



universität  
wien

# MASTERARBEIT

Titel der Masterarbeit

Picture-object recognition in pigeons: The use of visual features for classifying pictures of an unfamiliar natural category

Verfasserin ODER Verfasser

Martin Kysel

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2011

Studienkennzahl lt. Studienblatt:

A 066 013

Studienrichtung lt. Studienblatt:

MEI:CogSci

Betreuerin / Betreuer:

Dr. Ulrike Aust



# Contents

<b>Abstract</b>	<b>6</b>
<b>1 Introduction</b>	<b>9</b>
1.1 Theoretical Background . . . . .	9
1.1.1 Perception of images . . . . .	9
1.1.2 Avian perceptual system and stimulus presentation . . . . .	10
1.1.3 Object Recognition . . . . .	10
1.1.4 Categorization . . . . .	12
1.1.5 Picture-Object-Recognition (POR) . . . . .	14
1.2 Previous Studies . . . . .	16
1.2.1 Picture-to-picture transfer . . . . .	16
1.2.2 Bidirectional transfer . . . . .	17
1.2.3 Adapted behavior . . . . .	17
1.2.4 Cross-modality tests . . . . .	18
1.2.5 Recognition of humans in pictures, the Complementary Informa- tion Procedure (Aust & Huber) . . . . .	18
1.2.6 Influence of category familiarity . . . . .	19
1.3 Aim of the Study . . . . .	22
<b>2 Methods</b>	<b>25</b>
2.1 Subjects and Housing . . . . .	25

2.2	Apparatus . . . . .	27
2.3	Stimuli and Procedure . . . . .	28
2.3.1	General . . . . .	28
2.3.2	Training . . . . .	31
2.3.3	Generalization Test . . . . .	31
2.3.4	Picture-Object Recognition Test (POR) . . . . .	31
2.3.5	Head vs Shell vs Skin Test (HSS) . . . . .	32
2.3.6	Grayscale Test . . . . .	34
2.3.7	Blur Test . . . . .	34
2.3.8	Scramble Test . . . . .	34
2.3.9	Test Ammonites . . . . .	35
2.3.10	Test Spirals . . . . .	35
2.4	Data Analysis . . . . .	37
<b>3</b>	<b>Results</b>	<b>39</b>
3.1	Training . . . . .	39
3.2	Generalization Test . . . . .	42
3.3	Picture-Object Recognition Test . . . . .	43
3.4	Head vs Shell vs Skin Test (HSS) . . . . .	47
3.5	GrayScale Test . . . . .	50
3.6	Blur Test . . . . .	52
3.7	Scramble Test . . . . .	53
3.8	Ammonites and Spirals Tests . . . . .	53
<b>4</b>	<b>Discussion</b>	<b>57</b>
	<b>Acknowledgements</b>	<b>65</b>
	<b>References</b>	<b>65</b>

<b>A</b>	<b>Appendix</b>	<b>73</b>
A.1	Declaration . . . . .	73
A.2	Stimuli . . . . .	74
A.3	Summary . . . . .	75
A.4	Zusammenfassung . . . . .	77
	Cirriculum Vitae . . . . .	79

# Abstract

For over fifty years, researchers have been using pictures depicting natural scenes as stimuli in experimental studies. By now, we know that the avian perceptual system differs from that of humans. Differences in depth perception, flicker, pentachromatic basic colors, or an extension into the UV spectrum make the perception of pictorial stimuli different. The potential ability of animals to get beyond mere perceptual feature matching and actually create a link between the object in the image and the real, depicted object is called *Picture-Object Recognition* (POR) at the level of representational insight. Much simpler and cognitively less demanding mechanisms such as learning by rote, perceptual feature learning, or reliance on similarity to the training set could be considered as alternatives to true picture-object equivalence (i.e., insight). A recent study by Aust and Huber [1, 2] introduced the *Complementary Information Procedure* (CIP) to test POR in pigeons and widely exclude the influence of other, simpler mechanisms. The CIP tests animals with pictures of different parts of an object than presented during training. Namely, pigeons familiar with humans responded significantly more to missing, previously unseen parts of a human figure compared to arbitrary skin patches, but only if they had real life experience with humans. These results, ruling out the possibility of transfer being based merely on simple perceptual 2D-features, suggest the presence of a certain associative link between the image and its content. To further test this assumption, Grabner [3] conducted a study with pictures of snails, i.e., a category that the subjects were assumed to be unfamiliar with. The

