TESTS OF TRUE PICTORIAL COMPETENCE IN CHIMPANZEES: THE CASE OF DRAWINGS

TOMAS PERSSON

Abstract. For us human beings, interpreting pictures is as much an act of imagination as one of face-value recognition. Pictures are therefore a promising avenue for studying mental worlds. Can this be extended to the study of animal minds? True pictorial competence is here defined as a referential and differentiated apprehension of pictures, where reference is crucially dependent on visual resemblance. It is proposed that pictures can be appreciated in at least two other modes than the pictorial one, surface and reality modes, which have to be precluded in experimental investigations of true pictorial competence. One way to avoid reality mode is to use non-photographic stimuli. The few experiments on chimpanzees (*Pan troglodytes*) that have taken this approach are critically reviewed in this chapter.

Introduction

In science we have tried to understand ourselves by looking at the minds of nonhuman primates for well over a century. Today comparative cognitive studies from an evolutionary perspective usually take one of two forms. The first one can be called "cognitive cladistics." Cladistics is the reconstruction of ancestral relations by grouping e.g. species according to shared features, which is contrasted to groups that do not share the same features. In this 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 strives to find out which species can, and which cannot, perform a certain task in a certain way. The focus on contrasting averaged groups restricts this approach to being descriptive on a most general level and these studies seldom target cognitive *processes*.

Another way of doing comparative cognitive research is therefore not to focus on the typical, but on the *potential* of the ape mind. This can be studied on the level of individuals rather than species. A focus on typicality would for example not answer the question whether nonhumans can learn a human language in the positive. Typicality is interesting for many reasons, especially in combination with in-depth study of individual variation, but a single language-using chimpanzee can teach us more about language and language evolution than a hundred "typical" chimpanzees. A language-using chimpanzee can tell us that language is possible for a nonhuman and it provides us with necessary resources for understanding why this is so. Furthermore, it can help us guess whether these factors were active also in human prehistory.

For the above reasons comparative cognitive research on something as "artificial" and "human" as pictorial competence is not misdirected. Apes are in fact prime subjects because there seem to be individuals that can, and individuals that cannot, decode pictures as the typical human does (see Persson 2008). In this chapter a framework for studying pictorial competence in nonhumans will be presented together with a critical review of central studies in the field.

Pictures and animals

Pictures have been used as stimuli since the early days of animal cognition research, both as substitute for real-life objects, individuals and events, and as abstract stimuli. Pictures have been used to assess questions of spatial representation, social cognition, viewpoint consistency, and serial list learning (Fagot, Martin-Malivel and Dépy 2000), as well as in all sorts of discrimination and categorization tasks (Bovet and Vauclair 2000). Pictorial stimuli seem to have been used with such success that few scientists have asked the questions of why it works, and what it means. Since animals of all kinds readily accept pictures as examples of real-world objects, it must mean that pictures are simple and intuitive phenomena. This observation easily lends itself to the idea that there is only one way of viewing pictures: you either recognize them or you do not. This destiny was apparent already in the first reports on chimpanzee mentality:

> 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 (Köhler 1925/1957: 278).

The above quote is the closing paragraph of Köhler's classic *The Mentalities of Apes*, published in 1925. Köhler was a Gestalt psychologist who conducted cognitive experiments and observations on chimpanzees in the 1910s, and he is considered one of the founding fathers of cognitive primatology. He chose to look at chimpanzees' performance with pictures after having observed their reactions to stuffed toy animals, cardboard face masks, and mirrors. He observed that it seemed necessary for the toy animals to have some likeness to real animals, "nearness to life," in order to invoke a response, i.e. fear. The stuffed animals invoked even stronger responses than did most real animals. Köhler concluded that the stuffed animals, not being fully real, played on the imagination in a way that real animals did not. Uncertainty as opposed to experience was the key factor.

Köhler wondered whether the chimpanzees' ability to recognize 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 were much interested in the pictures of themselves and other chimpanzees, but only one showed suggestive signs that he recognized their content: he made social responses towards it.

