Panbanisha receptive DMTS

The following categories (lexigram samples) occurred as drawings and photograph targets in Panbanisha's computerised "receptive" DMTS session: banana, bird, blueberries, carrot, cherries, coconut, crayon, egg, fire, grapes, hammer, honeysuckle, hose, hotdog, key, kiwi, knife, lemon, lighter, marshmallow, melon, onion, orangutan, pineapple, pinecone, pomegranate, potato, rubber band, shoe, snake, spoon, toothbrush, umbrella and water.

As for Kanzi, training was not needed for Panbanisha in this variation of the DMTS task. In total, 44 pictures (16 drawings and 28 photographs) served as targets in Panbanisha's session. In addition, 37 pictures served as non-matches but did unfortunately never occur as matches in the randomisation process. Panbanisha was incorrect on 4 out of 67 trials (94% correct, X2 (1, 67) = 66.3, p<0.001). Three pictures (one drawing, two photographs) accounted for all of Panbanisha's incorrect responses (see *tables 8* and *9*). No significant difference was found between trials where drawings were the matches (95.7% correct, n = 23) and where photographs were the matches (95.5% correct, n = 44). Indication of the green circle accounted for one of Panbanisha's errors. One error occurred in correction trials.

When subjects show that they are aware of a choice situation (which will be argued in section 13.5 below) one must be wary of choices by exclusion, rather than true recognition of a match. Especially in receptive mode the non-matches are therefore not trivial in terms of their history as named pictures, and also whether they were photographs or drawings. See *table 10* (p. 261) for a full list of drawings that only appeared as non-matches in the receptive test (18 for Kanzi and Panbanisha each). None of them had been named in a previous session.

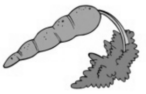
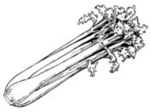
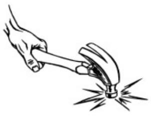
As seen in *table 8* and *9*, only a couple of them were chosen as the incorrect match to the lexigram plus voice samples. This means that the content of most of the comparison pictures was probably recognised. However, as for all the pictures in receptive testing this recognition could have been relative and scaffolded by the context set by the lexigram. Isolating the most e.g. "banana-like thing" in an array of three things might not require a full recognition of pictorial content. The bonobos could also have been helped by recognising some of the pictures and then make choices based on exclusion (see Beran & Washburn, 2002).[152] For this reason a “productive” version of the DMTS task was given.

Table 8 (below). “Receptive” testing. Delayed matching of pictures to sample “lexigram” + voice. Drawings. (+) Picture used and named in a previous session, (p) = only for Panbanisha, (k) = only for Kanzi. (bw) Drawing presented in black and white. (n) = number of exposures as match.

	Kanzi			Panbanisha		
”Lexigram” + target pic.	% corr. (n)	Corr. on first trial	Misidentifications	% corr. (n)	Corr. on first trial	Misidentifications
 “Pineapple” (+)	100 (1)	Yes				
 “Strawberry” (+, p)	100 (1)	Yes				
 “Apple” (bw)	100 (1)	Yes				
 “Banana” (bw)	100 (1)	Yes				
 “Banana”	100 (4)	Yes				
 “Bird”				100 (1)	Yes	
 “Bird” (bw)	100 (1)	Yes				

---

[152] Given the high success rate this point is moot, but worrying in principle. Furthermore, one want to be able to say exactly which individual pictures were recognised and which were not.

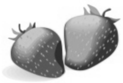
 "Carrot"	50 (2)	No (2nd)	 Onion	100 (1)	Yes	
 "Carrot"	100 (3)	Yes				
 "Carrot"	50 (2)	No (2nd)	*Green c.*			
 "Celery" (bw)	100 (2)	Yes				
 "Celery" (bw)	100 (1)	Yes				
 "Cherries" (bw)	100 (2)	Yes				
 "Cherries"	100 (1)	Yes		100 (1)	Yes	
 "Egg"	100 (2)	Yes		66.7 (3)	No (2nd)	*Green c.*
 "Fire"				100 (1)	Yes	
 "Hammer" (bw)	100 (2)	Yes				

 “Key” (bw)	100 (1)	Yes	100 (1)	Yes
 “Kiwi”	100 (1)	Yes		
 “Knife”			100 (1)	Yes
 “Lighter”	100 (1)	Yes	100 (2)	Yes
 “Mushroom” (bw)	100 (2)	Yes		
 “Onion”	100 (1)	Yes	100 (1)	Yes
 “Orangutan” (bw)	100 (1)	Yes	100 (1)	Yes
 “Pineapple”	100 (1)	Yes		
 “Potato”	100 (1)	Yes	100 (1)	
 “Shoe” (bw)			100 (3)	Yes

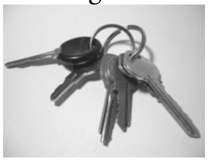
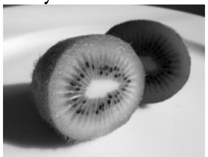
 “Shoe” (bw)	100 (2)	Yes				
 “Snake”	100 (3)	Yes		100 (2)	Yes	
 “Toothbrush” (bw)	66.7 (3)	Yes	 Carrot	100 (2)	Yes	
 “Umbrella”	100 (1)	Yes		100 (1)	Yes	
 “Water”	100 (2)	Yes				

Table 9 (below). Receptive testing. Delayed matching of pictures to sample “lexigram” + voice. Photographs. All photographs except “leaftree” and “apricot” (Panbanisha) and “apple” and “cherries” (Kanzi) had also been successfully identified at first trial in a pretest using physical prints and verbal requests. (+) Picture used and named in a previous session, p = only for Panbanisha, k = only for Kanzi.(bw) Photgraph presented in black and white. (n) = number of exposures as match.

	Kanzi			**Panbanisha**		
”Lexigram” + target pic.	% corr. (n)	Corr. on first trial	Misidentifications	% corr. (n)	Corr. on first trial	Misidentifications
 “Apple”	100 (2)	Yes				
 “Apricot”	100 (3)	Yes				

 “Banana”			100 (2)	Yes	
 “Banana”			100 (1)	Yes	
 “Bird”	100 (3)	Yes			
 “Blueberries”	100 (1)	Yes	100 (1)	Yes	
 “Coconut” (bw)	100 (2)	Yes	100 (1)	Yes	
 “Banana” (bw)	100 (1)	Yes			
 “Cherries” (bw)	100 (2)	Yes			
 “Grapes” (bw)	100 (4)	Yes	66.7 (6)	Yes	 Strawberries x2
 “Mushroom” (bw)	100 (1)	Yes			

 “Orangutan” (bw)				100 (1)	Yes
 “Pineapple” (bw)				100 (2)	
 “Tomato” (bw)	100 (4)	Yes			
 “Carrot”	100 (2)	Yes		100 (1)	Yes
 “Celery”				100 (1)	Yes
 “Crayon”				100 (1)	Yes
 “Dog”	100 (1)	Yes			
 “Fire”	80 (5)	No (2nd)	 Dog		
 “Hammer”	100 (2)	Yes		100 (2)	Yes

 “Honeysuckle”	100 (2)	Yes		100 (1)	Yes
 “Hose”	100 (2)	Yes		100 (1)	Yes
 “Hotdog”				100 (2)	Yes
 “Key”				100 (1)	Yes
 “Kiwi”	100 (1)	Yes		100 (1)	Yes
 “Knife”	100 (1)	Yes		100 (1)	Yes
 “Leaftree”	100 (5)	Yes			
 “Lemon” 1)	100 (1)	Yes			
 “Lighter”	33.3 (3)	No (3rd)	*Green c.*  Water	100 (1)	Yes

 “Marshmallow”	100 (1)	Yes	100 (3)	Yes	
 “Melon”	100 (2)	Yes	100 (1)	Yes	
 “Onion”	100 (3)	Yes	100 (1)	Yes	
 “Peas”	100 (1)	Yes			
 “Peanuts”	100 (1)	Yes			
 “Pinecone”	100 (3)	Yes	66.7 (3)	Yes	 Peanuts
 “Pineneedles”	100 (2)	Yes			
 “Pomegranade”			100 (3)	Yes	
 “Rock”	100 (1)	Yes			

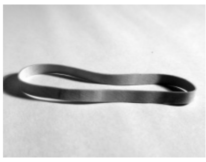
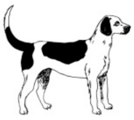
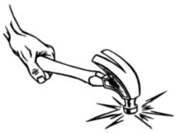
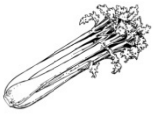
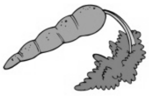
 "Rubberband"	100 (2)	Yes	100 (1)	Yes
 "Shoe"			100 (2)	Yes
 "Spoon"	100 (1)	Yes	100 (1)	Yes
 "Toothbrush"			100 (1)	Yes
 "Water"			100 (1)	Yes
 "Watermelon"	100 (1)	Yes		

Table 10 (below). “Receptive” testing. Delayed matching of pictures to sample “lexigram” + voice. Drawings that only occurred as non-matches. (bw) = black and white. (+) = picture used and named in a previous session.

Kanzi	 Apple (bw)	 Bird	 Knife	 Spoon
	 Apple	 Dog (bw)	 Lemon	 Tomato
	 Apple (bw)	 Fire	 Shoe (bw)	 Tomato (bw)
	 Banana (bw)	 Grapes	 Strawberry (bw)	 Tomato (+)
Panbanisha	 Apple (bw)	 Carrot (bw)	 Hammer (bw)	 Strawberry
	 Apple (bw)	 Celery (bw)	 Kiwi	 Strawberry (bw)
	 Banana (bw)	 Celery (bw)	 Mushroom (bw)	 Tomato (bw)
	 Banana	 Cherries (bw)	 Lemon	 Water
	 Carrot	 Dog	 Spoon	

13.4 Computerised “productive” DMTS

The same touchscreen system as in the previous experiment was used, but with different programming. This time the sample was a picture, either a photograph or a drawing, randomly picked by the computer from a pool of pictures. The sample was always presented in the middle square of the 3x3 touchscreen display. In order for the sample to come on-screen a large green filled circle had to be pressed. After the sample, in turn, had been pressed, the sample disappeared and the green circle returned in centre position. Simultaneously in the remaining 8 squares three choice items came on-screen, randomly distributed (see *fig. 26 a,b,c*)

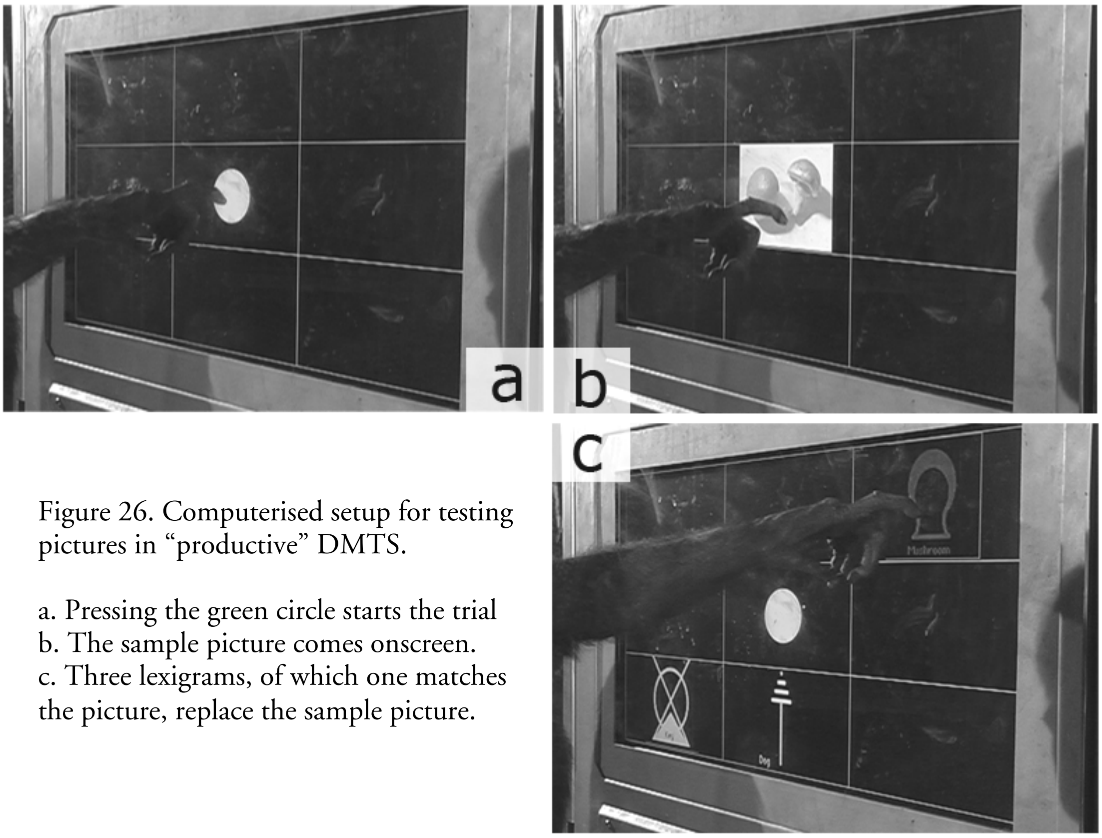


Figure 26. Computerised setup for testing pictures in “productive” DMTS.

a. Pressing the green circle starts the trial
b. The sample picture comes onscreen.
c. Three lexigrams, of which one matches the picture, replace the sample picture.

In the productive condition choice items were lexigrams, the non-matches randomly picked from the complete pool of nouns for pictures included in the test. The chosen lexigram, if correct, is read out by a pre-recorded voice and the screen is completely blacked out, including the green dot. If correct choice has been indicated the tune is played and next trial commences. If the choice was incorrect, the trial is repeated in silence (no naming of the lexigram), to a maximum of four attempts (i.e. three correction trials). That is, following blackout of the choice items the sample immediately appears again and needs to be pressed in order to see the choice items again, which are presented together with the green dot where the sample was. Sample and matches are thus never onscreen at the same time, i.e. a case of delayed matching to sample. On correction trials the choice items came onscreen in the same configuration as on the preceding trial.

Reused and novel photographs were randomly mixed with reused and novel drawings. The drawings were in a variety of styles (see *table 11*). The collection put into the randomisation program was considerable larger, but not all pictures were run in the automatic process during the trials given.

Since the subjects had not performed this type of matching in many years (Savage-Rumbaugh, pers. com.) training trials were included to make sure that they grasped the task. A criterion was not set. Rather, the readiness to move on to testing trials was decided by experimenter Savage-Rumbaugh. Only photographs were included in training, familiar and new ones. Some were re-used in subsequent testing. A combination of familiar and novel stimuli increases the chance of good performance, and consequently motivation, of the subjects.

Lexigrams (categories) used in the two sessions were: apple, banana, bird, carrot, celery, cherries, dog, fire, grapes, key, kiwi, knife, lemon, mushroom, onion, orangutan, pineapple, potato, shirt, shoe, snake, strawberries, tomato, umbrella.

13.4.1 Results, Kanzi productive DMTS

Kanzi made 69 errors in total (202 trials), which equals 65.8% correct.

54 pictures (26 drawings, 28 photographs) were tested of which 20 were incorrect on first exposure within the test. 11 of these were drawings. Taken together, 30 pictures accounted for all of Kanzi's errors. Of these 16 were drawings. There was thus no significant difference in performance on drawings and photographs when looking at the complete session, including training. However, it makes little sense to compare drawings and photographs in the total score since only photographs were included in training and some of them reoccurred in subsequent testing. Kanzi was correct on 71% of 31 training trials with photographs.

Kanzi made 60 errors in his 137 post-training trials = 65.3% correct (Photographs: 69.1%, Drawings: 62.5%).

He did not seem to have used the green circle in order to refresh his memory. Six of his thirteen uses occurred during training trials and in all of these he pressed the spot as a result of the touchscreen not reacting instantaneously to his indications (i.e. Kanzi had as yet not calibrated to the occasional lapses in the speed of the screen) or as a result of inattention to the fact that the sample had come onscreen. In the seven uses of the green circle during test trials four can again be attributed to Kanzi being out of sync with the touchscreen presentation. On three occasions, however, the choice of the green circle seems to have been more intentional. In two of these the sample was the bird drawing seen in *table 11* (p. 265), and in the third a photograph of an orangutan. The orangutan case had been preceded by an incorrect choice and was followed by a correct one. In the first bird case correct choice followed indication of the green circle and subsequent re-presentation of the sample. In the second bird case the correct choice followed after an incorrect choice. It can thus not be concluded that Kanzi used the green circle to refresh his memory of the sample (or that he needed to), or in any other way indicated that he understood the function of the circle. Pressing the green circle instead of a choice item has therefore been scored as an incorrect choice in the present calculations, as well as being included in the calculations of chance performance. When including the green circle there was thus a 25% chance of guessing the correct answer among the three choice items plus the circle.

There are two clear cases when Kanzi did not seem to pay attention when the sample picture was presented. However, since he still chose one of the choice items (rather than the green circle) these are scored as incorrect answers at first exposure. Luckily both images had been included in Kanzi's testing at earlier sessions, on which he had been correct on them both. Valuable data is therefore not lost by scoring Kanzi conservatively. The pictures were a drawing of keys and tomatoes respectively (see *table 11*). Both were named correctly on Kanzi's first correction trial.

Excluding correction trials and only looking at first-trial data, Kanzi made 48 errors in total on 133 trials = 63.9% correct.

Of these were 112 post training trials. On these he made 41 errors = 63.4% correct ($X^2(1, 112) = 33.5$, $p<0.001$).

Separating drawing and photograph trials (1st trial data, post-training), Kanzi made 25 errors on 63 trials with drawings = 60.3% correct (X^2 (1, 63) = 16.1, $p<0.001$), and 16 errors on 49 trials with photographs = 67.3% correct (X^2 (1, 49) = 17.7, $p<0.001$).

Difference between performance on drawings and photographs in productive mode is not significant ($X^2 = 0.59$, $p>0.1$).

13.4.2 Results, Panbanisha productive DMTS

Counting all trials, Panbanisha made 39 errors in total, distributed over 161 trials = 75.8% correct

49 pictures (25 drawings, 24 photographs) were tested, of which 6 were incorrect on first exposure within the test. Three of these were drawings. Among the 3 photographs that were incorrectly matched on first exposure, 2 were of birds (see *table 12,* p. 269*).*[153] 17 pictures accounted for all of Panbanisha's errors. Of these, 6 were drawings. Drawings did thus not pose more difficulties than did photographs in this experiment. This holds true also when excluding the training trials. (Panbanisha was 54.2% correct in training trials with photographs. Chance was 25% correct. 8 of 11 of her mistakes in the 24 training trials were elicited by one particular colour photograph of a bird.)

Panbanisha made 28 errors in 137 post-training trials = 79.6% correct (Photographs: 78.8%, Drawings: 80.3%).