data indeed indicated a lack of transfer to unseen parts and thus absence of POR in the absence of missing real life experience. However, the results were not always clearcut, which was partly due to small sample sizes. Also, it was not evident how and to what extent individual perceptual features controlled responding. Thus, as concluded by the author, additional experiments were needed.

I started the present study to extend Grabner’s results and possibly find a more comprehensive explanation for them, thereby supporting the notion of category familiarity as a crucial factor in POR. To this end, I trained pigeons to discriminate pictures of incomplete snails from pictures of non-snails and tested their ability to recognize complementary parts of representatives of such an unfamiliar concept. The POR Test showed no difference in the treatment of Unseen Parts and arbitrary Skin Patches. I therefore concluded that real life experience with a category is crucial for the subject’s ability to recognize the missing parts. Thereby, the present results strengthen those obtained by Grabner and are further support of representational insight in pigeons.

In the second part of the thesis, subjects were presented with images of varied informational content to assess the role of different perceptual features. Namely, the individual parts of a snail yielded different features that could be used for categorization. Especially Shells turned out to be very salient and their shape seemed to be the primary categorization cue. On the contrary, heads and arbitrary skin patches were significantly less important. Tests with scrambled stimuli showed that this kind of modification caused the most drastic decreases in response rates. Greyscaling came second, and blurred images seemed to be equivalent with original unmodified depictions. Artificial snail-like stimuli presented during the tests Spirals and Ammonites obviously failed to sufficiently mimic the snail category and were treated rather as negatives (i.e., non-snails). Even a previously unseen and unfamiliar concept, such as a snail is defined by category-specific perceptual features that can be integrated into a polymorphous response rule. Every single manipulation of the picture, be it frag-

mentation or a reduction of dimensionality, caused drops in peck rates. Nevertheless, pigeons that had never seen a real snail before could well discriminate snails from spirals and ammonites (despite that lacking real-life experience). This indeed suggests the abstraction of a set of class-defining features that were specific enough to allow for distinguishing between snails and members of related categories with similar features.



# Chapter 1

## Introduction

### 1.1 Theoretical Background

#### 1.1.1 Perception of images

For more than forty years, since the publication of the influential paper by Herrnstein and Loveland [4], researchers have been focussing on the ability of animals to perceive the object presented in an image. As I will elaborate on further, one of the main questions has thereby remained the same. Are human and avian vision similar? Do humans and birds view objects in pictures similarly [5]? We assume that the last common ancestor of humans and birds lived some 300 million years ago. It is therefore be plausible to expect that the neural mechanisms underlying the visual perception of humans and birds will differ. Sometimes we simply assume that birds perceive the world as we do. For example, Barraclough and colleagues [6] claimed that 69% of tested passerine birds are monomorphic in appearance. This would have an impact on the actual consequences of sexual dimorphism on both natural and sexual selection. Later in 2005 Eaton [7] showed that due to the extension of the perceptual field into the UV spectrum, less then 27% of the birds are actually monomorphic in the eye of the birds. The assumption that birds perceive images the same way as we do can thus

lead to an anthropocentric error.