Köhler then made a photograph of an empty crate and another photograph of a crate crammed full of bananas and pasted these on two boxes baited with fruit. The star pupil of the previous test chose the box with the banana picture on 10 successive trials. By now Köhler wanted to control for rote learning and created two new photographs: one of bananas and one of a rock. 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 of pictures.

When testing a second chimpanzee on the old pictures, 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 rock and the bananas, she made hardly any mistakes. Köhler ascribed these mixed results solely to the quality of the pictures. "I made no further tests […]" he reports (Köhler 1925/1957: 278).

Köhler's brief investigation of picture perception in chimpanzees gives the impression that he assumed that the correspondence between a picture and the real world could only be of one kind. His conclusion was that the higher the similarity between picture and referent, the easier it is for a chimpanzee to solve an object choice task guided by picture information.

But why do we not see the same correlation in adult human pictorial competence? A distorted photograph is not necessarily more difficult to recognize than one that depicts reality more truthfully. We readily decode non-photographs that are far removed from the real perceptual world, like cartoons. We can even intend to see likeness where there is not supposed to be any, like faces in the clouds. Are Köhler's chimpanzees and humans different points on one single competence continuum, with differences merely being the result of different experience of picture qualities? It will be argued below that it is not.

Three ways of looking at pictures

Following previous work (Persson 2008), I propose that there are at least three ways, or modes, in which pictures can be related to the world

(1) The first mode involves bypassing any estimation of what the picture might actually depict. What is perceived is rather the patterns, shapes, or colours on the surface of the picture, and it stays at that. Here this will be termed a *surface* mode 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 through generalization based on invariant features, one can judge correspondences while circumventing recognition on a categorical level. That is, one can e.g. sort pictures of apples on a level (e.g. colour) that does not involve realizing that it is apples that one sorts.

(2) Pictures can also be meaningful due to their likeness to the real world, but 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 the typical example, the effect is not necessarily limited to stimuli that seem realistic from a human perspective. Critical features can likewise elicit reality-guided responses. In *reality-based* picture processing an object is not seen as being anywhere else but in the picture, albeit perhaps in a stranger than usual form. In this mode one can solve tasks that depend on categorization, but it is not qualitatively different from categorizing real instances of the depicted objects. If an object in a picture is e.g. matched with a

similar object outside of the picture, it is an object relating to an object, not a picture of an object relating to an object.

(3.) In the third mode pictures become meaningful through likeness (iconicity) to the real world (and other pictures), but are still sufficiently differentiated from it in order not to be confused with it. Reference lies 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. In pictures where the likeness is directly perceived in virtue of mirroring the real-life experiences of the perceiver, an expectation of separation between picture and reality is crucial. 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. This type of picture use is distinguished from other referential instances by being called seeing pictures as *pictures*.

This three-part division is similar to that of Fagot et al. (2000) who propose independence, confusion, and equivalence modes of picture processing in a review of animal picture experiments. This is not surprising since "[t]he proposed classification into three modes of processing is intuitively obvious" (Fagot et al.: 297). Indeed, Premack (1976) makes a similar distinction between forms of picture competences in animals, as does Cabe (1980) indirectly. It also seems to follow naturally from the theory of (pictorial) semiotics of Sonesson (e.g. 1989, 2007).

The modes can be seen as standing in a hierarchical relation to each other, but they should not be equated with a developmental trajectory. The modes are not general competences but depend on interactions 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.

Only the third of the three modes entails understanding pictures as pictures, i.e. a referential understanding of pictures. This type of processing is interesting for many reasons. For example, only when pictures are viewed in a referential way does an interpreter have a reason to assume that the picture is informative about things that are not in the picture itself, such as movement, where an object is hidden in the real world, and that the orange thing in Donald Duck's face must be a beak. "Filling in" the content of a pictorial display is something akin to imagination. One reason to expect such competence in nonhumans at all is data on other iconic competences that has received more attention, such as the understanding of model replicas (e.g. Premack and Premack 1983; Kuhlmeier, Boysen and Mukobi 1999) and mirrors (e.g. Gallup 1970; Menzel, Savage-Rumbaugh and Lawson 1985). The three modes of pictorial competence are applicable to all iconic media.