---

[153] One of the birds was in training trials and can therefore not be found in *table 12.*

Panbanisha seemed to use the green circle differently from Kanzi. In total she used it 13 times, of which only three can be attributed to inattention or touching the screen out of tune with its speed and sensitivity. In 7 of the 13 trials Panbanisha touched the correct choice in the first correction trial. These all, except one, occurred in the later part of testing which might indicate that she had discovered that she could view the sample again if she was unsure of the correct choice. Her use of the green circle was elicited by only three categories of pictures: umbrella, lemon and bird, represented by two pictures each. A lemon drawing (see *table 11*) accounted for most of the uses. Since it is not clear that Panbanisha used the green circle to exclusively refresh her memory the circle has been included in one of the choice items in calculations of chance performance (25%). This also makes the results more comparable to Kanzi's.

Panbanisha's first-trial data (excluding correction trials) amounts to 126 trials, on which she made 21 errors = 83.3% correct.

Looking solely at first-trial, post-training trials, Panbanisha made 17 errors on 111 trials = 84.7% (X^2 (1, 111) = 79.8, $p<0.001$) correct.

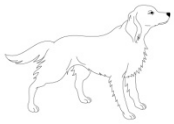
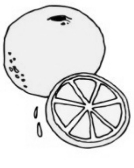
Of these are 7 errors made for 59 drawing trials = 88.1% (X^2 (1, 59) = 47.9, $p<0.001$) correct, and 10 errors for 52 photograph trials = 80.8% (X^2 (1, 52) = 32.5, $p<0.001$) correct

Difference between performance on drawings and photographs in productive testing is not significant for Panbanisha (X^2 = 1.16, $p>0.1$).

Table 11. (below). "Productive" testing. Delayed matching of "lexigrams" to sample picture. Drawings. (+) Picture used and named in a previous session, p = only for Panbanisha, k = only for Kanzi.(bw) drawing displayed in black and white. (n) = number of trials as sample. 1) No attention paid to sample on 1st trial.

	Kanzi			Panbanisha		
Sample Drawing	% corr.(n)	Corr. first trial	Misidentifications	% corr. (n)	Corr. first trial	Misidentifications
 Fire (+, p)	37.5 (8)	No (3rd)	"Snake" "Strawberry" "Cherries" "Onion" "Lemon"	100 (1)	Yes	
 Keys (+, bw)	66.7 (3)	No (2nd) 1)	"Banana"	100 (2)	Yes	

 Knife (+, p)	66.7 (3)	Yes	"Straw-berry"	100 (2)	Yes	
 Apple (+)	100 (2)	Yes		100 (1)	Yes	
 Banana (+, bw)	100 (2)	Yes		80 (5)	Yes	"Bird"
 Carrot (+, k)	60 (5)	Yes	"Shoe" "Shirt"	100 (5)	Yes	
 Celery (+, p, bw)	75 (4)	Yes	"Lemon"	100 (1)	Yes	
 Cherries (+, bw)				100 (2)	Yes	
 Kiwi (+, p)	100 (2)	Yes		100 (2)	Yes	
 Lemon (+)	100 (2)	Yes				
 Onion (+, p, bw)	33.3 (3)	No (3rd)	"Key" x2	100 (1)	Yes	
 Strawberry (+, p)	33.3 (3)	No (3rd)	"Dog" "Kiwi"	100 (1)	Yes	

 Tomato (+)	66.7 (3)	No (2nd) 1)	*Green c.*	100 (2)	Yes	
 Orangutan (+,bw)	75 (4)	Yes	“Dog”	100 (3)	Yes	
 Potato (+)				100 (2)	Yes	
 Apple (bw)	100 (4)	Yes		100 (2)	Yes	
 Banana	100 (2)	Yes		100 (4)	Yes	
 Bird (bw)	46.2 (13)	No (2nd)	“Onion” “Kiwi” “Carrot” “Knife” “Snake” *Green c.* x2	33.3 (6)	Yes	“Dog” x3 *Green c.*
 Dog (bw)	50 (4)	No (3rd)	“Apple” *Green c.*	100 (2)	Yes	
 Grapes	50 (4)	No (2nd)	“Apple” “Umbrella”	100 (2)	Yes	
 Lemon	42.9 (7)	Yes	“Key” “Dog” “Umbrella” “Grapes”	14.3 (7)	No (7th)	“Orangu-tan” *Green c.* x5

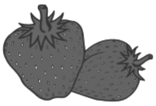
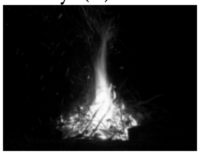
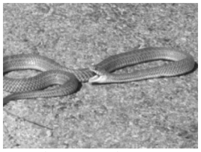
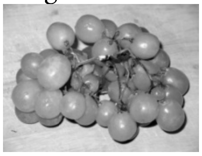
 Mushroom	40 (5)	No (4th)	“Snake” x2 “Kiwi”	66.7 (3)	No (2nd)	“Tomato”
 Pineapple	100 (2)	Yes		100 (5)	Yes	
 Snake	100 (1)	Yes		100 (3)	Yes	
 Strawberries	100 (5)	Yes				
 Shoe (bw)	100 (1)	Yes		66.7 (3)	No (2nd) 1)	*Green c.*
 Shirt	57.1 (7)	No (2nd)	“Shoe” x2			
 Umbrella	40 (5)	No (3rd)	“Key” “Tomato” “Apple”	75 (4)	Yes	*Green c.*

Table 12 (below). "Productive" testing. Delayed matching of "lexigrams" to sample picture. Photographs. (+) Picture used and named in previous session, p = only for Panbanisha, k = only for Kanzi. (bw) = black and white. (n) = number of trials picture served as sample.1) No attention paid to sample on 1st trial.2) Picture reused from receptive testing, but was then in a grayscale version.

	Kanzi			Panbanisha		
Sample photo (post-training)	% corr. (n)	Corr. first trial	Misidentifications	% corr. (n)	Corr. first trial	Misidentifications
 Carrot (+)	45.5 (11)	Yes	"Kiwi" "Grapes" x2 "Celery" "Dog" "Key"	80 (5)	Yes	"Strawberry"
 Celery (+)	100 (4)	Yes		100 (2)	Yes	
 Fire (+)	100 (2)	Yes		83.3 (6)	Yes	"Orangutan"
 Key (+)				100 (2)	Yes	
 Kiwi (+)	66.7 (3)	No (2nd)	"Knife"	100 (2)	Yes	
 Knife (+)	100 (2)	Yes		33.3 (3)	No (3rd)	"Dog" x2
 Onion (+)	100 (1)	Yes		100 (1)	Yes	
 Pineapple(+,p)	100 (2)	Yes		100 (2)	Yes	

 Shoe (+)	50 (4)	No (2nd)	"Snake" "Knife"			
 Snake (+, k)	100 (2)	Yes		100 (1)	Yes	
 Strawberry (+, p, bw)	66.7 (3)	No (2nd)	"Dog"	100 (2)	Yes	
 Tomato (+, p)	80 (5)	Yes	*Green c.*	80 (5)	Yes	"Strawberry"
 Bird (bw)				50 (2)	No (2nd)	*Green c.* x2
 Dog				100 (3)	Yes	
 Grapes	100 (1)	Yes		100 (5)	Yes	
 Lemon	50 (2)	No (2nd) 1)	"Dog"	62.5 (8)	Yes	"Carrot" *Green c.* "Pineapple"
 Mushroom (+,k) 2)	100 (2)	Yes		60 (5)	Yes	"Grapes" x2
 Orangutan	33.3 (9)	No (2nd)	"Celery" *Green c.* "Tomato" x2 "Apple"			

 Potato	100 (1)	Yes		100 (4)	Yes	
 Shirt	100 (3)	Yes		100 (1)	Yes	
 Umbrella	71.4 (7)	Yes	“Strawberry” “Kiwi”	60 (5)	Yes	“Fire” x2

13.5 Error analysis

The question, then, is: When the two bonobos did not match according to the intentions of the design, what did they do instead? An incorrect reply is still a reply. In the misidentification column of *table 11 (*p. 265) it is difficult to see a systematic pattern. However, for Panbanisha a couple of her misidentifications might have been of an iconic type, but between the picture and lexigram as such. For example she used the lexigram for dog (see *fig. 27,* p. 272) to name the photograph of a knife (see *table 12,* p. 269). Also, the lexigram for fire (see *fig. 27*) was used to name the photograph of a red umbrella (see *table 12*).In both cases were the wrong lexigram indicated on two consecutive trials before Panbanisha switched to the correct one. This suggests that her error was not a random guess but that she had a reason for her choice.

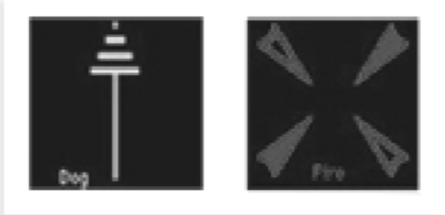


Figure 27. Possible iconic mistakes in response to a knife picture (lexigram “Dog,” left), and a red umbrella (lexigram “Fire,” right).

Lyn (2007), in her analysis of errors, reports that Kanzi and Panbanisha sometimes confuse two lexigrams due to visual similarity, or misidentify photographs. But translating photographs to lexigrams on an iconic basis is not reported.

When Kanzi indicated the wrong choice he often tested the remaining choices in succession. Systematic avoidance of previously incorrect choices suggests that subjects are aware of the fact that there is a choice involved and that only one of the choice items is correct. This is important because only when the subjects display that they know that there is a choice involved can we specifically invoke *guessing* as an explanation for seemingly random choices. The alternatives are choices according to reinforcement history, or some perceived cause and effect logic applicable to the task.

Given that Kanzi was motivated to perform correctly, going through the choices systematically indicates that for some pictures he truly did not know which match to choose. In the cases where Kanzi only needed one correction trial we can be less certain of this interpretation, but he used two correction trials to successfully eliminate the non-matches on at least eight occasions. This occurred for six pictures, of which only one was a photograph. For the drawing of fire and a lemon (see *table 11,* 265) it occurred twice. In contrast, Kanzi never needed to use this strategy in the receptive test, where he was always correct on first correction trial (see *tables 8,* p. 252, and *9,* p. 255*)*.

Panbanisha's errors looked somewhat different. Rather than changing her mind when in error she often repeated the incorrect choice, perhaps being stubborn and mistrusting the computer. This occurred on at least four occasions, excluding training trials and repeated use of the green circle. Two of the cases are the ones discussed above, where she seemed to match by the look of the lexigrams (dog for knife and fire for umbrella). The other two can be misidentifications. They were the repeated use of "dog" for a drawing of a bird (see *table 11,* p. 265*)* and "grapes" for a photograph of mushrooms (see *table 12*, p. 269). Panbanisha made only one repeated error of this kind in the receptive test, which was matching twice a colour drawing of strawberries to the sample lexigram "grapes," rather than the target black and white photograph of grapes. This is an understandable error.[154]

13.6 Free-response, Matata

The lack of free response data from Givskud made it the more interesting to gather such data from the bonobos at Great Ape Trust of Iowa. The spontaneous behaviour towards photographs of edible and inedible objects was informally tested with the adult bonobo Matata at GATI.[155] Matata was born in the wild in 1970 and came to the Yerkes laboratories in 1975 where Savage-Rumbaugh first met her. From 1980 Matata was involved in language-research at the Georgia State's Language Research Center, but showed slow learning of the lexigrams and was dropped from language research after about 5 years. She continues to participate in research as a non-linguistic control subject. Looking in magazines is described as one of Matata's favourite pastimes. She is also said to sometimes taste pictures of favourite fruit (Savage-Rumbaugh & Lewin, 1994).

In the session reported here (August 11th, 2006), Matata was handed, in each trial, a collection of 4 pictures, depicting various fruit and non-fruit objects. The fruit-pictures were photographs and colour line-drawings. The non-fruit objects were photographs (*fig. 28*). In a given trial all food pictures depicted the same type of fruit, but in various styles. All pictures of inedible objects, and many of the pic-

---

[154] That colour is not a necessary feature for recognition is not surprising. It is rather more surprising that recognition (of drawings) occur *despite* colour since the reproduced colours in most cases differ markedly from colouration of natural objects.

[155] This testing was conducted with the assistance of Takashi Yoshida.

tures of edible ones, were cut-outs against a white background with a strong pop-out feel to them.

A trial usually developed in the following way: After receiving the pictures, all at once and faced down, Matata either browsed through them directly by the mesh, or put them between her lips and carried them away to browse elsewhere. After some coaxing she returned and handed the pictures back one at a time. In return she received a piece of the same food that was displayed on the picture handed out. She always browsed the stack of pictures systematically, leaving no picture un-viewed. Some she mouthed, bit, sniffed, or visually inspected closer. Others were immediately dropped to the floor after just a casual glance. Five trials with unique sets of pictures were given in the session. Matata's spontaneous behaviour was recorded by handheld video camera and notes were taken which were later checked and corrected against the video recordings.

The five trials given in the session totalled the use of 20 pictures. Of these 6 pictures got bitten, targeted at the fruit in the picture. Only one of the 7 non-fruit photographs received oral contact, which was the brightly yellow flower (*fig. 28*). Of the bitten fruit pictures, 3 specifically interesting ones are shown in *fig. 29 (*p.274). Thus, not only photographic fruit was returned with bite marks. The picture with an apple in a pair of hands was the only one where the food was not presented in isolation.

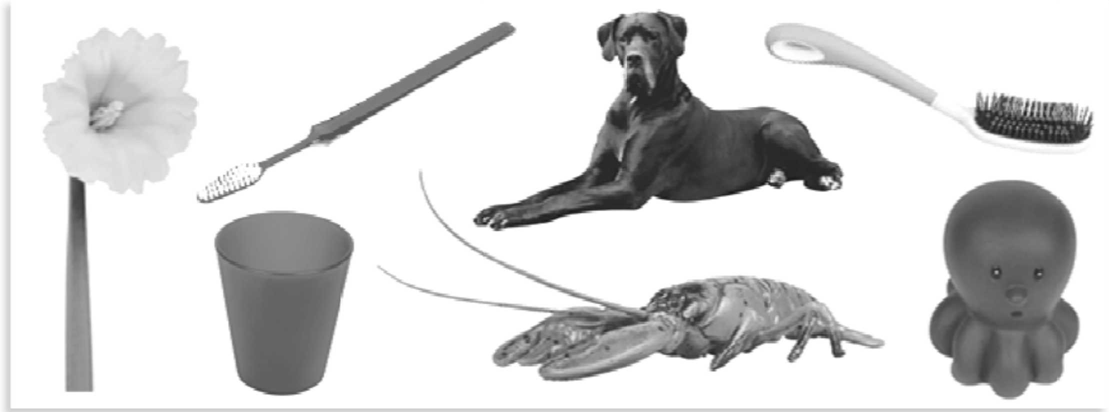


Figure 28. "Non-food" photographs used in a session of free response towards pictures in the bonobo Matata. Only the flower got "tasted."

The pictures included here are not copies of the actual test material. Bite marks can therefore not be seen. But the banana at the front cover of this book was included in a session with Matata and two of her children. It is unclear who is responsible for the damage to the picture. Besides the ripped out part there are also several bite marks targeted specifically at the banana. Glossy lamination was used and it seemed clear that they bit open the plastic to get to the content underneath. In a replication trial with only Matata in a similar session to the one described above, a new exemplar of the same banana photograph was offered. Matata bit into it several times, targeting the banana.

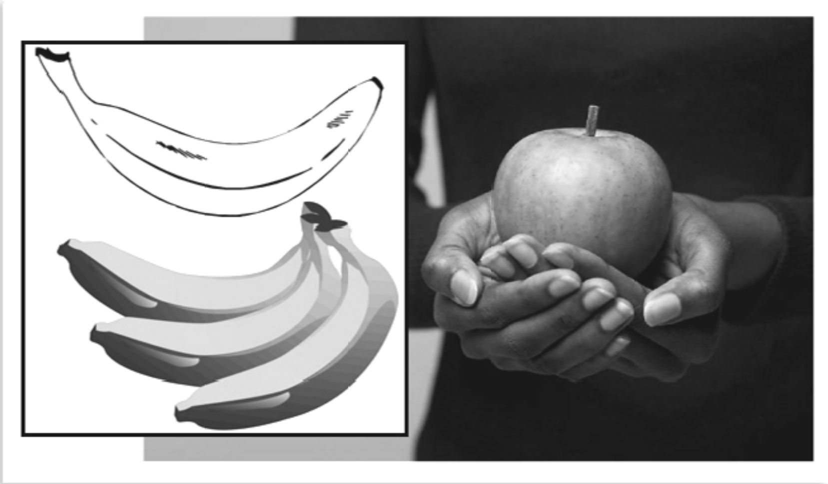


Figure 29. Three food pictures that were "tasted" by Matata in a session of free response towards pictures. Three additional photographs got bitten into.

13.7 Discussion, bonobos and drawings

Both Panbanisha and Kanzi performed exceedingly well with novel pictures that were low in realism. However, for Kanzi ($X^2 = 32.98$, $p<0.001$), but not for Panbanisha ($X^2 = 3.51$, $p>0.05$), there was a significant difference between performance in receptive and productive types of testing. Does this difference reflect pictorial ability, or differences in the DMTS task demands? The bonobos seemed for example to have problems with the sequencing of the task, i.e. getting to grips with what was matched with what. The trials tended to blend together when eager bonobo fingers got out of synchrony with the responsiveness of the touchscreen.

Kanzi did not improve during productive testing. He made 12 errors on his last 25 trials and 12 errors on his first 25 trials following training. Panbanisha, however, got better with time. She made 8 errors in her first 25trials, and only 2 in her last 25.

In Kanzi's case the overall result is congruent with a constructive view of picture decoding, where the absence of contextual constraints complicates recognition and interpretation. However, there was no significant difference between photographs and drawings in either of the experiments. If drawings specifically need an a priori view of pictures as representational in order for interpretation to take place, there would be a stronger difference between photographs and drawings in productive mode. Because this was not the case, Kanzi's difficulties with matching in the productive setup cannot be attributed to problems with interpreting drawings. However, it could possibly be attributed to problems with pictures in general in the absence of contextual scaffolding. But given the novelty of the experimental setup that conclusion is premature.

In Daehler et al. (1979), when 2-year-old humans were to match pictures to pictures, verbal labelling had a strong scaffolding effect. This effect was limited to boys, who almost doubled their success rate when the experimenter named pictures, as in "can you find another X," as opposed of pointing to a picture and say "can you find another one of these?" In the same study the children were significantly better at matching objects to objects than to make any type of match that involved a photograph of the sample object. This suggests that it was the appreciation of the pictorial motif that was facilitated by naming, at least for young boys.