### **1.1.2 Avian perceptual system and stimulus presentation**

Images usually used in experiments possess no olfactory cues; they are not three dimensional and even do not fit the perceptual color system [8, 9, 10] of pigeons, thus are static. CRT, LCD and LED monitors all use a three color system adapted to the human visual system. We know that birds are in general tetrachromats, but the vision of some is even pentachromatic. Their color perception is distributed differently and extended into the UV spectrum [11]. What we perceive as meaningful images, could, for pigeons only be a distorted chromaticity matrix. This is a problem that cannot be solved with current technology. Color perception is only one of the above-mentioned problems. Moving images (videos) are more closely related to its biological counterpart and therefore more ecologically valid than static ones. Videos, however, may cause even more problems (e.g., regarding temporal resolution and depth perception) than they actually solve [12, 13, 14]. Videos are not the main topic of this thesis, so I will elaborate on them no further. Seeing the problems that the presentation of images in pigeon experiments causes, one has to ask whether all of these have been solved for human subjects. We know that due to the difference in color perception between pigeons and humans, pigeons could have problems recognizing objects in images. According to this notion, is the use of images in humans always appropriate? Unexpectedly there has been negative evidence in past years [15, 16, 17]. For example, the often cited work by Herskovits [18] reported of a woman who was unable to recognize her son in a black-and-white photograph.

### **1.1.3 Object Recognition**

Recognition of objects is a natural and every-day problem that has to be solved by many animal species. Humans do this with apparently little or no effort. It is the higher

level of cognitive processing that enables us to recognize objects as being the same when seen from many different angles and distances, and/or under different lighting conditions.

Two particular theories of object recognition are of importance for this work. Due to Cerella [19], pigeons operate accordingly to the Particulate Feature Theory. Similarly to a perceptron (an artificial neural network based on bayesian statistics ), the brain of a pigeon identifies only small discontinued parts of an object. Attending only to the local features and specific patterns present in the object, the pigeon should ignore global information about the spatial relationship between these parts. Cerella found support for this claim in 1980 by showing that pigeons were insensitive to global scrambling of a drawing of Charlie Brown.

A year later, a contradictory theory was posed by Biederman [20]. At the core of his Recognition-by-Components Theory are *geons*, simple geometrical shapes such as cubes, spheres, cylinders or wedges. Representations of objects become spatial maps constructed of a few geons. If confronted with an object, pigeons should, based on local recognition of local geons, extract the spatial relationships among geons and if a match of this spatial relationship map with a known object occurs, recognize the object. In order to test the hypothesis, the author presented pigeons with simple line drawings of kettles, irons, desk lamps and sail boats. When some parts of the object were removed or dislocated, the recognition ratio did not change significantly. According to Cerella a removal of a part should result in a drop of recognition ratio, which was not the case.

Experiments by Cook [21] have contributed to the hypothesis that pigeons prefer local stimuli over global composition. One of the possible explanations put forward in that study is that the local information in the pigeons' memory is available faster. Contrasting results by Shimp [22] showed a more human-like preference for global features. All in all, it seems that pigeons are able to use both elemental (local) and configural (global) features and switch between these two. Support for this has been found by

Fremouw [23, 24]. The presentation and type of stimuli used obviously influences feature preference. We know, for example, that the presentation of simple line drawings as used by Cerella differs from naturalistic images [25, 26] in matters of feature selection.

#### 1.1.4 Categorization

I would like to point out the difference between Object Recognition and Object Categorization. As explained above Object Recognition is based on recognition of features that contribute to the identification of a spotted item. Experiments researching Object Recognition may, e.g., show stimuli under different view angles, modify the overall shape, or color information, or may scramble the image. Object Recognition is generally based on the importance of concrete features used to identify one single object, whereas Object Categorization operates on sets of objects. Two objects can look alike, have a similar function, share a reinforcement history or share any other set of common features. In order to speak of ecological relevance, objects have to be classified, for example, as edible/inedible, friend/foe or as human/nonhuman. The first series of experiments testing Object Categorization in pigeons by Herrnstein [4, 27] inspected the ability to discriminate between images containing or lacking trees, humans and bodies of water. Subjects readily learned these discriminations and transferred them to new previously unseen members of the learned categories.