When applying the above definitions on Köhler's picture experiments it is clear that what he tested was a reality mode competence. The social responses and the performance with pictures judged high versus low in realism points to that. The chimpanzees saw the photographs as objects in themselves. Had they seen the photographs as *views* of objects they would not so easily have been confused by a slightly distorted photograph.

Not making the distinction between different forms of competence with pictures is a mistake, especially when making claims that pertain to the pictorial mode. Many studies have indeed involved pictures (Bovet and Vauclair 2000; Cabe, 1980; Fagot et al. 2000), especially photographs, and several studies give the impression of referential use of pictures. But on close scrutiny, processing of pictures in a non-differentiated manner, i.e. in reality mode, or in surface modes, can often not be excluded (Persson 2008). If we are interested in true pictorial competence we must look for experiments designed specifically for testing pictorial mode competence at the expense of reality and surface modes. Drawings, for example, can be used in experiments that normally use photographs, but can be designed to exclude reality-based responses. In addition, drawings often require an a priori expectation of a motif for recognition to occur. Such an expectation comes with the pictorial mode. In the remainder of this chapter I will review all studies that have involved recognition of drawings by chimpanzees.

Recognising drawings

When picture-naïve humans describe pictures, slow and stepwise recognition sometimes takes place. According to Deregowski (1976) such recognition can go something like: "that is a tail, this is a foot, that is a leg joint, those are horns… it is a waterbuck". A parsimonious explanation for this process is one described by Gregory (1973, in Deregowski 1976) whereby the perception of a picture occurs in a series of "hypotheses." (In experienced viewers the process is often too quick for introspection.) A set of properties in the picture is the basis for a hypothesis which is then verified against further properties. If necessary, the hypothesis might be modified and retested against the features until a stable identification has settled. Parts and wholes thereby define each other with continuous feedback.

Some pictures might give only a small “whole” to start off the process, such as an eye or some other feature with high saliency from everyday life. Other times a larger but ill-defined whole can catch the eye. For example, one of the earlier recognitions that started the process of identifying the waterbuck might not have been a local feature, 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.

Successive approximation is the reason that we can perceive 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 influence the perception of the whole. Without this constructive process pictures that are not possible to interpret in a reality mode would 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: 59). It is true also for objects in the world at large. At each momentary glance only parts of objects are informative to our brains. Identifying an object is a question of using attention electively to complete the picture. This means that eye and head movements are not random, but are dependent on the viewer’s “perceptual purpose” (Hochberg 1980). If we expect to see an array of lines and colors we will not see a waterbuck. What one needs 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.

Let us now turn to those experiments that have specifically addressed recognition of drawings in chimpanzees.

Viki

The chimpanzee Viki was raised in the home of psychologist Keith Hayes and Catherine Hayes with the purpose of studying an infant chimpanzee growing up in a human social and material environment. Since Viki was raised as a human child she had early experience of pictures, but was never specifically trained to perceive them. She enjoyed picture books from the age of 6 months, but not until the age of 1.5 years did she spontaneously respond to pictures differentially (Hayes 1951).

When Viki was 4 years old she readily pointed to pictures of beverages and led the addressed person to the refrigerator. This indicates that she could relate pictures at least in some ways to the real world without confusing the two. But we need to know more about the context and generalisability of this behaviour before we can say that Viki knows that picture motifs are picture motifs and objects are objects. Viki could have discovered that she could barter those flat and flimsy *special cases* of drinks for more drinkable versions. If her performance was limited to certain categories, i.e. beverages, there is further reason to suspect that she had learned specific connections between pictures and their objects, rather than discovering the general nature of pictures and their communicative potential. It is possible that she used reality mode processing of pictures in a communicative context. This might in addition be limited to bartering situations.