In contrast to a socio-interactive case, note that the scaffolding context that *might* (it remains an empirical question) have produced better matching in the receptive than in the productive MTS test for Kanzi, was not primarily a social one. It was lexigrams on a computer screen, or the computerised voicing of them, that narrowed down reference. It was the words themselves that seemed to set the context and not the involvement of people in social constellations performing meaningful actions on objects in referential triangles. If it is suggested that it is the private manipulation of such context that allows us to "read" pictures in situations that lack external scaffolding, pictures can perhaps be a tool for studying internalised language in language apes.

A reason to suspect that there was a difference between the two types of experiments (but still not necessarily a pictorial one) was that the subjects seemed to make different errors in the two setups, going through more incorrect matchings, especially Kanzi, before hitting upon the correct match in the productive version. Panbanisha's "iconic errors" (see *fig. 27,* p.271) hints at a suggestive implication. These bonobo's can be argued to already be pictorial in their work with lexigrams. They recognise lexigrams flexibly appearing on various surfaces, including crisp ones on cardboard, and highly pixelated ones on a computer touchscreen. If they are indeed pictorial with regards to lexigrams they are expected to expect to see lexigrams á priori to actually be able to identify them, i.e. the opposite of pseudo-pictures. That, especially Panbanisha, are said to be able to spontaneously draw lexigrams supports this interpretation. Perhaps they then interpret some pictures as instances of low-iconic lexigrams. Even though lexigrams are constructed as visually arbitrary symbols, perhaps iconicity *do* enters the picture, between static visual displays as such. In addition to a connection between pictures and words, by way of the expectations posed on visual displays on e.g. touchscreens, a connection between pictures and lexigrams on a visual dimension might emerge. This suggests further studies.

This study also shows the potential of combining pictorality with touchscreens. Showing pictures on a touchscreen is of course an age-old practice, but if one would like to develop the work with pictorially competent animals, the touchscreen is a given interface.

Even with less pictorial subjects the touchscreen might be a way into picturehood, given what we learned about "connectedness" in the previous chapter. On a touchscreen the relations between stimuli is factorial in more than one sense. In virtue of being part of a single surface, stimuli are bridged in space. That is different from two laminated pictures on a floor.

Since the surface responds instantaneously to touch, stimuli are also bridged to the manual actions of the subject. That said, it is perceived connection in terms of cause and effect, rather than physical connection that might be the key component. Using a joystick can be as efficient as directly indicating the screen, given that the subject's attention is on the screen and not on the hand steering the joystick (Rumbaugh et al., 1989). In touchscreen setups incentives (e.g. food) can also be in instantaneous proximity to a correct indication on the screen, as can e.g. sound feedback and reinforcers.

In addition, a touchscreen display allows the subject to indicate any part of the picture and therefore preclude dissociation between the indication and the actual piece of information that is supposed to be attended. For macaques in a simple discrimination task, learning tends to be slowed down if the response locale is separated from the relevant cue with as little as half a centimetre (Iwai et al., 1986). Discrimination performance then tends to follow degree of physical separation between cue and location of response in a linear fashion, with larger gaps resulting in poorer results (McClearn & Harlow, 1954). On a touchscreen, on the other hand, manual investigation can go wherever attention goes, and can furthermore be recorded by the researcher.

Touchscreens can also help relieving problems of prominence. Most pictures are physical objects, and this fact can obscure the message. Displaying pictures on a touchscreen which is in itself an object, but one that the subject might already be quite familiar with, has the potential to greatly decrease the "objectness" of pictures and thus the effects of prominence. The screen surface, and not the images themselves, is in this case the primary physical obstacle.

The downside of touchscreens, though, is that a glass surface between the subject and the images also hinders exploration of the differences between real objects and their pictorial versions. Flatness cues are for example compromised, which could otherwise help with differentiation. Showing pictures behind a glass screen increases the likelihood of mistaking the screen as a kind of window rather than a depicting surface. It can therefore, in theory, facilitate a (mis)conception of pictures in reality mode rather than facilitate a pictorial mode.

With respect to the beneficial effect of direct contact between ones manual actions and the relevant stimuli in touchscreen work, it is worth noting that Kanzi and Panbanisha always targeted the motif in the pictures, although the whole picture area was sensitive to touch. Such direct interaction was often precluded in the Givskud studies by way of working through bars. Only the edges of the pictures were typically reachable.

Panbanisha and Kanzi, in these pilot studies of their comprehension of drawn material, demonstrate an impressive ability with novel clip-art. I strongly believe they answer the question whether apes can recognise novel line-drawings in the affirmative. But novel questions instantly spring up. What type of content can they interpret? What transformations can they understand? Can they match pictures equally well without linguistic input? Can pictures help us understand them better? Can we ask them questions through pictures that we cannot do through language? Or rather, through their language, can they tell us things about pictures that reflect their inner worlds in new ways? Especially interesting would the study of dynamic content be. The flow of events is halted in a static picture, but the interpreting mind must not be stalled by this. Dynamic content in pictures requires mental leaps where one must be able to fill in what has just happened, and what is about to happen, to be able to infer what is now happening.

Chapter 14

Summary and conclusions

This thesis has focused on pictorial competences in nonhuman primates. Understanding how animals view pictures is important for the conclusions we draw from experiments that utilise picture materials. In addition, primates' interpretations of pictures, when these are viewed as such, are potential windows to animal minds, just as good as any language. In humans, interpreting pictures can be argued to be both an act of recognition, and one of imagination.

The perspective in this thesis has been broad out of necessity. Before meaningful empirical work can be done the land needs to be charted, and the iconic landscape happens to be very wide. It spans from similarity judgements in everyday categorisation, to recognising oneself in mirrors, to playing with dolls, as well as pictures in many, many forms. The bulk of this thesis has therefore been review, with the purpose of asking new questions about old data.[156] This decision was made when it became clear that the Givskud gorillas would not mind doing a thousand trials more if I insisted on giving them grapes for every second time they pointed at plastic cards or wooden boxes.

To get to grips with the studies I found, I developed a framework that seemed to have been independently thought of several times already. At least three modes of pictorial competence have been suggested.

1. *Surface mode* bypasses recognition of motifs in a picture and focuses on local properties or combinations of local properties. Connection to the real world, if any, is learned. An example of performing in a surface mode would be to respond to the colour in apple pictures without seeing the apples.

2. *Reality mode* allows recognition, but the individual views pictorial displays (the typical example being photographs and video) as just other instances of reality. The result is an involvement with the display that would not have occurred if the view had been differentiated from reality, and especially not if it had been seen as being about reality.

---

[156] In order to find comparable experimental setups to the ones I have used in my own research I had to turn to sometimes very old sources indeed. It has been a good read, but I am especially grateful for looking up the studies of Murray Jarvik, pointed out to me, twice, by a colleague (see Preface). I predict that Jarvik's findings soon will have a revival.

3. *Pictorial mode* implies exerting expectations on pictures that are not exerted on everyday entities. Expectation of visual similarity to something, despite expectation of separation between the picture and the real world, is a tentative way of describing the approach needed towards a picture as a picture. Such expectations offer attentional resources that allow recognition in new ways.

The notion of "modes" was chosen instead of "levels" or "stages," because the three modes do not form a developmental trajectory, nor do they build on each other, although they imply processes that are nested. Appreciating marks on a surface is for example a requirement for all three modes, while appreciating recognisable forms is only a requirement for the last two, and the last mode exclusively requires an expectation of depiction that implies reference and differentiation.

A criterion for performance in a pictorial mode is thus that *a subject should be able to categorise novel depictions that cannot be confused with real instances of the depicted entities.* Four aspects in this criterion require further discussion: (1) categorisation in nonhumans, (2) what a picture is, (3) how reality mode can be precluded, and (4) the novelty requirement and other empirical considerations. The rest of this summary and discussion will be structured according to these areas, ending with suggestions for future directions in animal picture research.

14.1 Categorisation

Categories that are both stable and flexible are needed to recognise objects in nonrealistic pictures. Stable so as to have something to compare *to*, and flexible so that atypical displays are not excluded from this process. It is especially important for learning about how specific objects tend to appear in pictures. The ability to abstract specific features for comparison is arguably a central part of recognition, as well as the ability to make comparisons. Token use seems to facilitate comparison by making relations explicit, but is apparently not a necessity. Comparison involved in pictorality is directional, where the simpler has to be related to the more saturated. Having a picture concept greatly helps in directing comparison from picture to world.

The requirements of stable but flexible categories, and ability to abstract properties and make comparisons, seem to be fulfilled by apes (see Chapter 11). But it remains to specifically target potential limitations in these abilities with respect to similarity judgements. Can e.g. a red circle stand for an apple, where a red triangle and a blue circle cannot? What specific context is needed to evoke this meaning in a red circle? (In what context can red triangles and blue circles stand for apples?) Context is provided in the visual display of the picture, where this appears in the world, and how it is used, e.g. in social interaction. For it to remain an icon, though, context must never depend on conventionality, or habit, at the expense of iconicity.

Pictures, viewed in a pictorial mode, thus allow the study of both perception and categorisation in nonhumans in different ways than pictures viewed as just another instance of an entity, i.e. in reality mode. Also, because content in pictures often needs to be *filled in*, pictures open up the possibility to study animal imagination. Comparing language and non-language competent individuals can answer questions regarding language dependence in this regard.

Pretence was discussed in section 9.1 with the intent to find imaginative use of iconic objects in apes. It was found that such actions tended to be quite uniform and habitual (i.e. nursing dolls). But I would suggest that not until we know more about the understanding of the objects involved, can we separate non-pretence from pretence in play behaviours with iconic objects. It seems clear, though, that at least enculturated individuals make iconic judgements between objects (i.e. different types of dolls).

Likewise when it comes to the extensive research on the use of mirrors (see section 9.4) can conclusions regarding e.g. failures of self-recognition, not be drawn before one knows that the subjects indeed see mirrors as reflecting surfaces.

14.2 Picture perception and concepts

Using a semiotic framework (based on that of Sonesson, e.g. 1989), pictures have in this thesis been used in their function as icons (see Chapter 4). That is, a sign based on the principle of iconicity. The sign being further defined as an expression that is directly connected to the senses, in virtue of being physical, while the focus of the perceiving system is on the content. Expression and content must not be confused with each other, in which case recognition of a motif does not occur. But there is another important differentiation, which is that between the joint expression and content, and the referent that appears in the expression as content. Without this second differentiation, content and referent are one and the same thing, and the entities perceived in the picture cannot refer to external entities, at least not in the form of iconic reference.

Most pictures can be divided into two groups: pictures dependent on primary or secondary iconicity. These are interpretative processes that pertain to the perceiving subject. For primary iconic pictures, similarity precedes pictorality and is in fact the main reason to suspect that something is indeed a picture. All pictures that are possible to process in reality mode are primary iconic. Secondary iconicity is the result of approaching e.g. a surface as if it was a picture, and thus infer a similarity in its features which was not evident before this attention to the display. Even though it has been argued that the secondary iconic sign is not a picture, pictures rich in potentially secondary iconicity (it is not a black-and-white distinction) constitutes the most promising material for studying true pictorial competence. In fact, the most promising empirical work in this thesis has focused on pictures that border on secondary iconicity, arguably more so for naïve viewers than for experienced human viewers.

The picture must be attended to in an appropriate way. The way of attending determines both whether one can see a motif in a picture, and specifies the possibilities of this motif. Expectations on the picture object and its motif is central for how attention is applied.

In the cross-cultural research (Chapter 3), we found powerful examples of at least two related perceptual phenomena that are central for recognising non-naturalistic depictions. I tentatively argued that the phenomenon of "resemanticisation," i.e. that parts and wholes redefine each other in a picture, is part of a larger perceptual phenomenon that is called "successive approximation," i.e. the brain continuously updates its theory about what one is viewing until a percept has settled. Recognising one aspect feeds back to recognition of the whole, which allows further recognition of aspects, etc.

I suggested that what allows recognition of the type of picture that is not immediately recognisable in reality mode, are the expectations on a picture that follows from knowing that one is viewing a *depiction*. These expectations create "attentional purpose" and start the process of successive approximation/resemanticisation. Approaching a picture with such expectations can be said to constitute the pictorial mode. However, it must be noted that *depiction* entails more than expecting a picture to be a potentially informative surface. One must also expect the need to relate this surface to things external to it.

14.3 How real is reality mode, really?

The research on human children shows that young children recognise motifs in photographs as objects, which they try to investigate as if they were three-dimensional. Many examples have been given of reality-derived responses also in animals.

In this thesis it has been proposed that reality modes can be quite allowing. Several animals that see pictures for their very first time can still show recognition. Reductions in size, atypical colours, one thing being in two places at once, etc., are apparently not enough to preclude reality mode processing. As long as spontaneous recognition occurs, one cannot assume that subjects question what they see. But sometimes recognition do breaks down,[157] and it is possible that questioning is exactly what has just happened. Or rather, it was never "real enough" for reality mode to kick in.

When recognition remains despite the picture viewer being fully able to keep the two worlds of pictorality and reality apart, *differentiation* has taken place. Differentiation can be many things but is always partly dependent on cues from the medium. When differentiation fails, the animal can show responses towards pictures etc. in a way more appropriate to real world objects. Many examples of physical actions towards objects in pictures, or social responses towards individuals in pictures, have been given. However, confusion behaviours cannot define reality mode. Partly because they can be difficult to observe, but also because critical features can be lacking that usually triggers action. Even if animals habituate to pictures and stop acting out confusion behaviours while retaining recognition, it does not mean that they have changed modes of picture processing. It is likely that they habituate to individual pictures rather than to pictures as a class. This might be the reason the bonobo Matata still bites food pictures after all these years.

---

[157] Probably more often than the researcher thinks.

In human children (3 – 4 years) one can find a second type of reality responses towards pictures, which does not stem from perceptual or conceptual confusion, but from misconception. This can occur although a picture is used referentially. "Realist errors" (e.g. Thomas et al., 1999) occur when children believe that a representation (i.e. a photograph) updates as a referent updates.[158] However, realist errors seemingly happen independently of iconicity, and pertain to the *relation* between entities (Thomas et al., 1999). It is thus just not believing that a picture of an object and real objects share more properties than they in fact do, which is another common error, but more attributable to a reality mode. Realist errors probably stems from the fact that children get very little experience with pictures referring to specific physical objects. Referents are most of the time only mental. It remains to be investigated whether also animals would believe that changes in referents change their static representations. Interestingly, human children with autism are less prone to realist errors in picture tasks than are typical children (Leekam & Perner, 1991).

I have rather evoked the notion of "pseudo-pictures" (section 7.6) to account for the fact that some ape individuals use pictures referentially, despite probably processing them in a reality mode. Differently from the children above they are thus very much dependent on iconicity (or learning) in order to use pictures in e.g. communication. In a pseudo-picture there is no proper *stand-for* relation since the referential connection between picture and referent contains no "aboutness". The direction moves from what is appreciated in the picture to what can be done with this item, e.g. name it, exchange it, play with it, etc.

The other side of the coin, the first being differentiation, is thus (proper) *reference*. Pictorial reference is the establishment of an asymmetric comparison (from picture to referent), together with the specification of referent, inferred from context, which is external and/or internal to the pictorial display.

Like a word, a picture of a banana can refer to a second banana, or all other bananas. But differently from a word, a picture does this while it depicts a specific object, i.e. the one "in the picture." Attention to the banana in the picture must be transgressed in order for pictorial reference to occur, even when referring to the actual banana in the picture. It implies knowing (differentiation again) that the banana is not *really* there. Pictures affect reference mainly through which properties have been abstracted and selected for presentation. Again, the ability to attend to abstracted properties makes it possible for a red circle to be an apple where a blue triangle cannot.

Do children learn pictorial reference through social interaction? Learning from competent individuals is probably central for learning many things about the picture, for example about the picture as a manufactured object with an inherent intent. However, the common claim that pictures are always about communication can only be true for some pictures and cannot be a defining feature of neither pictures nor pictorial competence. But social interaction is a good way of both learning about the picture as a cultural artefact, and of scaffolding interpretation of pictures by supplying context. Language is an especially effective way of bridging entities in

---

[158] Here we can really talk about an equivalence relation. But as I have stressed, as have Cabe (1980), equivalence without full differentiation is not pictorality.

the real or imagined world with entities as they appear in pictures. But it was shown in Chapter 2 that language also interferes with attention to visual similarity by subjecting reliance on iconicity with reliance on verbal labels.

To conclude, a pictorial concept is partly expectations on a picture, and partly expectations on the contexts that usually contain pictures. Expectation is needed to decode many pictures, but not all. Some hits us directly and are accessible in reality mode. These can also be used in referential tasks, for example as iconic cues in an object-search task, but it is questionable if subjects need to apply *pictorial* reference in such cases. Objects can refer to other objects just as well as pictures can, although the effects of “prominence” make this less likely. That which is less prominent stands for that which is more prominent much better than the other way around. This has been evident in tests using miniature replicas versus tests that use photographs. It seems that the underlying process for this phenomenon is that attention is captured by salience. Also cross-cultural research has shown that the medium, if it is striking on its own accord, can obscure the message. With instruction, however, attention seems to quite easily be able to shift to the pictorial displays. Pseudo-pictures, which allow some degrees of differentiation between pictures and reality, would be more successfully used as cues in e.g. an object-choice task.

14.4 Empirical considerations

Many animal experiments claim pictorial recognition. The reviews in Chapters 5 to 10 suggest that recognition of content seems to occur in some of these experiments, while only the appearance of recognition occurs in others. Recognition of fully fledged pictorial reference is even less supported.

Extensive drilling and reinforcement seem to foster localised strategies and responses to invariant features, while spontaneous response paradigms show more signs of recognition of actual motifs. These often take the form of social responses or otherwise hint that processing in reality mode takes place. Several studies lack adequate controls for associative learning or response to invariant features. In both cases processing in surface mode cannot be ruled out. This has implications for conclusions drawn about animals’ perception and categorisation in the real world.

Especially interesting are the cases where photographs are used in seemingly referential tasks. The notion of pseudo-pictures needs more investigation. Perhaps they will turn out to be real pictures after all.

The original empirical work presented in this thesis took the forms of a match-to-sample (MTS) training study, discrimination tasks, and an object-choice task with gorillas at Givskud Zoo, Denmark, and two forms of sequential MTS (DMTS) tasks at Great Ape Trust of Iowa, USA.

14.4.1 Givskud

The intention at Givskud Zoo was to develop and train a method for testing pictorial competence in experimentally and picture-naïve gorillas. A simultaneous matching-to-sample (MTS) paradigm, as well as an object-choice task (sequential MTS with a problem-solving component) was implemented. Following training with photographs the ambition was to move on to drawings. However, a matching competence was never obtained.

In training of the MTS task the subjects were to match pictures to identical pictures in terms of colour, or colour plus shape, or to match photographs to edible objects. Performance only raised above chance when the correct choice item was indexically cued by placing the reward on or beneath it. Placing the sample picture closer to the matching pictures fostered the same effect in some subjects.