Natural Categories are "open-ended" as there is no limitation in the number of potential members. Categorization can be performed on the basis of simple perceptual cues, perceptual similarity, function, or an abstract relation among category members [28]. Category boundaries are fuzzy and membership is determined by means of polymorphous rules. The term *polymorphous* refers to the fact, that there are no both required and satisfactory single cues to define category membership [29]. This is, categorization is not based on one single attribute that reliable determines the result. Instead, various different features contribute to class membership. The more class-

defining attributes included in the polymorphous rule an object possesses, the more likely it will be regarded as a category member. For example, a typical instance of the category tree (used by Herrnstein) is green, has leaves and a trunk, is wooden and has branches. Other less typical members can lack some of these features: a tree in autumn is not green and conifers have no leaves like deciduous trees. In turn, there are also objects (e.g., flowers, wooden branchy bushes or simple wooden man-made artifacts) that share many of the category-defining features of the tree category, yet are not members. There is thus a degree of membership and not a member/not-member division. Prototype effects in pigeons were tested by Jitsumori in [30].

The features relevant for categorization are usually denoted positive compared to irrelevant features present in both, reinforced and non-reinforced images. In time subjects learn to focus attention to positive features and ignore weak category predictors [31]. Natural images vary in a large set of dimensions making the identification of relevant features rather problematic. One of the approaches to control for the amount and type of features is the Synthetic Approach. In [32] Huber and Lenz constructed human faces defined by their: distance between the eyes, height of the forehead, length of the nose and chin size. In order to categorize correctly, pigeons had to focus on all attributes at once. Because the dimensions were controlled, their relative contribution to category membership could be assessed. The disadvantage of the synthetic approach is, however, that respective tasks may not accurately reflect the structure and featural content of natural categories and thus, conclusions on pigeons' classification strategies in real life situations may be possible only to a limited extent.

Categories are unique and intrinsic to the perceiving agent (human, robot or animal). Among others, places or group members can belong to different categories across individuals, and different species may differ in where they draw the dividing lines between categories. For example, the category snail, used in this work, could in principle be broader than that of humans, including, e.g., also small bugs, or slimy surfaces.

Or the other way round, there could be many types of snail sub-categories based on experience, location, smell, taste, color or toxicity. All of this has to be borne in mind, if we try to imply from correct categorization, that there is Picture-Object-Recognition present (see below).

### 1.1.5 Picture-Object-Recognition (POR)

Pictures are always abstractions of an item present in the real world. They lack a lot of important dimensions that are of importance (as explained in Section 1.1.2) and animals are nonetheless able to categorize them correctly. What categorization per se does not imply is the understanding of the relationship between images and their referents. Correct performance can be achieved by simple 2D-feature recognition. Reported by Amato [33] cebus monkey readily learned to discriminate between images that contained humans and those that did not. Importantly they also classified images containing small pink-redish patches (open water melons, a flamingo) as members of the human category. Similar issues were reported by Herrnstein [4]. Picture-Object-Equivalence is, however, the ability to go beyond mere perceptual basis and categorize based on insight.

There are multiple levels of POR. According to Fagot [34] there are three distinct types of picture perception.

1. *The independence mode.* Pictures and their depicted real world counterparts are considered distinct entities. The processing of one is independent of the other and there is no transfer. Any relationships learned on pictures are not transferred to the real world. For example, an animal may learn to identify images of snakes eating its conspecifics and differentiate it from snakes eating eggs. However the subject would not acquire the fear of real snakes as it fails to understand the relationship between snakes in images and real snakes.
2. *The confusion mode.* Images are confused and interchanged with their real coun-

terparts. There is no difference perceived between 2D-pictures and the corresponding real object. For example a monkey may grab for a picture of a banana and try to eat it.

3. *The equivalence mode.* At this level animals understand the relationship between objects and their images. They are aware of the fact, that images are only representations and not the objects themselves. Facts newly learned from one type can thus be transferred to the other one. DeLoache defined this ability as the understanding of animals that pictures "stand for something other than themselves" [35]. Methods that support the presence of this mode in animals will be discussed later in this section.

Alternatively or in addition, we may distinguish between different types (or levels) of picture-object recognition apart from true representational insight. Thus, these types (level 1 and 2) do not imply that the animals understand the representational content of the image (level 3).