When she was 3.5 years old Viki's ability to imitate bodily actions from pictures was tested. Hayes and Hayes (1953) report that she did fairly well with stimuli ranging from movies, via black-and-white photographs, to "simple line drawings." Her successful interpretation of line drawings speaks for a pictorial competence, but there is no information on the novelty of the pictures and rote learning can therefore not be ruled out in the present analysis. Besides the performance with line drawings it is noteworthy that imitation of dynamic actions depicted in static pictures requires abductions on behalf of the interpreter (Sonesson 1989). 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 her responses we cannot know if Viki 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 and Nissen 1971). 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 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. Viki was rewarded for choosing from the same category throughout a session. Importantly, a given picture was only used once, precluding rote learning. The sessions were divided into four successive groups of picture types. In the first sessions "naturalistic" colour pictures were used, and in the following a mix of realistic photographs and stylized drawings. The third group consisted of realistic color as well as black-and-white photographs, and the last sessions involved only black-and-white line drawings. The pictures were a mix of magazine illustrations, photographs, and handmade line-drawings. Many of the drawings were freehand copies of photographs from earlier sessions, and can thus be criticized for lacking in novelty. Viki is reported to have been able to generalize discriminations to novel pictures of all the types mentioned above. Success rate on the four groups of problems were 85%, 75%, 82%, and 73% correct respectively. No data is given for the distribution within the second group (mixed), but the lower figures in this and the fourth group (line drawings) might be due to the abstractness of the stimuli. In the third group Viki was 95% correct on the color photographs but only 68% on the black-and-white ones. Picture "realism" thus seems to have had an effect on Viki's performance, but importantly she performed above chance on drawings. Note, however, that these stimuli were often traced from photographs. Furthermore, discrimination tasks are prone to responses based on invariant local features, bypassing recognition on a categorical level.

Following the discrimination work Viki was given matching tasks. She was shown an object and was made to choose from two pictures the one that depicted an object similar to the one that was shown. This experiment was also divided into four conditions. In the first one Viki received trials with realistic photographs reused from the discrimination tasks. She was correct on 95% of these trials. In the second condition she received picture pairs from the former group, but now the matching picture had become the non-match and vice versa, and new sample objects were used. She was correct on 85% of these trials. The third group of trials utilized line drawings, some of which were reused from the discrimination tasks and were thus not novel. Her performance dropped somewhat to a still good 80%. The last group consisted of rearrangements of the line-drawing pairs from the previous group. Viki was 91% correct. This figure might not have been as high if novel line drawings had been used.

Viki was also fond of drawing, but never seemed to make depicting pictures. However, she learned to connect multiple dots and twice she reacted to the shapes as if she recognized them. 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," and she spontaneously named a similarly drawn cup (one of the few words she could voice) and fetched one (Hayes and Nissen 1971). To other self-made drawings she was indifferent and, importantly, never tried to fetch an object through guesswork. The dog and the cup were thus not chance events. However, these observations lack control for contextual cueing. Self-made pictures make performance through reality mode unlikely, but the numerous occasions where Viki did not fetch objects in response to the dot connecting exercise amounts to arguments against her understanding the depicting potential of self-made drawings.

Cross-modal matching

By the 1970s the question regarding recognition of pictures by apes still seemed unresolved. Davenport and Rogers (1971) note with reference to the studies of Köhler (1925/1957) and Hayes and Hayes (1953) 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) criticize both studies for lacking controls for associative learning.

In cross-modal matching e.g. a visual sample has to be matched to haptic comparison stimuli, or a haptic sample is to be matched to visual comparisons. The subject typically puts its hand behind an occluder and feels an object that cannot be seen. It then has to match what it feels to visually accessible objects or pictures of objects. The matching can be simultaneous or delayed.

Davenport and Rogers (1971) report a cross-modal experiment with life-sized colour and black-and-white photographs of unfamiliar objects as visual stimuli. The subjects performed above chance and there was no difference between the two photograph conditions. In this study the subjects were naïve to pictures, and since the pictures were highly realistic and placed behind glass, a reality mode picture processing is the given candidate for the apes' performance. Placement behind glass potentially reduces the two-dimensionality of the pictures and retains an illusion suitable for reality mode. The glass also served as response surface. A study by Winner and Ettlinger (1979) failed to replicate these findings, and this may be because their pictures were not instrumentally connected to either locus of response or to the food rewards (cf. Persson 2008), but perhaps also because the subjects could experience the two-dimensionality of the pictures by being allowed to investigate them directly.