With the object-choice task the aim was to induce the gorillas to use pictorial information to guide their selection of one of two containers. Following extensive, unsuccessful, training it was judged that the subjects did not attend to the pictures pasted on top of the containers. Alternatively they did not utilise iconic information in their choices. However, they did use indexical information, i.e. picture placed on top of only the baited contained. It was reasoned that the experimental setup was mainly to blame. Disregarding the need for a means – end analysis, based on causality and physical connectedness that linked to iconicity, the object choice-task failed to elicit good performance.

14.4.2 Great Ape Trust of Iowa

With the language-competent bonobos Panbanisha and Kanzi at GATI a computerised DMTS was used in two versions: a “receptive” and a “productive” one.[159] The use of productive delayed experiments is preferred. In a receptive picture test the subject needs to pinpoint the correct picture among several, but in a productive version the picture is the question and is viewed independently of the answers. In productive testing there is therefore a greater need for spontaneous interpretation of the picture, dependent on an expectation of depiction. In receptive testing there is instead a risk that scaffolding context is given by relations to the non-matching pictures, the worst case being choice based on exclusion. Consequently, the target picture only gets a relative iconic meaning, e.g. the most banana-like thing in an array. Such estimations can occur even in surface mode.

Kanzi and Panbanisha displayed very good recognition of novel line-drawings on a category level that mapped to their lexigram use. A continued study to see whether pictures, viewed in pictorial mode, can map also to their non-lexigram world would be most interesting.

---

[159] The terms “receptive” and ”productive” are borrowed from language research where a difference is made between e.g. receptive and productive vocabulary. In this thesis I have used ”receptive” when the investigated picture is on the choice-item side in MTS, and ”productive” when the picture is on the sample side in MTS, because that is how they appear in language testing.

A note on the generalisability of data from language-competent apes is in order. My short answer is that we do not need to generalise to other apes, nor to common ancestors. We now know that a specifically human brain is not necessary to recognise drawing stimuli. Our real task is to understand the processes that made these specific apes into the picture viewers that they are. What sustains this aspect in them?

I have argued in this thesis that the study of ape potentiality is a worthwhile endeavour. Language-competent apes are inspiring co-workers in this regard. But remember for example that only one subjects among 27 in Herrmann et al. (2005) seemed to rely on iconicity in making her choice across all the indexical choice-tasks. This does not mean that 26 subjects did not have the potential, only that they settled for different solutions within the limited experience they had with the task. Sixteen subjects did not show any sensitivity to iconic information at all. What if these sixteen had been the only ones tested? Should it be concluded that iconic information is wasted on apes? From negative data we cannot even conclude that iconic information is difficult for apes to access. A useful question to ask in such a situation is: *What are the apes really attempting in order to solve the task, and what should be changed in order to elicit the use of iconic information instead?* This was how Kuhlmeier and Boysen (2001; 2002) created model-using chimpanzees. One could argue that such manipulation would be to shape an ability that did not exist before, but the very same interjection highlights the fact that it might not always be accurate to talk about abilities as fixed properties of the organism. Cognitive studies of ape minds are studies of the potential of ape minds. The relevant question is what it was that made that single gorilla female in Herrmann et al. (2005) to settle for an iconic solution to her problem.

14.5 Future directions

I second the call already made by Cabe (1980) that it remains to be investigated "[w]hat objects, features of objects, events, and/or aspects of environmental layout depicted in pictures can be recognized by animals" (p. 325) and that much, if not most, still remains to be done in the field, almost 30 years later.

An obvious area for investigation is thus to continue to test what types of content apes are able to appreciate in pictures, both when working in a reality mode and in a pictorial mode. The limitations are of course the questions that are possible to pose within a given experimental setup.

MTS seems to be the most open-ended method to study pictorality, where great control can be exerted over the stimuli and their relations to each other. It seems an especially fruitful method if it takes place on a computer with a touchscreen or joystick interface. A screen allows direct contact in time and space between the actions of the animal, the stimuli, and feedback. As mentioned above, a productive setup is preferred. In order to be able to test also non-language competent subjects the best setup seems therefore to be a DMTS setup where a drawing serves as sample, and photographs, movie-clips, or objects, serve as choice items. Dynamic content can be investigated by matching movies of events to pictures of events. Again the pictures need to be viewed independently and prior to the movies and preferably be non-realistic.[160]

If a reliable MTS competence is lacking in the subjects, I would suggest a sorting-by-sample paradigm. Sorting objects seems to have been employed successfully in several studies, with less training than MTS requires. A problem with sorting tasks, as with MTS and discrimination tasks, is to make sure that the subject attends to the accurate categorical level. Matching on the grounds of surface properties must be precluded. In this regard testing language-competent subjects, who have shown accurate symbol use in independent testing, is a great advantage since they can match to words in controls. With other subjects control trials with choice items that match on different categorical levels can be given.

If an object-choice task cannot be avoided, it should take into consideration the extra attentional demands posed by the equipment. Relations between sample, container, reward, and the actions of the subject, must allow for a causal analysis. Likewise the role of the experimenter needs to be transparent. This is perhaps best done by making the task altogether automatic.

The access to apes in enculturation contexts, who have an understanding of the testing procedures, and even the specific questions addressed to them, has opened up new possibilities for the study of pictures. But it remains to be investigated if, and how, their experience with humans has fed back to their pictorial competence. Such investigations necessarily start with comparing language-naïve subjects with language-competent ones. Can for example Matata perform comparably to Panbanisha and Kanzi? If language-naïve subjects can recognise drawings, it will be possible to directly pose them experimental problems on a categorical rather than instance based level. A drawing of e.g. a bird cannot be taken as an instance of a real, or almost real, individual bird. It must be recognised on a categorical level, very much like the word “bird.”

Furthermore, can language-competent subjects perform comparably in matching tasks that involve language and in purely pictorial tasks, where both sample and choice-items are pictures? It is possible, for example, that through language individual categories tend to converge with those of other individuals. We create shared concepts for that which we talk with each other about. Perhaps language-competent apes excel at categorising entities that they have lexigrams for in a photograph (match) to drawing (sample) condition. Comparing interpretation of drawings in a pictorial mode drawn from many types of categories can help us study this possibility.

It has been argued here that a picture is not one and the same thing in the three modes. From a cognitive perspective animals interaction with these three objects are highly interesting and worth studying on their own. Perception in surface and reality modes, and categorisation in reality mode, are classic areas. However, if true pictorial competence can be demonstrated in nonhumans, and I think that a few such

---

[160] From a pictorial perspective a possible downside with work on a computer screen is that the picture loses its object status and thus the medium as such does not assist with differentiation.

studies can be found in this review and research,[161] it opens up new avenues for exploring perception, categorisation, and imagination in great apes. I think Garner, Witmer, Köhler and Yerkes, among others, would have been intrigued.

> *As a further variation – to meet possible objections – I would suggest, in the crucial experiments, the use of pictures of another food – say the very popular oranges or thistles – if bananas were used in the preliminary tests.*

---

[161] Notably the works with the gorilla Koko and the GATI bonobos are suggestive in this regard. But the referential use of photographs by the gorilla King, the possible recognition of line-drawings by the chimpanzee Ai, and the likely recognition in the chimpanzee Viki, are also important findings. Add to this the use of scale models by e.g. the chimpanzees Sarah and Sheba.

References

Akshoomoff, N.A. & Stiles, J. (1995). Developmental trends in visuospatial analysis and planning: I. copying a complex figure. *Neuropsychology, 9*(3), 364-377.

Anderson, J.R. (2000). Primates and representations of self. In J. Fagot (Ed.), *Picture Perception in Animals* (pp. 373-396). Hove: Psychology Press Ltd.

Aust, U. & Huber, L. (2003). Elemental vesus configural perception in a people-present/people-absent discrimination task by pigeons. *Learning and Behavior, 31*(3), 213-224.

Baldwin, D.A. (1993). Early referential understanding: infants' ability to recognize referential acts for what they are. *Developmental Psychology, 29*(5), 832-843.

Baldwin, D.A. & Moses, L.J. (2001). Links between social understanding and early word learning: challenges to current accounts. *Social Development, 10*(3), 309-329.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5*, 617-629.

Bard, K.A. & Gardner, K.H. (1996). Influences on development in infant chimpanzees: enculturation, temperament, and cognition. In A.E. Russon, K.A. Bard & S.T. Parker (Eds.), *Reaching Into Thought. The Minds of the Great Apes* (pp. 235-256). Cambridge: Cambridge University Press.

Bard, K.A., Todd, B.K., Bernier, C., Love, J. & Leavens, D.A. (2006). Self-awareness in human and chimpanzee infants: what is measured and what is meant by the mark and mirror test? *Infancy, 9*(2), 191-219.

Barth, J. & Call, J. (2006). Tracking the displacement of objects: a series of tasks with great apes (*Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla*, and *Pongo pygmaeus*) and young children (*Homo sapiens*). *Journal of Experimental Psychology: Aninal Behavior Processes, 32*(3), 239-252.

Bauer, H.R. & Philip, M.M. (1983). Facial and vocal recognition in the common chimpanzee. *Psychological Record, 33*, 161-170.

Beilin, H. & Pearlman, E.G. (1991). Children's iconic realism: object versus property realism. *Advances in Child Development and Behavior, 23*, 73-111.

Benjamin, L.T., Jr. & Bruce, D. (1982). From bottle-fed chimp to bottlenose dolphin: a contemporary appraisal of Winthrop Kellogg. *The Psychological Record, 32*, 461-482.

Beran, M.J., Beran, M.M. & Menzel, C.R. (2005). Chimpanzees (*Pan troglodytes*) use markers to monitor the movement of a hidden item. *Primates, 46*, 255-259.

Beran, M.J., Savage-Rumbaugh, E.S., Pate, J.L. & Rumbaugh, D.M. (1999). Delay of gratification in chimpanzees (*Pan troglodytes*). *Developmental Psychobiology, 34*, 119-127.

Beran, M.J. & Washburn, D.A. (2002). Chimpanzee responding during matching-to-sample: control by exclusion. *Journal of the Experimental Analysis of Behavior, 78*(3), 497-508.

Bering, J.M. (2004). A critical review of the "enculturation hypothesis": the effects of human rearing on great ape social cognition. *Animal Cognition, 7*, 201-212.

Bhatt, R.S., Wasserman, E.A., Reynolds, W.F., Jr. & Knauss, K.S. (1988). Conceptual behavior in pigeons: categorization of both familiar and novel examples from four classes of natural and artificial stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 14*(3), 219-234.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94*(2), 115-147.

Bierens de Haan, J.A. (1927). Versuche über das Sehen der Affen. IV. Das Erkennen gleichförmiger und ungleichförmiger Gegenstände bei niederen Affen. V. Erkennen Affen in zweidimensionalen Abbildungen ihnen bekannte Gegenstände wieder? *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 5*(4), 699-729.

Bingham, H.C. (1932). *Gorillas in a Native Habitat.* Washington: Carnegie Institution of Washington.

Bloomsmith, M.A. & Lambeth, S.P. (2000). Videotapes as enrichment for captive chimpanzees (*Pan troglodytes*). *Zoo Biology, 19*, 541-551.

Bovet, D. & Vauclair, J. (1998). Functional categorization of objects and of their pictures in baboons (*Papio anubis*). *Learning and Motivation, 29*, 309-322.

Bovet, D. & Vauclair, J. (2000). Picture recognition in animals and humans. *Behavioural Brain Research, 109*(2), 143-165.

Bovet, D. & Vauclair, J. (2001). Judgement of conceptual identity in monkeys. *Psychonomic Bulletin and Review, 8*(3), 470-475.

Bower, T.G.R. (1972). Object perception in infants. *Perception, 1*, 15-30.

Boysen, S.T. (1994). Individual differences in the cognitive abilities of chimpanzees. In R.W. Wrangham, W.C. McGrew, F.B.M. de Waal & P.G. Heltne (Eds.), *Chimpanzee Cultures* (pp. 335-350). Cambridge, USA: Harvard University Press.

Boysen, S.T. & Berntson, G.G. (1986). Cardiac correlates of individual recognition in the chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology, 100*(3), 321-324.

Boysen, S.T. & Berntson, G.G. (1989). Conspecific recognition in the chimpanzee (*Pan troglodytes*): cardiac responses to significant others. *Journal of Comparative Psychology, 103*(3), 215-220.

Boysen, S.T., Berntson, G.G. & Prentice, J. (1987). Simian scribbles: a reappraisal of drawing in the chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology, 101*(1), 82-89.

Boysen, S.T., Bryan, K.M. & Shreyer, T.A. (1994). Shadows and mirrors: alternative avenues to the development of self-recognition in chimpanzees. In S.T. Parker, R.W. Mitchell & M.L. Boccia (Eds.), *Self-Awareness in Animals and Humans. Developmental Perspectives* (pp. 227-240). New York: Cambridge University Press.

Boysen, S.T. & Kuhlmeier, V.A. (2002). Representational capacities for pretence with scale models and photographs in chimpanzees (*Pan troglodytes*). In R.W. Mitchell (Ed.), *Pretending and imagination in animals and children* (pp. 210-228). Cambridge: Cambridge University Press.

Brakke, K.E. & Savage-Rumbaugh, E.S. (1995). The development of language skills in bonobo and chimpanzee – I. Comprehension. *Language and Communication, 15*(2), 121-148.

Brakke, K.E. & Savage-Rumbaugh, E.S. (1996). The development of language skills in *Pan* – II. Produciton. *Language and Communication, 16*(4), 361-380.

Bräuer, J., Kaminski, J., Riedel, J., Call, J. & Tomasello, M. (2006). Making inferences about the location of hidden food: social dog, causal ape. *Journal of Comparative Psychology, 120*(1), 38-47.

Brennan, B. & Visty, R. (Producers). (1999). *A Conversation with Koko* [Video]. Visty/Brennan Productions and Thirteen/WNET New York.

Brewster, J.M. & Siegel, R.K. (1976). Reinforced drawing in *Macaca mulatta*. *Journal of Human Evolution, 5*, 345-347.

Brown, D.A. & Boysen, S.T. (2000). Spontaneous discrimination of natural stimuli by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 114*(4), 392-400.

Bovet, D. & Vauclair, J. (2000). Picture recognition in animals and humans. *Behavioural Brain Research, 109*, 143-165.

Brooks-Gunn, J. & Lewis, M. (1984). The development of early visual self-recognition. *Developmental Review, 4*, 215-239.

Bruce, C. (1982). Face recognition by monkeys: absence of an inversion effect. *Neuropsychologia, 20*(5), 515-521.

Buttelmann, D., Call, J. & Tomasello, M. (2008). Behavioral cues that great apes use to forage for hidden food. *Animal Cognition*, *11*(1), 117-128.

Byrne, R. (2004). The manual skills and cognition that lie behind hominid tool use. In A.E. Russon & D.R. Begun (Eds.), *The Evolution of Thought. Evolutionary Origins of Great Ape Intelligence* (pp. 31-44). Cambridge: Cambridge University Press.

Byrnit, J. (2005). *Primate Theory of Mind: A Comparative Psychological Analysis*. Århus: Psykologisk Institut, Aarhus Universitet.

Cabe, P.A. (1980). Picture perception in nonhuman subjects. In M.A. Hagen (Ed.), *The Perception of Pictures, Vol. 2* (pp. 305-343). New York: Academic Press, Inc.

Cacchione, T. & Krist, H. (2004). Recognizing impossible object relations: intuitions about support in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 118*(2), 140-148.

Call, J. (2001). Object permanence in orangutans (*Pongo pygmaeus*), chimpanzee (*Pan troglodytes*), and children (*Homo sapiens*). *Journal of Comparative Psychology, 115*(2), 159-171.

Call, J. (2003). Spatial rotations and transpositions in orangutans (*Pongo pygmaeus*) and chimpanzees (*Pan troglodytes*). *Primates, 44*, 347-357.

Call, J. (2006a). Inference by exclusion in the great apes: the effect of age and species. *Animal Cognition, 9*, 393-403.

Call, J. (2006b). Descarte's two errors: reason and reflection in the great apes. In S. Hurley & M. Nudds (Eds.), *Rational Animals?* (pp. 219-234). New York: Oxford University Press.

Call, J. & Tomasello, M. (1996). The effect of humans on the cognitive development of apes. In A.E. Russon, K.A. Bard & S.T. Parker (Eds.), *Reaching Into Thought. The Minds of the Great Apes* (pp. 371-403). Cambridge: Cambridge University Press.

Callaghan, T.C. (1999). Early understanding and production of graphic symbols. *Child Development, 70*(6), 1314-1324.

Callaghan, T.C. (2000). Factors affecting children's graphic symbol use in the third year. Language, similarity, and iconicity. *Cognitive Development, 15*, 185-214.

Callaghan, T.C. & Rankin, M.P. (2002). Emergence of graphic symbol functioning and the question of domain specificity: a longitudinal training study. *Child Development, 73*(2), 359-376.

Callaghan, T.C., Rochat, P., MacGillivray, T, & MacLellan, C. (2004). Modeling referential actions in 6- to 18-month-old infants: a precursor to symbolic understanding. *Child Development, 75*(6), 1733-1744.

Candland, D.K. (1993). *Feral Children and Clever Animals. Reflections on Human Nature*. New York: Oxford University Press.

Capitanio, J.P., Boccia, M.L. & Colaiannia, D.J. (1985). The influence of rank on affect perception by pigtailed macaques (*Macaca nemestrina*). *American Journal of Primatology, 8*, 53-59.

Caron-Pargue, J. (1989). Is pictorial space "perceived" as real space? *Behavioral and Brain Sciences, 12*(1). 75-76.

Cerella, J. (1977). Absence of perspective processing in the pigeon. *Pattern Recognition, 9*, 65-68.

Cerella, J. (1980). The pigeon's analysis of pictures. *Pattern Recognition, 12*, 1-6.

Chance, M.R.A. (1960). Kohler's chimpanzees - how did they perform? *Man*, no. 179, 130-135.

Charman, T. & Baron-Cohen, S. (1995). Understanding photos, models, and beliefs: a test of the modularity thesis of theory of mind. *Cognitive Development, 10*, 287-298.

Coetzee, J.M. (1997). The lives of animals. *The Tanner Lectures on Human Values*, delivered at Princeton University, Oct. 15-16, 1997 [PDF document]. URL http://www.tannerlectures.utah.edu/lectures/coetzee99.pdf

Cole, P.D. & Honig, W.K. (1994). Transfer of a discrimination by pigeons (*Columba livia*) between pictured locations and the represented environments. *Journal of Comparative Psychology, 108*(2), 189-198.

Cook, R.G., Katz, J.S. & Kelly, D.M. (2000). Pictorial same-different categorical learning and discrimination in pigeons. In J. Fagot (Ed.), *Picture Perception in Animals* (pp. 181-218). Hove: Psychology Press Ltd.

Corbetta, M. & Shulman, G.L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*, 201-215.