1. *Perceptual Level.* This level of understanding attributes the least cognitive abilities to pigeons (or animals in general). Based solely on simple visual 2D-features, objects are sorted into categories without understanding their correspondence with real 3D-objects. It can be one or many features defining the category, yet still the animal is unable to recognize the represented object. The transfer to novel stimuli can also be based on these perceptual features or on similarity to the learning set. Perceptual similarity can thereby lead to assigning, by mistake, also non-members to the category in questions (such as a red flamenco or a watermelon to the human category).
2. *Associative Level.* Individual parts of real objects become associated, and these relationships can be transferred from the real world representatives to images and vice versa. The link between objects and their pictures is, however, not

understood in the strictest sense.

3. *Abstract Level.* The representational content of the image is understood and a link to its real-life representative is created. This is, the picture is recognized as a representation of the real object, i.e., representational insight is gained. Therefore, this mechanism clearly goes beyond the mere recognition of simple 2D-features and also beyond the mere association of individual parts. This level matches the level of picture-object equivalence as put forward by Fagot.

Whenever one claims having found evidence of representational insight (or picture-object equivalence) all lower levels of POR have to be ruled out. Pigeons possess the cognitive ability to learn large sets of picture-response pairs by rote [36, 37]. Up to 1.600 images randomly assigned to groups can be learned without any obvious categorization cues. Alternatively, there is plenty of evidence for them being able to abstract perceptual rules [4, 15, 38, 26, 39]. Correct categorization can thus lead to false belief in picture-object equivalence. This is, transfer from familiar to novel stimuli or between objects and images may not be based on the animal’s recognition of the correspondence between objects and pictures of them. Furthermore, it has to be made sure that the animal can differentiate between the image and the depicted object before true equivalence can be inferred.

## 1.2 Previous Studies

### 1.2.1 Picture-to-picture transfer

The pioneering studies by Herrnstein and Loveland [4] showed that pigeons and monkeys are able to generalize beyond the scope of mere exemplar learning. The tested subjects were able to generalize to novel pictures of trees, humans or bodies of water. The pictures used in the generalization tests were never seen by the animals before,



but the possibility of learning by simple perceptual cues could not be ruled out.

### **1.2.2 Bidirectional transfer**

One of the approaches to investigate the link between objects and their pictures is to train animals on pictures and assess their behavior on the transfer to real objects. The same experimental design can be done the other way round. The experiments by Delius [39] and Watanabe [40] yielded good transfer with stimuli that had similar shapes in both presentation types. Similarly, the pigeons of Cabe [41] performed well when confronted with no or little variation in viewing angle. It therefore seems that animals tested with this paradigm used some kind of perceptual or associative cues rather than truly abstract processing that would have involved the creation of 3D-representations.

### **1.2.3 Adapted behavior**

Animals are shown images or videos and their behavioral responses towards these stimuli are recorded. Either new behavioral reactions are learned or the direct reaction is assessed. Recognition of conspecifics in videos [42], acquired fear reactions [43] or predator-prey behavior towards video images [44] are some examples from the 90's. This approach clearly identifies the ability of the tested animals to extract the relevant conceptual information from the 2D representation even though the presentation of the stimuli is perceptually distorted compared to the real object. Yet animals still show the tendency to treat pictures like the real represented things. In the experiments of Bovet and Vauclair [15] baboons successfully categorized food and non-food items. They also successfully transferred this categorization to images of food/non-food items but did not understand the difference between images and real objects. Baboons tried to put depicted bananas into their mouths or peel pictures of seeds. This error in cognition can be explained with the confusion mode of picture processing.

### 1.2.4 Cross-modality tests

This class of experiments tests the transfer between different perceptual modalities. For example, primates tested by Winner and Ettlinger [45] were allowed to explore the categorized object by touch. Afterwards they had to match the explored object to its pictorial representation. The animals failed the visual/haptic matching and the authors concluded that "the photographs were meaningless 2D objects." A successful cross-modality transfer was shown by Davenport [46].