Davenport, Rogers and Russell (1975) introduced non-photographs in their paradigm. Five classes of pictures of unfamiliar objects were tested in a simultaneous matching condition, given in the order 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. In addition, colour photographs were used in a delayed version of the task. 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 one-trial problems, completely novel stimuli or novel combinations of stimuli were given for each picture condition. However, Davenport et al. (1975) seem to have neglected to control for order effects. At the time the subjects received the line drawing trials they had already had extensive experience with the previous types of pictures. The line-drawings might have been perceived as special cases of silhouette drawings, which in turn can be argued to be recognizable as shapes in a reality mode.

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. In the delayed matching condition with colour photographs, four out of five subjects performed well, but in the critical tests with novel stimuli only two performed above chance.

The fact that the testing was made in the context of cross-modal matching might very well have been a significant factor for the picture performance shown. Shape matching rather than identity matching might have taken place through comparison of the remembered sample shapes to the pictured shapes, which probably remained intact in all picture conditions. If the animals had hit upon this strategy in the silhouette condition transfer to line drawings might have been a simple task. The fact that most sample objects in the study were unknown to the subjects strengthens the advantage of matching based on shape similarity rather than object identity. That the pictures were behind glass further enhanced their aspects as shapes as opposed to pigment on two-dimensional surfaces. The poor performance with novel stimuli in the delayed condition suggests that the subjects could not reliably identify the objects and therefore had problems remembering them. If, instead, the task had been to match identifiable objects, memory demands might have been smaller. At the same time pictorial performance might have declined since the pictures would have to be interpreted with their identity taken into account. Line drawings might then fall short. The experiments by Davenport and colleagues therefore most likely involved combinations of picture processing in the surface and reality modes.

The Kyoto Studies

Itakura (1994) tested black and white drawings on the female chimpanzee Ai at the Primate Research Institute in Kyoto, Japan. Ai was 12 years old at the time and versed in the use of the abstract visual symbols called lexigrams. Itakura (1994) does not report what previous experience Ai might have had with non-photographic pictures, but at least she had extensive training on discriminating figure from ground in drawing-like stimuli since e.g. her lexigrams are all black-and-white patterns (Matsuzawa 2003).

Testing stimuli included pictures of three humans, three chimpanzees, and one orangutan that Ai could recognize in movies and photographs and name with letters from the alphabet. Colour photographs and black-and-white line drawings were used as representations of the individuals and were to be named with their respective letter. (Two extra distracter letters were included.) All line drawings had been constructed from photographic templates and were quite detailed.

The presentation and apparatus used was computerized. In each trial, one of the seven individuals was the target and after the sample had been presented there was a two second delay until the letters came onscreen. A food item was dispensed each time Ai made a correct choice among the nine letters. It seems from the published descriptions that the matching was simultaneous, i.e. the sample picture remained in view during the response. If an incorrect choice was made, a signal sounded and the trial restarted after a brief timeout. Ai was 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 consisted 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 7 all seven stimuli-individuals were represented by two photographs and one drawing. Photographs were changed for each session but the report does not say whether the drawings were also changed or remained the same throughout testing. That would have meant that in the 7th 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 were high for most drawings.

Ai was correct on 100% of the photographs but not above chance on the first trial on any of the line drawings. She tended to choose the letter that had been reinforced in the previous trial, and responded randomly during correction trials. However, 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 good.

In a second phase of the study each stimulus-individual was represented by two photographs and one line drawing from the onset. Again 84 presentations (not counting correction trials) were given per session and novel line drawings and photographs were continuously introduced. But still, each picture was used as target four times within a session. Ai was much better in this phase 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, suggesting learning rather than recognition. 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 by a human control group.)

That the orangutan sticks out as stimulus is not surprising. It is not implausible that the orangutan drawings show some invariant features that the other drawings lack. However, such an analysis would have to be made from 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 wore glasses. The third person also wears glasses, but is a woman. Combining the glasses feature with e.g. 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 even without attributing “glasses” and “hair” to them, and thus solve the problem completely in a surface mode. (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 Ai’s errors one finds that it was more common for her to confuse individuals within the chimpanzee category than between species. 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 patterns are recognizable as classes, i.e. “this is one of the human-patterns”? This latter alternative might explain the hardships of keeping individual patterns apart while still succeeding in keeping the species 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 them. This can all be done without actually seeing "humanness" or "chimpanzeeness" in the pictures themselves.