Cowles, J.T. & Nissen, H.W. (1937). Reward-expectancy in delayed responses of chimpanzees. *Journal of Comparative Psychology, 24*, 345-358.

Custance, D. & Bard, K.A. (1994). The comparative and developmental study of self-recognition and imitation: the importance of social factors. In S.T. Parker, R.W. Mitchell & M.L. Boccia (Eds.), *Self-awareness in Animals and Humans. Developmental Perspectives* (pp. 207-226). New York: Cambridge University Press.

Custance, D.M., Whiten, A. & Bard, K.A. (1995). Can young chimpanzees (*Pan troglodytes*) imitate arbitrary actions? Hayes & Hayes (1952) revisited. *Behaviour, 132*(11-12), 837-859.

Daehler, M.W., Lonardo, R. & Bukato, D. (1979). Matching and equivalence judgements in very young children. *Child Development, 50*, 170-179.

D'Amato, M.R. & Van Sant, P. (1988). The person concept in monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes, 14*(1), 43-55.

Dasser, V. (1987). Slides of group members as representations of the real animals (*Macaca fascicularis*). *Ethology, 76*, 65-73.

Dasser, V. (1988). A social concept in Java monkeys. *Animal Behavior, 36*, 225-230.

Davenport, R.K. & Rogers, C.M. (1970). Intermodal equivalence of stimuli in apes. *Science, 168*, 279-280.

Davenport, R.K. & Rogers, C.M. (1971). Perception of photographs by apes. *Behaviour, 39*(2), 318-320.

Davenport, R.K., Rogers, C.M. & Russell, I.S. (1973). Cross modal perception in apes. *Neuropsychologia, 11*, 21-28.

Davenport, R.K., Rogers, C.M. & Russell, I.S. (1975). Cross-modal perception in apes: altered visual cues and delay. *Neuropsychologia, 13*, 229-235.

Dawkins, M.S., Guilford, T., Braithwaite, V.A. & Krebs, J.R. (1996). Discrimination and recognition of photographs of places by homing pigeons. *Behavioural Processes, 36*, 27-38.

Day, C. (1920/1941). *This Simian World*. New York: Alfred A. Knoff.

D'eath, R.B. (1998). Can video images imitate real stimuli in animal behaviour experiments? *Biological Review, 73*, 267-292.

de Veer, M.W., Gallup, G.G., Jr., Theall, L.A., van den Bos, R. & Povinelli, D. (2003). An 8-year longitudinal study of mirror self-recognition in chimpanzees (*Pan troglodytes*). *Neuropsychologia, 41*, 229-234.

de Waal, F.B.M. (1998). No imiation without identification. *Behavioral and Brain Sciences, 21*(5), 689.

de Waal, F.B.M. (2001). *The Ape and the Sushi Master: Cultural Reflections by a Primatologist*. New York: Basic Books.

de Waal, F.B.M., Dindo, M., Freeman, C.A. & Hall, M.J. (2005). The monkey in the mirror: hardly a stranger. *Proceedings of the National Academy of Sciences of the United States of America, 102*(32), 11140-11147.

Delius, J.D., Emmerton, J., Hörster, W., Jäger, R. & Ostheim, J. (2000). Picture-object recognition in pigeons. In J. Fagot (Ed.), *Picture Perception in Animals* (pp. 1-35). Hove: Psychology Press Ltd.

DeLoache, J.S. (1986). Memory in very young children: exploitation of cues to the location of a hidden object. *Cognitive Development, 1*(2), 123-137.

DeLoache, J.S. (1987). Rapid change in the symbolic functioning of very young children. *Science, 238*, 1556-1557.

DeLoache, J.S. (2000). Dual representation and young children's use of scale models. *Child Development, 71*(2), 329-338.

DeLoache, J.S. (2004). Becoming symbol minded. *Trends in Cognitive Sciences, 8*(2), 66-70.

DeLoache, J.S. & Brown, A.L. (1983). Very young children's memory for the location of objects in a large-scale environment. *Child Development, 54*, 888-897.

DeLoache, J.S. & Burns, N.M. (1994). Early understanding of the representational function of pictures. *Cognition, 52*, 83-110.

DeLoache, J.S. & Sharon, T. (2005). Symbols and similarity: you can get too much of a good thing. *Journal of Cognition and Development, 6*(1), 33-49.

DeLoache, J.S., Peralta de Mendoza, O.A. & Anderson, K.N. (1999). Multiple factors in early symbol use: instructions, similarity, and age in understanding a symbol-referent relation. *Cognitive Development, 14*, 299-312.

DeLoache, J.S., Pierroutsakos, S.L., Uttal, D.H., Rosengren, K.S. & Gottlieb, A. (1998a). Grasping the nature of pictures. *Psychological Science, 9*(3), 205-210.

DeLoache, J.S., Simcock, G. & Marzolf, D.P. (2004a). Transfer by very young children in the symbolic retrieval task. *Child Development, 75*(6), 1708-1718.

DeLoache, J.S., Uttal, D.H. & Pierroutsakos, S.L. (1998b). The development of early symbolization: educational implications. *Learning and Instruction, 8*(4), 325-339.

DeLoache, J.S., Uttal, D.H. & Pierroutsakos, S.L. (2000). What's up? The development of an orientation preference for picture books. *Journal of Cognition and Development, 1*, 81-95.

DeLoache, J.S., Uttal, D.H. & Rosengren, K.S. (2004b). Scale errors offer evidence for a perception-action dissociation early in life. *Science, 304*, 1027-1029.

Demaria, C. & Thierry, B. (1988). Responses to animal stimulus photographs in stumptailed macaques (*Macaca arctoides*). *Primates, 29*(2), 237-244.

Deregowski, J. (1968). Difficulties in pictorial depth perception in Africa. *British Jorunal of Psychology, 59*, 195-204.

Deregowski, J. (1969). Perception of the two-pronged trident by two- and three-dimensional perceivers. *Journal of Experimental Psychology, 82*, 9-13.

Deregowski, J. (1970). Responses mediating pictorial recognition. *Journal of Social Psychology, 84*, 27-33.

Deregowski, J.B. (1976). On seeing a picture for the first time. *Leonardo, 9*, 19-23.

Deregowski, J.B. (1989). Real space and represented space: cross-cultural perspectives. *Behavioral and Brain Sciences, 12*(1), 51-74.

Deregowski, J.B., Muldrow, E.S. & Muldrow, W.F. (1972). Pictorial recognition in a remote Ethiopean population. *Perception, 1*, 417.

Dewsbury, D.A. (2003). Conflicting approaches: operant psychology arrives at a primate laboratory. *Behavior Analyst, 26*, 253-265.

Diamond, A., Churchland, A., Cruess, L. & Kirkham, N.Z. (1999). Early development in the ability to understand the relation between stimulus and reward. *Developmental Psychology, 35*(6), 1507-1517.

Diamond, R. & Carey, S. (1986). Why faces are and are not special: an effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107-117.

Dittrich, W. (1990). Representation of faces in longtailed macaques (*Macaca fascicularis*). *Ethology, 85*, 265-278.

Dittrich, W. (1994). How monkeys see others: discrimination and recognition of monkeys' shape. *Behavioural Processes, 33*, 139-154.

Dodwell, P.C., Muir, D. & DiFranco, D. (1976). Responses of infants to visually presented objects. *Science, 194*, 209-211.

Dugdale, N. & Lowe, C.F. (2000). Testing for symmetry in the conditional discriminations of language-trained chimpanzees. *Journal of the Experimental Analysis of Behavior, 73*, 5-22.

Eddy, T.J., Gallup, G.G., Jr., Povinelli, D.J. (1996). Age differences in the ability of chimpanzees to distinguish mirror-images of self from video images of others. *Journal of Comparative Psychology, 110*(1), 38-44.

Fagot, J. & Deruell, C. (2002). Perception of pictorial eye gaze by baboons (*Papio papio*). *Journal of Experimental Psychology: Animal Behavior Processes, 28*(3), 298-308.

Fagot, J., Martin-Malivel, J. & Dépy, D. (2000). What is the evidence for an equivalence between objects and pictures in birds and nonhuman primates? In J. Fagot (Ed.), *Picture Perception in Animals* (pp. 295-320). Hove: Psychology Press Ltd.

Fagot, J. & Tomonaga, M. (1999). Global and local processing in humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*): use of a visual search task with compound stimuli. *Journal of Comparative Psychology, 113*(1), 3-12.

Fagot, J. & Tomonaga, M. (2001). Effects of element separation on perceptual grouping by humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*): perception of Kanizsa illusory figures. *Animal Cognition, 4*, 171-177.

Fletcher, K.L. & Reese, E. (2005). Picture book reading with young children: a conceptual framework. *Developmental Review, 25*, 64-103.

Fouts, R. (1997). *Next of Kin. My Conversations with Chimpanzees*. London: Penguin Books.

Fraser, A.K. (1932). *Teaching Healthcraft to African Women.* London: Longmans.

Friedman, A., Spetch, M.L. & Ferrey, A. (2005). Recognition by humans and pigeons of novel views of 3-d objects and their photographs. *Journal of Experimental Psychology: General, 134*(2), 149-162.

Friedman, S.L. & Stevenson, M.B. (1980). Perception of movement in pictures. In M.A. Hagen (Ed.), *The Perception of Pictures, Vol. 1* (pp. 225-255). New York: Academic Press, Inc.

Friends of Washoe (2004). Meet the family – Washoe's biography [WWW page]. URL http://www.friendsofwashoe.org/washoe_bio.shtml

Fuglesang, A. (1970). Communication with illiterates. *Markeds Kemmunikasjon, 3*, 24-66.

Fujita, K. (1987). Species recognition by five macaque monkeys. *Primates, 28*(3), 353-366.

Fujita, K. (1993a). Role of some physical characteristics in species recognition by pigtail monkeys. *Primates, 34*(2), 133-140.

Fujita, K. (1993b). Development of visual preference for closely related species by infant and juvenile macaques with restricted social experience. *Primates, 34*(2), 141-150.

Fujita, K. & Matsuzawa, T. (1986). A new procedure to study the perceptual world of animals with sensory reinforcement: recognition of humans by a chimpanzee. *Primates, 27*(3), 283-291.

Fujita, K. & Matsuzawa, T. (1990). Delayed figure reconstruction by a chimpanzee (*Pan troglodytes*) and humans (*Homo sapiens*). *Journal of Comparative Psychology, 104*(4), 345-351.

Fujita, K., Watanabe, K., Widarto, T.H. & Suryobroto, B. (1997). Discrimination of macaques by macaques: the case of Sulawesi species. *Primates, 38*(3), 233-245.

Furness, W.H. (1916). Observations on the mentality of chimpanzees and orang-utans. *Proceedings of the American Philosophical Society, 55*(3), 281-290.

Gallup, G.G., Jr. (1970). Chimpanzees: self-recognition. *Science, 167*, 86-87.

Gallup, G.G., Jr. (1977). Self-recognition in primates. A comparative approach to the bidirectional properties of consciousness. *American Psychologist, 32*, 329-338.

Garner, R.L. (1892). *The Speech of Monkeys*. London: William Heinemann.

Garner, R.L. (1896). *Gorillas and Chimpanzees*. London: Osgood, McIlvaine & Co.

Gardner, B.T. & Gardner, R.A. (1971) Two-way communication with an infant chimpanzee. In A.M. Schrier & F. Stollnitz (Eds.), *Behavior of Nonhuman Primates, Vol. 4.* (pp. 117-183) New York & London: Academic Press.

Gardner, B.T. & Gardner, R.A. (1985). Signs of intelligence in cross-fostered chimpanzees. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 308*, 159-176.

Gardner, H. & Wolf, D. (1983). Waves and streams of symbolization: notes on the development of symbolic capacities in young children. In D. Rogers & J.A. Sloboda (Eds.), *The Acquisition of Symbolic Skills* (pp. 19-42). London: Plenum Press.

Gardner, R.A. & Gardner, B.T. (1969). Teaching sign language to a chimpanzee. *Science, 165*, 664-672.

Gardner, R.A. & Gardner, B.T. (1984). A vocabulary test for chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 98*(4), 381-404.

Gardner, R.A. & Gardner, B.T. (1989). A cross-fostering laboratory. In R.A. Gardner, B.T. Gardner & T.E. Van Cantfort (Eds.), *Teaching Sign Language to Chimpanzees* (pp. 1-28). Albany: State University of New York.

Gardner, R.A. & Gardner, B.T. (1998). *The Structure of Learning – From Sign Stimuli to Sign Language*. Mahwah: Lawrence Erlbaum Associates, Inc.

Gardner, R.A., Gardner, B.T. & Van Cantfort, T.E. (Eds.) (1989). *Teaching Sign Language to Chimpanzees*. Albany: State University of New York.

Gellermann, L.W. (1933). Form discrimination in chimpanzees and two-year-old children: I. Form (triangularity) per se. *Journal of Genetic Psychology, 42*, 3-27.

Gelman, S.A. & Bloom, P. (2000). Young children are sensitive to how an object was created when deciding what to name it. *Cognition, 76*, 91-103.

Gelman, S.A. & Ebeling, K.S. (1998). Shape and representational status in children's early naming. *Cognition, 66*, B35-B47.

Gelman, S.A., Chesnick, R.J. & Waxman, S.R. (2005). Mother – child conversations about pictures and objects: referring to categories and individuals. *Child Development, 76*(6), 1129-1143.

Gibeault, S. & MacDonald, S.E. (2000). Spatial memory and foraging competition in captive western lowland gorillas (*Gorilla gorilla gorilla*). *Primates, 41*(2), 147-160.

Gibson, B.M. & Wasserman, E.A. (2003). Pigeons learn stimulus identity and stimulus relations when both serve as redundant, relevant cues during same-different discrimination training. *Journal of Experimental Psychology: Animal Behavior Processes, 29*(1), 84-91.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Glover, S. (2004). What causes scale errors in children? *Trends in Cognitive Sciences, 8*(10), 440-442.

Golomb, C. (1992). *The Child's Creation of a Pictorial World*. Berkely and Los Angeles: University of California Press.

Gómez, J.C. & Martín-Andrade, B. (2002). Possible precursors of pretend play in nonpretend actions of captive gorillas (*Gorilla gorilla*). In R.W. Mitchell (Ed.), *Pretending and imagination in animals and children* (pp. 255-268). Cambridge: Cambridge University Press.

Gorilla Foundation, The (2007). KokoPix of the day [WWW page]. URL http://www.koko.org/world/kokopix.php

Gothard, K.M., Erickson, C.A. & Amaral, D.G. (2004). How do rhesus monkeys (*Macaca mulatta*) scan faces in a visual paired comparison task? *Animal Cognition, 7*, 25-36.

Gregory, R.L. (1970). *The Intelligent Eye*. New York: McGraw-Hill.

Gregory, R.L. (1973). The confounded eye. In R.L. Gregory & E.H. Gombrich (Eds.), *Illusion in Nature and Art*. London: Duckworth.

Gross, J. & Hayne, H. (1999). Young children's recognition and description of their own and others' drawings. *Developmental Science 2*(4), 476-489.

Haber, R.N. (1980). Perceiving space from pictures: a theoretical analysis. In M.A. Hagen (Ed.), *The Perception of Pictures, Vol. 1* (pp. 3-31). New York: Academic Press, Inc.

Hamilton, C.R. & Vermeire, B.A. (1983). Discrimination of monkey faces by split-brain monkeys. *Behavioural Brain Research, 9*, 263-275.

Hare, B. & Tomasello, M. (2004). Chimpanzees are more skilful in competetive than in cooperative cognitive tasks. *Animal Behaviour, 68*, 571-581.

Harlow, H.F. (1945a). Studies in discrimination learning by monkeys: III. Factors influencing the facility of solution of discrimination problems by rhesus monkeys. *Journal of General Psychology, 32*, 213-?.

Harlow, H.F. (1945b). Studies in discrimination learning by monkeys: IV. Relative difficulty of discriminations between stimulus-objects and between comparable patterns with homogenous and with heterogenous grounds. *Journal of General Psychology, 32*, 317-?.

Harlow, H.F. (1949). The formation of learning sets. *Psychological Review, 56*, 51-65.

Harlow, H.F. (1951). Primate learning. In C.P. Stone (Ed.), *Comparative Psychology* (3rd ed.) (pp. 183-238). New York: Prentice Hall, Inc.

Harrison, B. (1964). *Orang-utan ("Orang-Utan", 1962)*. Stockholm: Rabén & Sjögren.

Hashiya, K. & Kojima, S. (1997). Auditory-visual intermodal matching by a chimpanzee (Pan troglodytes). *Japanese Psychological Research, 39*(3), 182-190.

Hashiya, K. & Kojima, S. (2001a). Hearing and auditory-visual intermodal recognition in the chimpanzee. In T. Matsuzawa (Ed.), *Primate Origins of Human Cognition and Behavior* (pp. 155-189). Tokyo: Springer.

Hashiya, K. & Kojima, S. (2001b). Acquisition of auditory-visual intermodal matching-to-sample by a chimpanzee (*Pan troglodytes*): comparison with visual-visual intramodal matching. *Animal Cognition, 4*, 231-239.

Haun, D.B.M, Call, J., Janzen, G. & Levinson, S.C. (2006). Evolutionary psychology of spatial representations in the Hominidae. *Current Biology, 16*, 1736-1740.

Hayes, C. (1951). *The Ape in Our House*. New York: Harper & Brothers.

Hayes, K.J. & Hayes, C. (1952). Imitation in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology, 45*, 450-459.

Hayes, K.J. & Hayes, C. (1953). Picture perception in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology, 46*, 470-474.

Hayes, K.J. & Hayes, C. (1954). The cultural capacity of chimpanzee. *Human Biology; An International Record of Research, 26*(3), 288-303.

Hayes, K.J & Nissen, C.H. (1971). Higher mental functions of a home-raised chimpanzee. In A.M. Schrier & F. Stollnitz (Eds.), *Behavior of Nonhuman Primates, Vol. 4*. (pp. 59-115) New York & London: Academic Press.

Hayes, K.J., Thompson, R. & Hayes, C. (1953a). Discrimination learning set in chimpanzees. *Journal of Comparative and Physiological Psychology, 46*(2), 99-104.

Hayes, K.J., Thompson, R. & Hayes, C. (1953b). Concurrent discrimination learning in chimpanzees. *Journal of Comparative and Physiological Psychology, 46*(2), 105-107.

Herrmann, E., Call, J., Hernández-Lloreda, M.V., Hare, B. & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science, 317*, 1360-1366.

Herrmann, E., Melis, A.P. & Tomasello, M. (2005). Apes' use of iconic cues in the object-choice task. *Animal Cognition, 9*(2), 118-130.

Herrnstein, R.J. (1990). Levels of stimulus control: a functional approach. *Cognition, 37*, 133-166.

Herskovits, M. (1959). Art and value. In R. Redfield, M. Herskovits & G.F. Ekholm (Eds.), *Aspects of Primitive Art*. New York: Museum of Primitive Art.