### 1.2.5 Recognition of humans in pictures, the Complementary Information Procedure (Aust & Huber)

In order to exclude the possibility of transfer being merely based on perceptual 2D-feature matching, Aust and Huber [1] introduced a study design in which the same set of visual features never occurred in the training and in the test. The Complementary Information Procedure (CIP) is a modification of the picture-to-picture matching scenario. Objects, in this case humans, are separated into logical physiological parts. The subjects, ten pigeons were separated into two distinct groups. Both groups were trained on humans as the positive class (S+) and on images showing something else as the negative class (S-). The first group, NoHead was trained on images showing humans without heads. This part of the human body was either digitally deleted or simply removed by smart image cropping. Hands, torso and legs remained intact and were shown as part of the training stimuli. In the generalization test the animals were only shown the complementary part, the head. They had never seen pictures of heads before. The same procedure was applied in an analogous way to the second, the NoHands group.

The part shown in the training set was denoted the Seen Part (SP) and the complementary part the Unseen Part (UP). Both groups of animals responded to the UP with

above average peck rates. To show that this was not due to the textual or color information related to skin which was present in both hands and heads a third stimuli class Skin Patches (SK) was constructed. The SK consisted of arbitrarily shaped patches of human skin that were similar in size and shape to the true, representative, parts. Importantly, animals did not respond emit high response rates to this class of stimuli. Namely, they were significantly lower, than those emitted to the complementary parts, the UP. The authors concluded that the subjects must have used a higher-level abstract matching of the different parts of the same concept. This is, they displayed POR at least at the associative level.

If the logic behind the CIP is correct (and the method is valid), then animals should fail to show transfer to UPs with members of categories they had never experienced before in real life. The study by Aust and Huber [2] introduced visually restricted pigeons that had no pre-experience with heads of live humans. Thus, they should not be able to match the complementary parts as they were unfamiliar with heads and their spatial relations with the remainder of the human body. To them, the heads shown as UP should be unfamiliar entities. Indeed, the experiment showed that there was no preference of unfamiliar UP over arbitrarily shaped skin patches (SP). Hence, there was a difference between pigeons with and without real life experience with humans (or at least parts of them). The results contributed to the conclusion made by the authors in their previous work that pigeons are able to extract the abstract content from images and therefore understand the correspondence between objects and their pictures.

### **1.2.6 Influence of category familiarity**

The studies by Aust and Huber [2] showed that category familiarity is a precondition for successful POR (which is actually plausible). Similar results were obtained by Watanabe [47] who trained pigeons to discriminate between pictures of familiar objects

(i.e., feeder) and unfamiliar objects (i.e., mug). Only animals trained with familiar objects as positive were able to generalize to unusual viewpoints. Animals that were trained with unfamiliar objects failed. The author of this work concludes a certain level of picture-object recognition that enabled the recognition of familiar objects. A comparable effect of similarity was reported by Wilkie, Willson and Kardal [48] who trained pigeons with pictures of familiar/unfamiliar places. The "familiar" group was significantly faster than the "unfamiliar" group.

In line with the experiments by Aust and Huber, Grabner [3] applied the CIP to an unfamiliar category for her pigeons. If there was indeed an effect of familiarity in the human study [2], then there should also be no significant difference between Unseen Parts (UP) and Skin Patches (SK), if pigeons are tested with yet another unfamiliar category. Her decision was to use a natural category whose members could be divided into visually distinct parts (such as a head) and bore a certain degree of similarity among each other, namely snails.

Grabner divided her animals into two experimental groups, NoHead and NoShell, and trained them on complementary parts of snails (Heads or Shells). Two birds (one from each group) passed the generalization test that involved the presentation of novel instances of the learned task and could be subjected to a subsequent Picture-Object Recognition test (POR).