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 also of the drawings, but could not decode the transformations in most of them. However, we cannot be sure of what the names actually encode for Ai, especially after having subjected her to stimuli that unreliably match to the names, i.e. the drawings. One way to ease her task would be by matching between drawings and photographs, doing away with naming altogether. The matching could also be made the other way around, so that she is given the sample name or photograph, and then require her to pick the matching drawing from an array of pictures. In this experimental format she could compare pictures with each other and pinpoint the drawing that is *most* like the referent. In this way there is no need to invoke an absolute resemblance, only a relative one. Although this would be a case of scaffolded recognition, it is still an iconic one, as well as referential and differentiated. For now we must unfortunately conclude that Ai's performance leaves us with inconclusive evidence.

A study that directly targets the questions raised in this chapter was recently published. Tanaka (2007) tested Ai, three other adult females, and three 4 to 5 year old juvenile chimpanzees at the Kyoto Primate Research Institute on a computerized generalization task involving photographs and drawings of various degree of realism. The subjects were required to choose three 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 and generalization to novel colour photographs was above chance for all but one subject. However, the subjects sometimes chose non-flower pictures that contained colourful patches. Whether they mistook these for flowers or did not choose in accordance with a flower concept is unknown.

Next step was generalization to realistic colour sketches and less realistic color cartoons, or computer clip-art. The criteria for "realism" are not given so we have to assume that it was the subjective judgment of 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 readily choose correctly above chance in both categories at first exposure. Chloe, another adult, 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 one subject, Chloe, showed a clear decline with decrease in "realism." In contrast to the variable performance of the adults all three juveniles performed above chance from the very first trial on both types of pictures. The juveniles, together with Ai, also displayed strong learning effects with repeated exposures.

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. 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. It is difficult to say whether Ai suddenly grasped, by the third exposure, that also the line-drawings were flowers, or whether she had formed a new category perhaps not remotely connected to flowers. She could also just have memorized the correct pictures, which were only 12. However, when tested with novel Kanji (Chinese) characters 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.

Tanaka (2007) would be a more convincing case if possible generalization from photographs to drawings had been precluded, perhaps by testing the abstracted pictures before photographs. If transfer would occur in this direction the claim for recognition of drawings as flowers would be stronger. Also, repeated response to flower-like stimuli, regardless of level of abstraction, will not do as an indication of pictorial competence. Flowers are unfortunately a type of stimuli that take highly canonical forms in most cartoons and clipart. An analysis of those particular flower pictures that proved problematic for the subjects would be required in order to see if this was indeed the case. In order to make a study like this more convincing a requirement would be a conditional discrimination, as in a matching-to-sample paradigm, where flowers are sometimes correct and sometimes wrong, and where several categories of drawn objects are tested at the same time. Of course, critical trials with novel pictures are a necessity, and correct first trial performance needs to be demonstrated.

Conclusions

This chapter has explored the experimental work that has so far been conducted with chimpanzees and drawings. It has been suggested that drawings, in virtue of being less realistic than e.g. photographs, are a promising avenue for studying true pictorial competence in animals. True pictorial competence has been defined as a referential and differentiated apprehension of pictures, where reference is crucially dependent on visual resemblance. A necessity for demonstrating true pictorial competence is to control for associative learning of picture-object relations, as well as performance in “surface” and “reality” modes. The studies reviewed do not convince in this regard, unfortunately.

Studying the interpretation of pictures when it takes place in a pictorial mode, in turn, supplies a tool for studying categorization and imagination in non-linguistic nonhumans. This has the potential not only to help us understand the minds of other species, but also to cast light on our own cognitive prehistory. Although the picture is a fairly recent cultural creation the pictorial mind might not be, and might not even be restricted to human brains.