Hess, L. (1954). *Christine the Baby Chimp*. London & Southampton: The Camelot Press, Ltd.

Hirata, S. (2007). A note on the responses of chimpanzees (*Pan troglodytes*) to live self-images on television monitors. *Behavioural Processes, 75*(1), 85-90.

Hochberg, J. (1980). Pictorial functions and perceptual structures. In M.A. Hagen (Ed.), *The Perception of Pictures, Vol. 2* (pp. 47-93). New York: Academic Press, Inc.

Hochberg, J. & Brooks, V. (1962). Pictorial recognition as an unlearned ability: a study of one child's performance. *The American Journal of Psychology, 75*(4), 624-628.

Hoffman, K.L., Gothard, K.M., Schmid, M.C. & Logothetis, N.K. (2007). Facial-expression and gaze-selective responses in the monkey amygdala. *Current Biology, 17*, 766-772.

Hopkins, W.D. & Savage-Rumbaugh, E.S. (1991). Vocal communication as a function of differential rearing experiences in *Pan paniscus*: a preliminary report. *International Journal of Primatology, 12*(6), 559-583.

Hopkins, W.D. & Washburn, D.A. (2002). Matching visual stimuli on the basis of global and local features by chimpanzees (*Pan troglodytes*) and rhesus monkeys (*Macaca mulatta*). *Animal Cognition, 5*, 27-31.

Horn, H.A. & Angrist Myers, N. (1978). Memory for location and picture cues at ages two and three. *Child Development, 49*, 845-856.

Horn Ratner, H. & Angrist Myers, N. (1980). Related picture cues and memory for hidden-object locations at age two. *Child Development, 51*, 561-564.

Hornaday, W.T. (1922). *The Minds and Manners of Wild Animals*. New York: Charles Scribner's Sons.

Hoyt, A.M. (1941). *Toto and I*. Philadelphia & New York: J.B. Lippincott Company.

Hudson, W. (1960). Pictorial depth perception in sub-cultural groups in Africa. *Journal of Social Psychology, 52*, 183-208.

Hudson, W. (1962). Cultural problems in pictorial perception. *South African Journal of Science, 58*, 189-195.

Humphrey, N.K. (1974). Species and individuals in the perceptual world of monkeys. *Perception, 3*, 105-114.

Huxley, J.S. (1942). Origins of human graphic art. *Nature, 149*, 637.

Hyatt, C.W. & Hopkins, W.D. (1994). Self-awareness in bonobos and chimpanzees: a comparative perspective. In S.T. Parker, R.W. Mitchell & M.L. Boccia (Eds.), *Self-awareness in Animals and Humans. Developmental Perspectives* (pp. 248-253). New York: Cambridge University Press.

Inoue-Nakamura, N. (1997). Mirror self-recognition in nonhuman primates: a phylogenetic approach. *Japanese Psychological Research, 39*(3), 266-275.

Inoue-Nakamura, N. (2001). Mirror self-recognition in primates: an ontogenetic and a phylogenetic approach. In T. Matsuzawa (Ed.), *Primate Origins of Human Cognition and Behavior* (pp. 297-312). Tokyo: Springer.

Itakura, S. (1987a). Mirror guided behavior in Japanese monkeys (*Macaca fuscata fuscata*). *Primates, 28*(2), 149-161.

Itakura, S. (1987b). Use of a mirror to direct their responses in Japanese monkeys (*Macaca fuscata fuscata*). *Primates, 28*(3), 343-352.

Itakura, S. (1994). Recognition of line-drawing representations by a chimpanzee (*Pan troglodytes*). *The Journal of General Psychology, 121*(3), 189-197.

Ittelson, W.H. (1996). Visual perceptions of markings. *Psychonomic Bulletin and Review, 3*(2), 171-187.

Iversen, I. & Matsuzawa, T. (1997). Model-guided line drawing in the chimpanzee (*Pan troglodytes*). *Japanese Psychological Research, 39*(3), 154-181.

Iversen, I. & Matsuzawa, T. (2001). Establishing line tracing on a touch monitor as a basic drawing skill in chimpanzees (*Pan troglodytes*). In T. Matsuzawa (Ed.), *Primate Origins of Human Cognition and Behavior* (pp. 235-268). Tokyo: Springer.

Iwai, E., Yaginuma, S. & Mishkin, M. (1986). Acquisition of discrimination learning of patterns identical in configuration in macaques (*Macaca mulatta* and *M. fuscata*). *Journal of Comparative Psychology, 100*(1), 30-36.

Izumi, A. & Kojima, S. (2004). Matching vocalization to vocalizing faces in a chimpanzee (*Pan troglodytes*). *Animal Cognition, 7*, 179-184.

Jarvik, M.E. (1953). Discrimination of colored food and food signs by primates. *Journal of Comparative and Physiological Psychology, 46*, 390-392.

Jarvik, M.E. (1956). Simple color discrimination in chimpanzees: effect of varying contiguity between cue and incentive. *Journal of Comparative and Physiological Psychology, 49*, 492-495.

Jitsumori, M. & Delius, J.D. (2001). Object recognition and object categorization in animals. In T. Matsuzawa (Ed.), *Primate Origins of Human Cognition and Behavior* (pp. 269-293). Tokyo: Springer.

Jitsumori, M. & Matsuzawa, T. (1991). Picture perception in monkeys and pigeons: transfer of rightside-up versus upside-down discrimination of photographic objects across conceptual categories. *Primates, 32*(4), 473-482.

Jones, R.K. & Hagen, M.A. (1980). A perspective on cross-cultural picture perception. In M.A. Hagen (Ed.), *The Perception of Pictures, Vol. 2* (pp. 193-226). New York: Academic Press, Inc.

Kanazawa, S. (1996). Recognition of facial expressions in a Japanese monkey (*Macaca fuscata*) and humans (*Homo sapiens*). *Primates, 37*(1), 25-38.

Katz, J.S. & Wright, A.A. (2006). Same/different abstract-concept learning by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 32*(1), 80-86.

Katz, J.S., Wright, A.A. & Bachevalier, J. (2002). Mechanisms of same/different abstract-concept learning by rhesus monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes, 28*(4), 358-368.

Kawai, N. & Matsuzawa, T. (2001). Reproductive memory processes in chimpanzees: homologous approaches to research on human working memory. In T. Matsuzawa (Ed.), *Primate Origins of Human Cognition and Behavior* (pp. 226-234). Tokyo: Springer.

Keating, C.F. & Keating, E.G. (1993). Monkeys and mug shots: cues used by rhesus monkeys (*Macaca mulatta*) to recognize a human face. *Journal of Comparative Psychology, 107*(2), 131-139.

Kellogg, W.N. (1968). Chimpanzees in experimental homes. *The Psychological Record, 18*, 489-498.

Kellogg, W.N. & Kellogg, L.A. (1933/1967). *The Ape and The Child. A Study of Environmental Influence Upon Early Behavior*. New York: Hafner Publishing Company, Inc.

Kidd, D. (1904). *The essential Kafir*. London: Adam and Charles Black.

Kilbridge, P.L. & Robbins, M.C. (1969). Pictorial depth perception and acculturation among the Baganda. *American Anthropologist, 71*, 293-301.

Kirkpatrick-Steger, K., Wasserman, E.A. & Biederman, I. (1998). Effects of geon deletion, scrambling , and movement on picture recognition in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 24*(1), 34-46.

Kirkpatrick-Steger, K., Wasserman, E.A. & Biederman, I. (2000). The pigeon's discrimination of shape and location information. *Visual cognition, 7*(4), 417-436.

Kitchen, A., Denton, D. & Brent, L. (1996). Self-recognition and abstraction abilities in the common chimpanzee studied with distorting mirrors. *Proceedings of the National Academy of Sciences of the United States of America, 93*, 7405-7408.

Köhler, W. (1922). Zur Psychologie des Schimpansen. *Psychological Research, 1*(1), 2-46.

Köhler, W. (1925/1957). *The Mentality of Apes*. London & Tonbridge: Penguin Books.

Kojima, S. (2003). Experimental analyses of body image in the chimpanzee. *Primates, 44*, 231-238.

Kojima, S., Izumi, A. & Ceugniet, M. (2003). Identification of vocalizers by pant hoots, pant grunts and screams in a chimpanzee. *Primates, 44*, 225-230.

Kuhlmeier, V.A. & Boysen, S.T. (2001). The effect of contingencies on scale model task performance by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 115*(3), 300-306.

Kuhlmeier, V.A. & Boysen, S.T. (2002). Chimpanzees (*Pan troglodytes*) recognize spatial and object correspondences between a scale model and its referent. *Psychological Science, 13*(1), 60-63.

Kuhlmeier, V.A., Boysen, S.T. & Mukobi, K.L. (1999). Scale-model comprehension by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 113*(4), 396-402.

Kyes, R.C., Mayer, K.E. & Bunnell, B.N. (1992). Perception of stimuli presented as photographic slides in cynomolgus macaques (*Macaca fascicularis*). *Primates, 33*(3), 407-412.

Ladygina-Kohts, N.N. (1935/2002). Infant chimpanzee and human child. In F.B.M. de Waal (Ed.), *Infant Chimpanzee and Human Child. A Classic 1935 Comparative Study of Ape Emotions and Intelligence* (pp. 1-407). New York: Oxford University Press.

Law, L.E. & Lock, A.J. (1994). Do gorillas recognize themselves on television? In S.T. Parker, R.W. Mitchell & M.L. Boccia (Eds.), *Self-Awareness in Animals and Humans. Developmental Perspectives* (pp. 308-312). New York: Cambridge University Press.

Leekam, S.R. & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition, 40*, 203-218.

Lenain, T. (1997). *Monkey Painting*. London: Reaktion Books Ltd.

Lechelt, D.P. & Spetch, M.L. (1997). Pigeon's use of landmarks for spatial search in a laboratory arena and in digitized imagses of the arena. *Learning and Motivation, 28*, 424-448.

Lewis, C., Russell, C. & Berridge, D. (1993). When is a mug not a mug? Effects of content, naming, and instruction on children's drawings. *Journal of Experimental Child Psychology, 56*, 291-302.

Liben, L.S. (2003). Beyond point and shoot: children's developing understanding of photographs as spatial and expressive representations. *Advances in Child Development and Behavior, 31*, 1-42.

Lin, A.C., Bard, K.A. & Anderson, J.R. (1992). Development of self-recognition in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 106*(2), 120-127.

Lloyd, A.B. (1904). Acholi country, Part 2. *Uganda Notes, 5*, 18-22.

Looney, T.A. & Cohen, P.S. (1974). Pictorial target control of schedule-induced attack in white Carneaux pigeons. *Journal of the Experimental Analysis of Behavior, 21*, 571-584.

Lorincz, E.N., Gómez, J-C. & Perrett, D.I. (2003). *Spontaneous attention following: a comparative study in human adults and rhesus macaques*. Poster session presented at the spring meeting of the Primate Society of Great Britain, University of St. Andrews, St. Andrews.

Loughlin, K.A. & Daehler, M.W. (1973). The effects of distraction and added perceptual cues on the delayed reaction of very young children. *Child Development, 44*, 384-388.

Loveland, K.A. (1986). Discovering the affordances of a reflecting surface. *Developmental Review, 6*(1), 1-24.

Lyn, H. (2007). Mental representation of symbols as revealed by vocabulary errors in two bonobos (*Pan paniscus*). *Animal Cognition, 10*(4), 461-475.

Lyn, H., Greenfield, P. & Savage-Rumbaugh, S. (2006). The development of representational play in chimpanzees and bonobos: evolutionary implications, pretence, and the role of interspecies communication. *Cognitive Development, 21*, 199-213.

Malone, D.R., Tolan, J.C. & Rogers, C.M. (1980). Cross-modal matching of objects and photographs in the monkey. *Neuropsychologia, 18*, 693-697.

Maple, T.L. & Hoff, M.P. (1982). *Gorilla Behavior*. New York: Van Nostrand Reinhold.

Martin-Malivel, J. & Fagot, J. (2001). Perception of pictorial human faces by baboons: effects of stimulus orientation on discriminative performance. *Animal Learning and Behavior, 29*(1), 10-20.

Martin-Malivel, J., Mangini, M.C., Fagot, J. & Biederman, I. (2006). Do humans and baboons use the same information when categorizing human and baboon faces? *Psychological Science, 17*(7), 599-607.

Matevia, M.L., Patterson, F.G.P. & Hillix, W.A. (2002). Pretend play in a signing gorilla. In R.W. Mitchell (Ed.), *Pretending and imagination in animals and children* (pp. 285-304). Cambridge: Cambridge University Press.

Marzolf, D.P., DeLoache, J.S. & Kolstad, V. (1999). The role of relational similarity in young children's use of a scale model. *Developmental Science, 2*(3), 296-305.

Matsukawa, A., Inoue, S. & Jitsumori, M. (2004). Pigeon's recognition of cartoons: effects of fragmentation, scrambling, and deletion of elements. *Behavioural Processes, 65*, 25-34.

Matsuzawa, T. (2003). The Ai project: historical and ecological contexts. *Animal Cognition, 6*, 199-211.

McClearn, G.E. & Harlow, H.H. (1954). The effect of spatial contiguity on discrimination learning by rhesus monkeys. *Journal of Comparative and Physiological Psycholog, 47*(5), 391-394.

Menzel, E.W. (1973). Chimpanzee spatial memory organization. *Science, 182*, 943-945.

Menzel, E.W., Jr., & Davenport, R.K., Jr. (1961). Preference for clear versus distorted viewing in the chimpanzee. *Science, 134*, 1531.

Menzel, E.W., Jr., Premack, D. & Woodruff, G. (1978). Map reading by chimpanzees. *Folia Primatologica, 29*(4), 241-249.

Menzel, E.W., Jr., Savage-Rumbaugh, E.S. & Lawson, J. (1985). Chimpanzee (*Pan troglodytes*) spatial problem solving with the use of mirrors and televised equivalents of mirrors. *Journal of Comparative Psychology, 99*(2), 211-217.

Meyer, D.R., Treichler, F.R. & Meyer, P.M. (1965). Discrete-trial training techniques and stimulus variables. In A.M. Schrier, H.F. Harlow & F. Stollnitz (Eds.), *Behavior of Nonhuman Primates, Vol. 1*. (pp. 1-49) New York & London: Academic Press.

Miles, H.L.W. (1990). The cognitive foundations for reference in a signing orangutan. In S.T. Parker & K.R. Gibson (Eds.), *"Language" and Intelligence in Monkeys and Apes: Comparative Developmental Perspectives* (pp. 511-539). New York: Cambridge University Press.

Miles, H.L.W. (1994). Me Chantek: the development of self-awareness in a signing orangutan. In S.T. Parker, R.W. Mitchell & M.L. Boccia (Eds.), *Self-awareness in Animals and Humans. Developmental Perspectives* (pp. 254-272). New York: Cambridge University Press.

Miles, H.L., Mitchell, R.W. & Harper, S.E. (1996). Simon says: the development of imitation in an enculturated orangutan. In A.E. Russon, K.A. Bard & S.T. Parker (Eds.), *Reaching Into Thought. The Minds of the Great Apes* (pp. 278-299). Cambridge: Cambridge University Press.

Miller, R.J. (1973). Cross-cultural research in the perception of pictorial materials. *Psychological Bulletin, 80*(2), 135-150.

Mitchell, R.W. (1990). A theory of play. In M. Bekoff & D. Jamieson (Eds.), *Interpretation and Explanation in the Study of Animal Behavior, Vol. 1* (pp. 197-227). Boulder: Westview Press.

Mitchell, R.W. (1997). Kinesthetic-visual matching and the self-concept as explanations of mirror-self-recognition. *Journal of the Theory of Social Behaviour 27*(1), 17-39.

Mitchell, R.W. (1999). Scientific and popular conceptions of the psychology of great apes from the 1790s to the 1970s: déjà vu all over again. *Primate Report, 53*, 3-118.

Mitchell, R.W. (2002). Imaginative animals, pretending children. In R.W. Mitchell (Ed.), *Pretending and imagination in animals and children* (pp. 3-22). Cambridge: Cambridge University Press.

Moore, D.S. (2003). *The Dependent Gene: The Fallacy of "Nature vs Nurture"*. New York: Henry Holt & Company, LLC.

Morimura, N. & Matsuzawa, T. (2001). Memory of movies by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 115*(2), 152-158.

Morris, D. (1963). *Apor och konst. En studie i människoapornas bildskapande ("The Biology of Art", 1962).* Falköping: Natur och Kultur.

Morris, D. (1980). *Animal Days*. New York: William Morrow & Company.

Murai, C., Kosugi, D., Tomonaga, M., Tanaka, M., Matsuzawa, T. & Itakura, S. (2005). Can chimpanzee infants (*Pan troglodytes*) form categorical representations in the same way as human infants (*Homo sapiens*)? *Developmental Science, 8*(3), 240-254.

Myowa-Yamakoshi, M., Tomonaga, M., Tanaka, M. & Matsuzawa, T. (2003). Preference for human direct gaze in infant chimpanzees (*Pan troglodytes*). *Cognition, 89*, B53-B64.

Myowa-Yamakoshi, M., Yamaguchi, M.K., Tomonaga, M., Tanaka, M. & Matsuzawa, T. (2005). Development of face recognition in infant chimpanzees (*Pan troglodytes*). *Cognitive Development, 20*, 49-63.

Nakamura, T., Croft, D.B. & Westbrook, R.F. (2003). Domestic pigeons (*Columba livia*) discriminate between photographs of individual pigeons. *Learning and Behavior, 31*(4), 307-317.

Neiworth, J.J., Parsons, R.R. & Hassett, J.M. (2004). A test of the generality of perceptually based categories found in infants: attentional differences toward natural kinds by New World monkeys. *Developmental Science, 7*(2), 185-193.

Neiworth, J.J. & Wright, A.A. (1994). Monkeys (*Macaca mulatta*) learn category matching in a nonidentical same-different task. *Journal of Experimental Psychology: Animal Behavior Processes, 20*(4), 429-435.

Newman, C.V. (1969). Children's size judgements in a picture with suggested depth. *Nature, 223*, 418-420.

Nissen, H.W. (1931). A field study of the chimpazee. Observations of chimpanzee behavior and environment in Western French Guinea. *Comparative Psychology Monographs, 8*(1).

Nissen, H.W., Riesen, A.H. & Nowlis, V. (1938). Delayed response and discrimination learning by chimpanzees. *Journal of Comparative Psychology, 26*, 361-386.

Oden, D.L., Thompson, R.K.R. & Premack, D. (1988). Spontaneous transfer of matching by infant chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes, 14*(2), 140-145.

Oden, D.L., Thompson, R.K.R. & Premack, D. (1990). Infant chimpanzees spontaneously perceive both concrete and abstract same/different relations. *Child Development, 61*, 621-631.