One of the animals (Josef) showed no significant difference between Unseen Parts (UP) and arbitrary skin patches (SK). This animal contributes to the hypothesis by Aust and Huber, that there is a difference between animals trained on familiar and unfamiliar categories. The other animal (Daisy) pecked significantly more at UP compared to SK. Such a result would indicate that the subject was able to recognize an unseen part of a member of an unfamiliar category and therefore point to a flaw in the study design. The CIP decreases the amount of perceptual cues that are shared by both training and test images. However, the results of Daisy may show that another

type of transfer can cause abstract-like categorization. No explanatory conclusion has yet been made about the nature of the features relevant for categorization or about the outlying results of this subject. The response rates of both animals showed a high degree of variation and the absolute peck rates were very low. Therefore, the results must be interpreted with caution.

At the end of her thesis Grabner proposed a study on snails versus non-snails that became the basis of this work.

## 1.3 Aim of the Study

This study contains two main tests and several additional tests. One of the main goals was to extend the Picture Object Recognition Test by Grabner. For this I decided to take two of her animals that would have passed the Generalization Test if the criterion had been less conservative. I retrained them and performed the POR test. Both this study, and the study of Grabner worked with unfamiliar categories and therefore were in a contrasting position to the original work by Aust & Huber. Should more of the tested animals respond indifferently to the Unseen Parts and Skin Patches, one could assume a familiarity effect. By contrast, a flaw in the study design would have to be considered should all of the tested animals respond similarly to Daisy. As explained above, the CIP assumes that the animals are in possession of a certain abstract concept based on experience with real life representatives of the tested (familiar) category (humans). There should, however, no such concept be formed in the case of unfamiliar objects, and thus no transfer to the UP should happen. This test was expected to strengthen the case of POR at a level beyond the recognition of simple 2D-features by demonstrating the need for real life experience.

Any category representation present on the subject's mind is purely intrinsic to the animal. We may only assume that the object is divisible into parts that are of biological relevance to or at least perceptually salient for the subject. The aim of the second test was to extend our limited knowledge about the relative importance of the individual parts. Actually, the contribution to and thus the importance for object categorization can differ among parts. The test that examines these relative roles of different parts of a snail used in the POR test was called the Head Vs Shell Vs Skin Test (HSS Test). Due to the significantly longer training duration and lower training success of the No Shell animals by Grabner, I expected that animals would be generally more attentive to the visual patterns of the Shell rather than to the color and texture of the skin. As explained above, two of her No Shell animals completely failed to learn the initial

discrimination task. The No Head animals that were trained on Shells were faster and showed better categorization results compared to the opposite group. I therefore assumed that shells would more readily be categorized as members of the snail category than bodies, heads and arbitrary skin patches.

In three additional tests (Grayscale, Blur, and Scrambled) the informational content of the snail-pictures was systematically manipulated. All three kinds of images were variations of both positive Full Snail images and negative images showing something else. The more the mean standardized peck rate would drop compared to the original image, the more important the tested feature would be. Grayscale images removed the color content from the stimuli and performance on them could therefore be compared to that on the colored originals in order to assess the relative importance of color. Blurring the images distorted the local information and was thereby a test for the importance of texture. Scrambling spatially rearranged parts of the image and therefore reduced global information content while retaining the surface information (like color and texture) and local details. The shape of bigger objects present in the stimuli, such as the shell of a snail would be destroyed. If this type of image manipulation caused the strongest drop in classification performance, I would assume that the shell is the most important part that the animals attended to. The term scrambling was used for different types of image manipulation. A clear distinction between the approaches was made by Matsukawa and colleagues ([49], page 26). "... dividing the original pictures into small squares that were then distributed arbitrarily over the whole picture area will be referred to as *fragmentation*." This approach was also used by Aust & Huber [50]. The other method as used in earlier studies by Cerella [19], Watanabe [51] or Aust & Huber [25] scrambles only some components of an image (e.g., humans or, cartoon figures, but not the background). The term scrambling as used throughout this thesis refers to the former definition.

Based on the results of these tests I devised additional artificial snail-like stimuli.