Of further interest are the common denominators that foster pictorial potential in minds. It has been suggested that a true pictorial competence is part of a more general “symbolic function” or a “sign function” (cf. Sonesson 2007) that cuts across domains such as play, imagination, and of course verbal and gestural communication. Again, turning towards nonhumans seems a worthwhile endeavour. Comparing the pictorial abilities of apes that have a background in language projects to that of apes with a similar upbringing but without language-training, could reveal possible connections between pictures and language (see Persson 2008). Perhaps language-competent apes can decode pictures that language-naïve apes cannot, or perhaps there is no difference. Contrary to Köhler (1925/1957: 278) the suggestion would therefore be that “anyone who may take the trouble to experiment on other chimpanzees” has a lot left to investigate. Some steps in this direction were taken in my dissertation (Persson 2008), but a lot remains to be done. The framework of different modes of apprehending pictorial stimuli seems to be a necessary prerequisite for such research.

References

Cabe, P.A. 1980. Picture perception in nonhuman subjects. In M.A. Hagen (Ed.), *The Perception of pictures, Vol. 2*, 305-343. New York: Academic Press, Inc.

Bovet, D. and J. Vauclair. 2000. Picture recognition in animals and humans. *Behavioural Brain Research* 109(2): 143-165.

Davenport, R.K. and C.M. Rogers. 1971. Perception of photographs by apes. *Behaviour* 39(2): 318-320.

Davenport, R.K., C.M. Rogers and I.S. Russell. 1975. Cross-modal perception in apes: altered visual cues and delay. *Neuropsychologia* 13: 229-235.

Deregowski, J.B. 1976. On seeing a picture for the first time. *Leonardo* 9: 19-23.

Fagot, J., J. Martin-Malivel and D. Dépy. 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*, 295-320. Hove: Psychology Press Ltd.

Gallup, G.G., Jr. 1970. Chimpanzees: self-recognition. *Science* 167: 86-87.

Gregory, R.L. 1973. The confounded eye. In R.L. Gregory and E.H. Gombrich (Eds.), *Illusion in nature and art*. London: Duckworth.

Hayes, C. 1951. *The Ape in Our House*. New York: Harper and Brothers.

Hayes, K.J. and C. Hayes. 1953. Picture perception in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology* 46: 470-474.

Hayes, K.J and C.H. Nissen. 1971. Higher mental functions of a home-raised chimpanzee. In A.M. Schrier and F. Stollnitz (Eds.), *Behavior of Nonhuman Primates, Vol. 4*, 59-115. New York and London: Academic Press.

Hochberg, J. 1980. Pictorial functions and perceptual structures. In M.A. Hagen (Ed.), *The perception of pictures, Vol. 2*, 47-93. New York: Academic Press, Inc.

Itakura, S. 1994. Recognition of line-drawing representations by a chimpanzee (Pan troglodytes). *The Journal of General Psychology* 121(3): 189-197.

Kuhlmeier, V.A., S.T. Boysen and K.L. Mukobi. 1999. Scale-model comprehension by chimpanzees (Pan troglodytes). *Journal of Comparative Psychology* 113(4): 396-402.

Köhler, W. 1925/1957. *The mentality of apes*. London and Tonbridge: Penguin Books.

Matsuzawa, T. 2003. The Ai project: historical and ecological contexts. *Animal Cognition* 6: 199-211.
Menzel, E.W., Jr., E.S. Savage-Rumbaugh and J. Lawson 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.
Persson, T. 2008. *Pictorial primates: A search for iconic abilities in great apes*. Lund: Lund University Cognitive Studies, 136.
Premack, D. 1976. *Intelligence in Ape and Man*. New Jersey: Lawrence Erlbaum Associates, Inc.
Premack, A.J. and D. Premack. 1972. Teaching language to an ape. *Scientific American* 227(4): 92-99.
Premack, D. and A.J. Premack. 1983. *The mind of an ape*. New York: W.W. Norton and Company, Inc.
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. 2007. From the meaning of embodiment to the embodiment of meaning: A study in phenomenological semiotics. In T. Ziemke, J. Zlatev and R. Frank (Eds.) *Body, language and mind. Vol 1: Embodiment*, 242-283. Berlin: Mouton de Gruyter.
Tanaka, M. 2007. Recognition of pictorial representations by chimpanzees (Pan troglodytes). *Animal Cognition* 10: 169-179.
Winner, E. and G. Ettlinger. 1979. Do chimpanzees recognize photographs as representations of objects? *Neuropsychologia* 17: 413-420.