Oden, D.L., Thompson, R.K.R. & Premack, D. (2001). Can apes reason analogically? Comprehension and production of analogical problems by Sarah, a chimpanzee (*Pan troglodytes*). In D. Gentner, K.J. Holyoak & B.N. Kokinov (Eds.), *Analogy: Theory and Phenomena* (pp. 472-497). Cambridge, USA: MIT Press.

Olson, R., Yonas, A. & Cooper, R. (1980). Development of pictorial perception. In M.A. Hagen (Ed.), *The Perception of Pictures, Vol. 2* (pp. 155-192). New York: Academic Press, Inc.

Omari, I.M. & MacGinitie, W.H. (1974). Some pictorial artifacts in the studies of African children's pictorial depth perception. *Child Development, 45*, 535-539.

Overman, W.H., Jr. & Doty, R.W. (1982). Hemispheric specialization displayed by man but not macaques for analysis of faces. *Neuropsychologia, 20*(2), 113-128.

Parker, S.T. (1994). Incipient mirror self-recognition in zoo gorillas and chimpanzees. In S.T. Parker, R.W. Mitchell & M.L. Boccia (Eds.), *Self-awareness in Animals and Humans. Developmental Perspectives* (pp. 301-307). New York: Cambridge University Press.

Parr, L.A. (2001). Cognitive and physiological markers of emotional awareness in chimpanzees (*Pan troglodytes*). *Animal Cognition, 4*, 223-229.

Parr, L.A. (2004). Perceptual biases for multimodal cues in chimpanzee (*Pan troglodytes*) affect recognition. *Animal Cognition, 7*, 171-178.

Parr, L.A., de Waal, F.B.M. (1999). Visual kin recognition in chimpanzees. *Nature, 399*, 647-648.

Parr, L.A., Dove, T. & Hopkins, W.D. (1998). Why faces may be special: evidence of the inversion effect in chimpanzees. *Journal of Cognitive Neuroscience, 10*(5), 615-622.

Parr, L.A., Heintz, M. (2006). The perception of unfamiliar faces and houses by chimpanzees: influence of rotation angle. *Perception, 35*, 1473-1483.

Parr, L.A., Heintz, M. & Akamagwuna, U. (2006). Three studies on configural face processing by chimpanzees. *Brain and Cognition, 62*, 30-42.

Parr, L.A. & Hopkins, W.D. (2000). Brain temperature asymmetries and emotional perception in chimpanzees, *Pan troglodytes*. *Physiology and Behavior, 71*, 363-371.

Parr, L.A., Winslow, J.T. & Hopkins, W.D. (1999). Is the inversion effect in rhesus monkeys face-specific? *Animal Cognition, 2*, 123-129.

Parr, L.A., Winslow, J.T., Hopkins, W.D. & de Waal, F.B.M. (2000). Recognizing facial cues: individual discrimination by chimpanzees (*Pan troglodytes*) and rhesus monkeys (*Macaca mulatta*). *Journal of Comparative Psychology, 114*(1), 47-60.

Pascalis, O., Petit, O., Kim, J.H. & Campbell, R. (2000). Picture perception in primates: the case of face perception. In J. Fagot (Ed.), *Picture Perception in Animals* (pp. 263-194). Hove: Psychology Press Ltd.

Patterson, F.G. (1978a). The gestures of a gorilla: language acquisition in another pongid. *Brain and Language, 5*, 72-97.

Patterson, F.G. (1978b). Conversations with a gorilla. *National Geographic Magazine, 154*(4), 438-465.

Patterson, F.G. (1980). Innovative uses of language by a gorilla: a case study. In K.E. Nelson (Ed.), *Children's Language, vol. 2* (pp. 497-561). New York: Gardner Press.

Patterson, F.G.P. & Cohn, R.H. (1990). Language acquisition by a lowland gorilla: Koko's first ten years of vocabulary development. *Word, 41*(2), 97-143.

Patterson, F.G.P. & Cohn, R.H. (1994). Self-recognition and self-awareness in lowland gorillas. In S.T. Parker, R.W. Mitchell & M.L. Boccia (Eds.), *Self-Awareness in Animals and Humans. Developmental Perspectives* (pp. 273-290). New York: Cambridge University Press.

Patterson, F.G.P. & Gordon, W. (2001). Twenty-seven years of project Koko and Michael. In B.M.F. Galdikas, N.E. Briggs, L.K. Sheeran, G.L. Shapiro & J. Goodall (Eds.), *All Apes Great and Small Volume 1: African Apes* (pp. 165-176). New York: Kluwer Academic/Plenum Publishers.

Patterson, F. & Linden, E. (1981). *The Education of Koko*. New York: Holt, Rinehart, and Winston.

Peissig, J.J., Kirkpatrick, K., Young, M.E., Wasserman, E.A. & Biederman, I. (2006). Effects of varying stimulus size on object recognition in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 32*(4), 419-430.

Peissig, J.J., Young, M.E., Wasserman, E.A. & Biederman, I. (2000a). Seeing things from a different angle: the pigeon's recognition of single geons rotated in depth. *Journal of Experimental Psychology: Animal Behavior Processes, 26*(2), 115-132.

Peissig, J.J., Young, M.E., Wasserman, E.A. & Biederman, I. (2000b). The pigeon's perception of depth-rotated shapes. In J. Fagot (Ed.), *Picture Perception in Animals* (pp. 37-70). Hove: Psychology Press Ltd.

Peralta de Mendoza, O.A. & Marcela Salsa, A. (2003). Instruction in early comprehension and use of a symbol-referent relaiton. *Cognitive Development, 18*, 269-284.

Perkins, D.N. & Hagen, M.A. (1980). Convention, context and caricature. In M.A. Hagen (Ed.), *The Perception of Pictures, Vol. 1* (pp. 257-286). New York: Academic Press, Inc.

Petit, O. & Thierry, B. (1993). Reactions to shadows in captive Guinea baboons. *Human Evolution, 9*(3), 257-260.

Phelps, M.T. & Roberts, W.A. (1994). Memory for pictures of upright and inverted primate faces in humans (*Homo sapiens*), squirrel monkeys (*Saimiri sciureus*), and pigeons (*Columba livia*). *Journal of Comparative Psychology, 108*(2), 114-125.

Pierroutsakos, S.L. (1998). Infants' manual investigation of pictures as a function of exposure to the objects depicted. *Infant Behavior and Development, 21*, 622.

Pierroutsakos, S.L. & DeLoache, J.S. (2003). Infants' manual exploration of pictorial objects varying in realism. *Infancy, 4*(1), 141-156.

Pierroutsakos, S.L., DeLoache, J.S., Gound, M. & Bernard, E.N. (2005). Very young children are insensitive to picture- but not object-orientation. *Developmental Science, 8*(4), 326-332.

Pierroutsakos, S.L. & Troseth, G.L. (2003). Video Verité: infants' manual investigation of objects on video. *Infant Behaviour and Development, 26*, 183-199.

Plimpton, E.H. & Swartz, K.B. & Rosenblum, L.A. (1981). Responses of juvenile bonnet macaques to social stimuli presented through color videotapes. *Developmental Psychobiology, 14*, 109-115.

Plotnik, J.M., de Waal, F.B.M & Reiss, D. (2006). Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences of the United States of America, 103*(45), 17053-17057.

Posada, S. & Colell, M. (2007). Another gorilla (*Gorilla gorilla gorilla*) recognizes himself in a mirror. *American Journal of Primatology, 69*, 1-8.

Poss, S.R. & Rochat, P. (2003). Referential understanding of videos in chimpanzees (*Pan troglodytes*), orangutans (*Pongo pygmaeus*), and children (*Homo sapiens*). *Journal of Comparative Psychology, 117*(4), 420-428.

Povinelli, D.J., Rulf, A.B., Landau, K.R. & Bierschwale, D.T. (1993). Self-recognition in chimpanzees (*Pan troglodytes*): distribution, ontogeny, and patterns of emergence. *Journal of Comparative Psychology, 107*(4), 347-372.

Povinelli, D.J. & Simon, B.B. (1998). Young children's understanding of briefly versus extremely delayed images of the self: emergence of the autobiographical stance. *Developmental Psychology, 34*(1), 188-194.

Preissler, M.A. & Bloom, P. (2007). Two-year olds appreciate the dual nature of pictures. *Psychological Science, 18*(1), 1-2.

Preissler, M.A. & Carey, S. (2004). Do both pictures and words function as symbols for 18- and 24-month-old children? *Journal of Cognition and Development, 5*(2), 185-212.

Preissler, M.A. & Carey, S. (2005). The role of inferences about referential intent in word learning: evidence from autism. *Cognition, 97*, B13-B23.

Premack, A.J. & Premack, D. (1972). Teaching language to an ape. *Scientific American, 227*(4), 92-99.

Premack, D. (1971). Language in chimpanzee? *Science, 172*, 808-822.

Premack, D. (1975). Putting a face together. *Science, 188*, 228-236.

Premack, D. (1976). *Intelligence in Ape and Man*. New Jersey: Lawrence Erlbaum Associates, Inc.

Premack, D. (1983). The codes of man and beasts. *The Behavioral and Brain Sciences, 6*, 125-167.

Premack, D. & Premack, A.J. (1983). *The Mind of an Ape*. New York: W.W. Norton & Company, Inc.

Premack, D. & Woodruff, G. (1978a). Chimpanzee problem-solving. A test for comprehension. *Science, 202*, 532-535.

Premack, D. & Woodruff, G. (1978b). Does the chimpanzee have a theory of mind? *The Behavioral and Brain Sciences, 4*, 515-526.

Premack, D. & Woodruff, G. (1979). Chimpanzee problem comprehension: insufficient evidence. *Science, 206*, 1202.

Premack, D., Woodruff, G. & Kennel, K. (1978). Paper-marking test for chimpanzee: simple control for social cues. *Science, 202*, 903-905.

Quinn, P.C., Eimas, P.D. & Tarr, M.J. (2001). Perceptual categorization of cat and dog silhouettes by 3- to 4-month-old infants. *Journal of Experimental Child Psychology, 79*, 78-94.

Radick, G. (2005). Primate language and the playback experiment, in 1890 and 1980. *Journal of the History of Biology, 38*(3), 461-493.

Reid, S.L. & Spetch, M.L. (1998). Perception of pictorial depth cues by pigeons. *Psychonomic Bulletin and Review, 5*(4), 698-704.

Reiss, D. & Marino, L. (2001). Mirror self-recognition in the bottlenose dolphin: a case of cognitive emergence. *Proceedings of the National Academy of Sciences in the United States of America, 98*(10), 5937-5942.

Ringo, J.L., Lewine, J.D. & Doty, R.W. (1986). Comparable performance by man and macaque on memory for pictures. *Neuropsychologia, 24*(5), 711-717.

Roberts, W.A. & Mazmanian, D.S. (1988). Concept learning at different levels of abstraction by pigeons, monkeys, and people. *Journal of Experimental Psychology: Animal Behavior Processes, 14*(3), 247-260.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology, 7*, 532-547.

Rose, S.A., Gottfried, A.W. & Bridger, W.H. (1982). Infant's cross-modal transfer from solid objects to their graphic representations. *Child Development, 54*, 686-694.

Rosenfeld, S.A. & Van Hoesen, G.W. (1979). Face recognition in the rhesus monkey. *Neuropsychologia, 17*, 503-509.

Rumbaugh, D.M. (Ed.) (1977). *Language Learning by a Chimpanzee. The LANA Project*. New York: Academic Press, Inc.

Rumbaugh, D.M. (1997). The psychology of Harry F. Harlow: a bridge from radical to rational behaviorism. *Philosophical Psychology, 10*(2), 197-211.

Rumbaugh, D.M. & Gill, T.V. (1976). Language and the acquisition of language-type skills by a chimpanzee. *Annals of the New York Academy of Sciences, 270*(1), 90-123.

Rumbaugh, D.M., Richardson, W.K., Washburn, D.A., Savage-Rumbaugh, E.S. & Hopkins, W.D. (1989). Rhesus monkeys (*Macaca mulatta*), video tasks, and implications for stimulus-response spatial contiguity. *Journal of Comparative Psychology, 103*(1), 32-38.

Rumbaugh, D.M & Washburn, D.A. (2003). *Intelligence of Apes and Other Rational Beings*. New Haven: Yale University Press.

Ryan, C.M.E. & Lea, S.E.G. (1994). Images of conspecifics as categories to be discriminated by pigeons and chickens: slides, video tapes, stuffed birds and live birds. *Behavioural Processes, 33*, 155-176.

Sackett, G.P. (1965). Response of rhesus monkeys to social stimulation presented by means of colored slides. *Perceptual and Motor Skills, 20*, 1027-1028.

Sackett, G.P. (1966). Monkeys reared in isolation with pictures as visual input: evidence for an innate releasing mechanism. *Science, 154*, 1468-1473.

Saito, A., Mikami, A., Hasegawa, T., Koida, K., Terao, K., Koike, S., Onishi, A., Takenaka, O., Teramoto, M. & Mori, Y. (2003). Behavioral evidence of color vision deficiency in a protanomalia chimpanzee (*Pan troglodytes*). *Primates, 44*, 171-176.

Salsa, A.M. & Peralta de Mendoza, O. (2007). Routes to symbolization: intentionality and correspondence in early understanding of pictures. *Journal of Cognition and Development, 8*(1), 79-92.

Sato, N. & Nakamura, K. (2001). Detection of directed gaze in rhesus monkeys (*Macaca mulatta*). *Journal of Comparative Psychology, 115*(2), 115-121.

Savage-Rumbaugh, E.S. (1986). *Ape Language. From Conditioned Response to Symbol.* New York: Columbia University Press.

Savage-Rumbaugh, S., Fields, W.M. & Taglialatela, J.P. (2001). Language, speech, tools and writing. A cultural imperative. *Journal of Consciousness Studies, 8*(5-7), 273-292.

Savage-Rumbaugh, S. & Lewin, R. (1994). *Kanzi. The Ape at the Brink of the Human Mind.* London: Doubleday.′

Savage-Rumbaugh, S., McDonald, K., Sevcik, R.A., Hopkins, W.D. & Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *Journal of Experimental Psychology: General, 115*(3), 211-235.

Savage-Rumbaugh, E.S., Murphy, J., Sevcik, R.A., Brakke, K.E., Williams S.L., Rumbaugh, D.M. & Bates, E. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development, 58*(3/4), i-252.

Savage-Rumbaugh, E.S. & Rubert, E. (1986). Video representation of reality. In E.S. Savage-Rumbaugh, *Ape Language. From Conditioned Response to Symbol* (pp. 299-323). New York: Columbia University Press.

Savage-Rumbaugh, E.S. & Rumbaugh, D.M. (1979). Chimpanzee problem comprehension: insufficient evidence. *Science, 206*, 1201-1202.

Savage-Rumbaugh, E.S., Rumbaugh, D.M., Fields, W.M. (2006). Language as a window on rationality. In S. Hurley & M. Nudds (Eds.), *Rational Animals?* (pp. 513-552). New York: Oxford University Press.

Savage-Rumbaugh, E.S., Rumbaugh, D.M., Smith, S.T. & Lawson, J. (1980). Reference: the linguistic essential. *Science, 210*, 922-925.

Savage-Rumbaugh, S., Sevcik, R.A. & Hopkins, W.D. (1988). Symbolic cross-modal transfer in two species of chimpanzees. *Child Development, 59*, 617-625.

Savage-Rumbaugh, S., Shanker, S.G. & Taylor, T.J. (1998). *Apes, Language, and the Human Mind.* New York: Oxford University Press.

Schroeder, B. (Director). (1978/2006). *Koko: A Talking Gorilla* [DVD]. Les Filmes du Losange/The Criterion Collection.

Schwartz, B.L, Colon, M.R., Sanchez, I.C., Alexa Rodriguez, I. & Evans, S. (2002). Single-trial learning of "what" and "who" information in a gorilla (*Gorilla gorilla gorilla*): implications for episodic memory. *Animal Cognition, 5*, 85-90.

Schwartz, B.L., Hoffman, M.L. & Evans, S. (2005). Episodic-like memory in a gorilla: a review and new findings. *Learning and Motivation, 36*, 226-244.

Schwartz, B.L., Meissner, C.A., Hoffman, M., Evans, S. & Frazier, L.D. (2004). Event memory and misinformation effects in a gorilla (*Gorilla gorilla gorilla*). *Animal Cognition, 7*, 93-100.

Schwartz, D.L. (1995). Reasoning about the referent of a picture versus reasoning about the picture as the referent: an effect of visual realism. *Memory and Cognition, 23*(6), 709-722.

Segall, M.H., Campbell, D.T. & Herskovits, M.J. (1963). Cultural differences in the perception of geometric illusions. *Science, 139*, 769-771.

Segall, M.H., Campbell, D.T. & Herskovits, M.J. (1966). *The Influence of Culture on Visual Perception*. Indianapolis: Bobbs-Merrill.

Segerdahl, P., Fields, W. & Savage-Rumbaugh, S. (2005). *Kanzi's Primal Language. The Cultural Initiation of Primates into Language*. Basingstoke: Palgrave Macmillan.

Shapiro, G.L. & Galdikas, B.M.F. (1999). Early sign performance in a free-ranging, adult orangutan. In S.T. Parker, R.W. Mitchell & H.L. Miles (Eds.), *The Mentalities of Gorillas and Orangutans. Comparative Perspectives* (pp.265-179). Cambridge: Cambridge University Press.

Sharon, T. & DeLoache, J.S. (2003). The role of perseveration in children's symbolic understanding and skill. *Developmental Science, 6*(3), 289-296.

Shaw, B. (1969). *Visual Symbols Survey*. London: Centre for Educational Development Overseas.

Shyne, A. (2006). Meta-analytic review of the effects of enrichment on stereotypic behavior in zoo mammals. *Zoo Biology, 25*, 317-337.

Schiller, P.H. (1951). Figural preferences in the drawings of a chimpanzee. *Journal of Comparative and Physiological Psychology, 44*, 101-111.

Schillito, D.J., Gallup, G.G., Jr. & Beck, B.B. (1999). Factors affecting mirror behaviour in western lowland gorillas, *Gorilla gorilla. Animal Behaviour, 57*, 999-1004.

Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W. & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of Experimental Analysis of Behavior, 37*, 23-44.

Sidman, M. & Tailby, W. (1982). Conditional discrimination vs. matching to sample: an expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37*, 5-22.

Slater, A., Rose, D. & Morrison, V. (1984). Newborn infants' perception of similarities and differences between two- and three-dimensional stimuli. *British Journal of Developmental Psychology, 2*, 287-294.

Slater, M., Antley, A., Davison, A., Swapp, D., Guger, C., Barker, C., Pistrang, N. & Sanchez-Vives, M.V. (2006). A virtual reprise of the Stanley Milgram obedience experiments. *PloS ONE, 1*(1), e39.

Slocombe, K.E. & Zuberbühler, K. (2005). Functionally referential communication in a chimpanzee. *Current Biology, 15*, 1779-1784.

Smith, D.A. (1973). Systematic study of chimpanzee drawing. *Journal of Comparative and Physiological Psychology, 82*(3), 406-414.

Smith, C.M. & DeLoache, J.S. (1996). Not picture perfect: young children's use of a picture as a symbol for themselves. *Infant Behavior and Development, 19*, suppl. 1, 747.

Smith, H.J., King, J.E., Witt, E.D. & Rickel, J.E. (1975). Sameness-difference matching from sample by chimpanzees. *Bulletin of the Psychonomic Society, 6*(5), 469-471.

Sonesson, G. (1989). *Pictorial Concepts. Inquiries into the semiotic heritage and its relevance for the analysis of the visual world.* Lund: Aris/Lund University Press.

Sonesson, G. (1992). The semiotic function and the genesis of pictorial meaning. In E. Tarasti (Ed.), *Center/Periphery in representations and institutions. Proceedings from the 3rd Annual Meeting and Congress of The International Semiotics Institute, Imatra, Finland, July 16-21, 1990* (pp. 211-256). Imatra: International Semiotics Institute.

Sonesson, G. (2003a). The Symbolic Species revisited. Considerations on the semiotic turn in cognitive science and biology. *Working Papers of the Project "Language, Gesture, and Pictures from the Point of View of Semiotic Development", Lund University, Lund*, SGBWP3.

Sonesson, G. (2003b). Why the mirror is a sign – and why the television picture is no mirror. Two episodes in the critique of the iconicity critique. *S. European Journal for Semiotic Studies, 15*(2-4), 217-232.

Sonesson, G. (in press a). From iconicity to pictorality. A view from ecological semiotics. *VISIO*, thematic issue: *Iconicity revisited/Iconicité revisité.*

Sonesson, G. (in press b). Prologomena to a general theory of iconicity. Considerations on language, gesture, and pictures. *Logos and Language*, thematic issue: *Naturalness and Iconicity in Linguistics.*

Spencer, J.P., Clearfield, M., Corbetta, D., Ulrich, B., Buchanan, P. & Schöner, G. (2006). Moving toward a grand theory of development: in memory of Esther Thelen. *Child Development, 77*(6), 1521-1538.

Spetch, M.L. & Friedman, A. (2006). Pigeons see correspondence between objects and their pictures. *Psychological Science, 17*(11), 966-972.

Spetch, M.L., Kelly, D.M. & Reid, S. (2000). Recognition of objects and spatial relations in pictures across changes in viewpoint. In J. Fagot (Ed.), *Picture Perception in Animals* (pp. 107-141). Hove: Psychology Press Ltd.

Spielman, K.S. (1976). Development of the perception and production of line forms. *Child Development, 47*, 787-793.

Spinozzi, G. (1993). Development of spontaneous classificatory behavior in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 107*(2), 193-200.

Spinozzi, G. (1996). Categorization in monkeys and chimpanzees. *Behavioural Brain Research, 74*, 17-24.

Striano, T., Tomasello, M. & Rochat, P. (2001). Social and object support for early symbolic play. *Developmental Science 4*(4), 442-455.

Suarez, S.D. & Gallup, G.G., Jr. (1981). Self-recognition in chimpanzees and orangutans, but not gorillas. *Journal of Human Evolution, 10*, 175-188.

Suddendorf, T. (2003). Early representational insight: twenty-four-month-olds can use a photo to find an object in the world. *Child Development, 74*(3), 896-904.

Sugihara, T., Edelman, S. & Tanaka, K. (1998). Representation of objective similarity among three-dimensional shapes in the monkey. *Bilogical Cybernetics, 78*, 1-7.

Sutton, P.J. & Rose, D.H. (1998). The role of strategic visual attention in children's drawing development. *Journal of Experimental Child Psychology, 68*, 87-107.

Swaisgood, R.R. & Shepherdson, D.J. (2005). Scientific approaches to enrichment and stereotypies in zoo animals: what's been done and where should we go next? *Zoo Biology, 24*, 499-518.

Swartz, K.B. (1982). Species discrimination in infant pigtail macaques with pictorial stimuli. *Developmental Psychobiology, 16*(3), 219-231.

Swartz, K.B., Evans, S. (1994). Social and cognitive factors in chimpanzee and gorilla mirror behavior and self-recognition. In S.T. Parker, R.W. Mitchell & M.L. Boccia (Eds.), *Self-awareness in Animals and Humans. Developmental Perspectives* (pp. 189-206). New York: Cambridge University Press.

Swartz, K.B., Sarauw, D. & Evans, S. (1999). Comparative aspects of mirror self-recognition in great apes. In S.T. Parker, R.W. Mitchell & H.L. Miles (Eds.), *The Mentalities of Gorillas and Orangutans. Comparative Perspectives* (pp.283-294). Cambridge: Cambridge University Press.

Taglialatela, J.P., Savage-Rumbaugh, S. & Baker, L.A. (2003). Vocal production by a language-competent *Pan paniscus*. *International Journal of Primatology, 24*(1), 1-17.

Tanaka, M. (1995). Object sorting in chimpanzees (*Pan troglodytes*): classification based on physical identity, complementarity, and familiarity. *Journal of Comparative Psychology, 109*(2), 151-161.

Tanaka, M. (1996). Information integration about object-object relationships by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 110*(4), 323-335.

Tanaka, M. (1997). Formation of categories based on functions in a chimpanzee (*Pan troglodytes*). *Japanese Psychological Research, 39*(3), 212-225.

Tanaka, M. (2001). Discrimination and categorization of photographs of natural objects by chimpanzees (*Pan troglodytes*). *Animal Cognition, 4*, 201-211.

Tanaka, M. (2003). Visual preference by chimpanzees (*Pan troglodytes*) for photos of primates measured by a free choice-order task: implication for influence of social experience. *Primates, 44*, 157-165.

Tanaka, M. (2006). Spontaneous categorization of natural objects in chimpanzees. In T. Matsuzawa, M. Tomonaga & M. Tanaka (Eds.), *Cognitive Development in Chimpanzees* (pp. 340-367). Tokyo: Springer.

Tanaka, M. (2007a). Development of the visual preference of chimpanzees (*Pan troglodytes*) for photographs of primates: effect of social experience. *Primates, 48*(4), 303-309.

Tanaka, M. (2007b). Recognition of pictorial representations by chimpanzees (*Pan troglodytes*). *Animal Cognition, 10*, 169-179.

Tanaka, M., Tomonaga, M. & Matsuzawa, T. (2003). Finger drawing by infant chimpanzees (*Pan troglodytes*). *Animal Cognition, 6*, 245-251.

Tanner, J.E. & Byrne, R.W. (1996). Representation of action through iconic gesture in a captive lowland gorilla. *Current Anthropology, 37*(1), 162-173.

Temerlin, M.K. (1976). *Lucy: Growing Up Human*. London: Souvenir Press.

Terrace, H.S. (1980). *Nim*. London: Eyre Methuen.

Terrace, H.S. (2005). The simultaneous chain: a new approach to serial learning. *Trends in Cognitive Sciences, 9*(4), 202-210.

Thomas, G.V., Jolley, R.P., Robinson, E.J. & Champion, H. (1999). Realist errors in children's responses to pictures and words as representations. *Journal of Experimental Child Psychology, 74*, 1-20.

Thomas, R.K. (1996). Investigating cognitive abilities in animals: unrealized potential. *Cognitive Brain Research, 3*, 157-166.

Thomson, J. (1885). *Through Masai Land.* London: Sampson, Low, Marston, Searle and Rivington.

Thompson, R.K.R. & Contie, C.L. (1994). Further reflections on mirror usage by pigeons: lessons from Whinnie-the-Pooh and Pinocchio too. In S.T. Parker, R.W. Mitchell & M.L. Boccia (Eds.), *Self-awareness in Animals and Humans. Developmental Perspectives* (pp. 392-409). New York: Cambridge University Press.

Thompson, R.K.R. & Oden, D.L. (1995). A profound disparity revisited: perception and judgement of abstract identity relations by chimpanzees, human infants, and monkeys. *Behavioural Processes, 35*, 149-161.

Thompson, R.K.R. & Oden, D.L. (2000). Categorical perception and conceptual judgements by nonhuman primates: the paleological monkey and the analogical ape. *Cognitive Science, 24*(3), 363-396.

Thompson, R.K.R., Oden, D.L. & Boysen, S.T. (1997). Language-naïve chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processes, 23*(1), 31-43.

Tobach, E., Skolnick, A.J., Klein, I. & Greenberg, G. (1997). Viewing of self and nonself images in a group of captive orangutans (*Pongo pygmaeus abellii*). *Perceptual and Motor Skills, 84*(2), 355-370.

Tolan, J.C., Rogers, C.M. & Malone, D.R. (1981). Cross-modal matching in monkeys: altered visual cues and delay. *Neuropsychologia, 19*(2), 289-300.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, USA: Harvard University Press.

Tomasello, M. & Call, J. (1997). *Primate Cognition*. New York: Oxford University Press.

Tomasello, M. & Call, J. (2004). The role of humans in the cognitive development of apes revisited. *Animal Cognition, 7*, 213-215.

Tomasello, M., Call, J. & Gluckman, A. (1997). Comprehension of novel communicative signs by apes and human children. *Child Development, 68*(6), 1067-1080.

Tomasello, M., Striano, T. & Rochat, P. (1999). Do young children use objects as symbols? *British Journal of Developmental Psychology, 17*, 563-584.

Tomonaga, M. (1994). How laboratory-raised Japanese monkeys (*Macaca fuscata*) perceive rotated photographs of monkeys: evidence for an inversion effect in face perception. *Primates, 35*(2), 155-165.

Tomonaga, M. (1999). Inversion effect in perception of human faces in a chimpanzee (*Pan troglodytes*). *Primates, 40*(3), 417-438.

Tomonaga, M. (2007). Visual search for orientation of faces by a chimpanzee (*Pan troglodytes*): face-specific upright superiority and the role of facial configural properties. *Primates, 48*, 1-12.

Trillmich, F. (1976). Learning experiments on individual recognition in budgerigars (*Melopsittacus undulatus*). *Zeitschrift für Tierpsychologie, 41*, 372-395.

Troseth, G.L. (2003). Getting a clear picture: young children's understanding of a televised image. *Developmental Science, 6*(3), 247-253.

Troseth, G.L. & DeLoache, J.S. (1998). The medium can obscure the message: young children's understanding of video. *Child Development, 69*(4), 950-965.

Troseth, G.L., Saylor, M.M. & Archer, A.H. (2006). Young children's use of video as a source of socially relevant information. *Child Development, 77*(3), 786-799.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327-352.

Tyrrell, D.J., Anderson, J.T., Clubb, M. & Bradbury, A. (1987). Infant recognition of the correspondence between photographs and caricatures of human faces. *Bulletin of the Psychonomic Society, 25*(1), 41-43.

Uttal, D., Schreiber, J.C. & DeLoache, J.S. (1995). Waiting to use a symbol: the effects of delay on children's use of models. *Child Developments, 66*, 1875-1889.

Van der Stigchel, S., Meeter, M. & Theeuwes, J. (2006). Eye movement trajectories and what they tell us. *Neuroscience and Biobehavioral Reviews, 30*, 666-676.

Vauclair, J., Rollins, H.A., Jr. & Nadler, R.D. (1983). Reproductive memory for diagonal and nondiagonal patterns in chimpanzees. *Behavioural Processes, 8*(3), 289-300.

Vermeire, B.A. & Hamilton, C.R. (1998). Inversion effect for faces in split-brain monkeys. *Neuropsychologia, 36*(10), 1003-1014.

Vokey, J.R., Rendall, D., Tangen, J.M., Parr, L.A. & de Waal, F.B.M. (2004). Visual kin recognition and family resemblance in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology, 118*(2), 194-199.

Vonk, J. (2002). Can orangutans and gorillas acquire concepts for social relationships? *International Journal of Comparative Psychology, 15*, 257-277.

Vonk, J. (2003). Gorilla (*Gorilla gorilla gorilla*) and orangutan (*Pongo abelii*) understanding of first- and second-order relations. *Animal Cognition, 6*, 77-86.

Vonk, J. & MacDonald, S.E. (2002). Natural concepts in juvenile gorilla (*Gorilla gorilla gorilla*) at three levels of abstraction. *Journal of the Experimental Analysis of Behavior, 78*(3), 315-332.

Vonk, J. & MacDonald, S.E. (2004). Levels of abstraction in orangutan (*Pongo abelii*) categorization. *Journal of Comparative Psychology, 118*(1), 3-13.

Vonk, J. & Povinelli, D.J. (2006). Similarity and difference in the conceptual systems of primates: the unobservability hypothesis. In E.A. Wasserman & T.R. Zentall (Eds.), *Comparative Cognition: Experimental Explorations of Animal Intelligence* (pp. 363-387). New York: Oxford University Press.

Walraven, V.; van Elsacker, Linda & Verheyen, R. (1995). Reactions of a group of pygmy chimpanzees (*Pan paniscus*) to their mirror-images: evidence of self-recognition. *Primates, 36*(1), 145-150.

Ware, E.A., Uttal, D.H., Wetter, E.K. & DeLoache, J.S. (2006). Young children make scale errors when playing with dolls. *Developmental Science, 9*(1), 40-45.

Washburn, D.A. (1993). The stimulus movement effect: allocation of attention or artifact? *Journal of Experimental Psychology: Animal Behavior Processes, 19*(4), 380-390.

Washburn, D.A., Gulledge, J.P. & Rumbaugh, D.M. (1997). The heuristic and motivational value of video reinforcement. *Learning and Motivation, 28*, 510-520.

Washburn, D.A., Hopkins, W.D. & Rumbaugh, D.M. (1989). Video-task assessment of learning and memory in macaques (*Macaca mulatta*): effects of stimulus movement on performance. *Journal of Experimental Psychology: Animal Behavior Processes, 15*(4), 393-400.

Washburn, D.A. & Putney, R.T. (2001). Attention and task difficulty: when is performance facilitated? *Learning and Motivation, 32*, 36-47.

Wasserman, E.A., Fagot, J. & Young, M.E. (2001). Same-different conceptualization by baboons (*Papio papio*): the role of entropy. *Journal of Comparative Psychology, 115*(1), 42-52.

Wasserman, E.A., Gagliardi, J.L., Cook, B.R., Kirkpatrick-Steger, K., Astley, S.L. & Biederman, I. (1996). The pigeon's recognition of drawings of depth-rotated stimuli. *Journal of Experimental Psychology: Animal Behavior Processes, 22*(2), 205-221.

Wasserman, E.A. & Miller, R.R. (1997). What's elementary about associative learning? *Annual Review of Psychology, 48*, 573-607.

Watanabe, S. (1993). Object-picture equivalence in the pigeon: an anlysis with natural concept and pseudoconcept discriminations. *Behavioural Processes, 30*, 225-232.

Watanabe, S. (1997). Visual discrimination of real objects and pictures in pigeons. *Animal Learning and Behavior, 25*(2), 185-192.

Watanabe, S. (2001). Discrimination of cartoons and photographs in pigeons: effects of scrambling of elements. *Behavioural Processes, 53*, 3-9.

Whiten, A., Horner, V., Litchfield, C.A. & Marshall-Pescini, S. (2004). How do apes ape? *Learning and Behavior, 32*(1), 36-52.

Whiten, A., Goodall, J., McGrew, W.C., Nishida, T., Reynolds, V., Sugiyama, Y., Tutin, C.E.G., Wrangham, R.W. & Boesch, C. (2001). Charting cultural variation in chimpanzees. *Behaviour, 138*, 1481-1516.

Wilkie, D.M. (2000). Use of pictures to investigate aspects of pigeons' spatial cognition. In J. Fagot (Ed.), *Picture Perception in Animals* (pp. 91-106). Hove: Psychology Press Ltd.

Williams, S.L., Brakke, K.E. & Savage-Rumbaugh, E.S. (1997). Comprehension skills of language-competent and nonlanguage-competent apes. *Language and Communication, 17*(4), 301-317.

Winner, E. & Ettlinger, G. (1979). Do chimpanzees recognize photographs as representations of objects? *Neuropsychologia, 17*, 413-420.

Witmer, L. (1909). A monkey with a mind. *Psychological Clinic, 3*, 179-205.

Wrangham, R. & Peterson, D. (1997). *Demonic Males. Apes and the Origins of Human Violence*. London: Bloomsbury Publishing.

Yerkes, R.M. (1916a). Provision for the study of monkeys and apes. *Science, 43*, 231-234.

Yerkes, R.M. (1916b). The mental life of monkeys and apes: a study of ideational behavior. *Behavior Monographs, 3*, 1-145.

Yerkes, R.M. (1926). *Människan närmast ("Almost Human", 1925).* Stockholm: Hugo Gebers förlag.

Yerkes, R.M. (1927a). The mind of a gorilla. *Genetic Psychology Monographs, 2*(1-2), 1-193.

Yerkes, R.M. (1927b). The mind of a gorilla: part 2. Mental development. *Genetic Psychology Monographs, 2*(6), 377-551.

Yerkes, R.M. (1929). The mind of a gorilla: part 3. Memory. *Comparative Psychology Monographs, 5*(2), 1-92.

Yerkes, R.M. (1943/1945). *Chimpanzees. A Laboratory Colony*. New Haven: Yale University Press.

Yerkes, R.M. (1951). Gorilla census and study. *Journal of Mammalogy, 32*(4), 429-436.

Yerkes, R.M. & Nissen, H.W. (1939). Pre-linguistic sign behavior in chimpanzee. *Science, 89*, 585-587.

Yerkes, R.M. & Petrunkevitch, A. (1925). Studies of chimpanzee vision by Ladygin-Kohts. *Journal of Comparative Psychology, 5*(1), 99-108.

Yerkes, R.M. & Sykes Child, M. (1927). Anthropoid behavior. *The Quarterly Review of Biology, 2*(1), 37-57.

Yerkes, R.M. & Yerkes, A.W. (1929/1953). *The Great Apes. A Study of Anthropoid Life*. New Haven: Yale University Press.

Yoshikubo, S. (1985). Species discrimination and concept formation by rhesus monkeys (*Macaca mulatta*). *Primates, 26*(3), 285-299.

Zaitchik, D. (1990). When representations conflict with reality: the preschooler's problem with false beliefs and "false" photographs. *Cognition, 35*, 41-68.

Zimmerman, R.R. & Hochberg, J. (1970). Responses of infant monkeys to pictorial representations of a learned visual discrimination. *Psychonomical Science, 18*(5), 307-308.

Zlatev, J. (2007). Embodiment, language and mimesis. In T. Ziemke, J. Zlatev & R. Frank (Eds.), *Body, Language and Mind, Vol 1: Embodiment.* Berlin: Mouton.

Zlatev, J., Persson, T. & Gärdenfors, P. (2005). Bodily mimesis as "the missing link" in human cognitive evolution. *Lund University Cognitive Studies, 121.*