The test Spirals added simple black and white spirals to negative previously unseen images. If categorization was based purely on this simple shape cue, spirals should also be classified as members of the snail category. Otherwise, they should be classified as negative images, and we would see that either the snail category has more membership criteria in addition to a simple spiral-feature (which would then be insufficient to determine membership), or that the latter is generally unimportant. The other test included fossilized ammonites that, to the human eye, look very snail-like. I wanted to test if this is the case for pigeons as well. Other than the Spirals test, this test did not investigate only one feature, but varied along many different dimensions. Ammonites are similar to and still different from snails by means of shape, texture and color. Therefore it is of importance to see, whether deviation in the perceptual space is small enough to allow for generalization (i.e., for categorization as snails). It could be, for example, that the pigeons would classify not according to a snail/non-snail rule but rather according to an animal/non-animal rule. In principle the class snails needs not be restricted to snails only, but it might, for example, include several types of small animals (such as bugs), or all animals with shells. For humans fossilized ammonites are very similar to snails and could belong to the same shelled category. Here, I tested whether shelled snails and shelled fossils belonged to the same category for pigeons.



# Chapter 2

## Methods

### 2.1 Subjects and Housing

Six pigeons (*Columbia livia*), obtained from local breeder or hatched in our aviaries, were used as subjects. Five of them were homing pigeons (Anatol, Hannes, Niki, Olivia, Rhonda). One of them was of a local Austrian breed, called "Strasser" (Lukas). Furthermore, two homing pigeons that had already been trained by Grabner were tested (Cordula and Franz). All birds were maintained at approximately 90% of their free-feeding weight. Water and grit were freely available in the aviary but food was only available during the experimental sessions. Subjects were housed together with several con-specifics in 3 outdoor aviaries (approximate size 3 m x 1m x 3 m ) with 10-14 birds each.

At the beginning of these experiments the birds were familiar with the Go/No-go procedure and the apparatus. They were naive to the actual task even though they had already participated in other visual categorization experiments.

Subjects were divided into three experimental groups. Groups NoHead and NoShell both contained 4 animals that had already been trained and tested by Grabner in [3], but for reasons of comparison and clarity their results are also included in the present report. In addition, I increased the number of birds in the Picture-Object Recognition

Test (POR) by testing two of Grabner’s pigeons that she had abandoned when they failed to perform at criterial level in the Generalization Test. The chosen criterion of mastery was however quite conservative (i.e.: significance in an U-test on the 1%level), and both birds would have passed at a more tolerant criterion of 5%. For higher result stability I allowed the birds that failed with the strict criterion (Cordula and Franz) to enter the POR-test after some re-training and repetition of the Generalization Test. The group assignments: of Grabner’s birds (Groups NoHead, NoShell) and of the birds that were newly trained for this thesis (Group FullSnails) as well as the number of animals in each group can be found in Table 2.1.

Table 2.1: Experimental Groups and the tests they were subjected to (X)

Group	Nb of Ss	Tests				
		Gen	POR	HSS	GS,BI,Sc	Amm,Sp
<b>NoHead</b>	3	X	X			
<b>NoShell</b>	1	X	X			
<b>FullSnails</b>	6	X		X	X	X
<i>ni NoHead</i>	1	X				
<i>ni NoShell</i>	3	X				

*Note.* ni = not included in this work, animals that did not pass the Generalization Test by [3] (and would not even have passed according to the more lenient criterion). Ss = subjects. Gen = Generalization test, POR = Picture-Object recognition test, HSS = Head vs Shell vs Skin test, GS = Grayscale test, BI = Blur test, Sc = Scramble test, Amm = Ammonites Test, Sp = Spirals test

## 2.2 Apparatus

The experiments were performed in wooden experimental indoor chambers (Skinner boxes). Subjects could enter the apparatus one by one from their respective outdoor aviaries through a connecting tunnel. The interior size of the chamber was 39 x 40 x 53 cm. Stimuli were presented on a 15.0 XGA Color TFT-LCD monitor placed behind an infrared touch-frame. Directly below the touchscreen there was the aperture of the automated grain feeder ( $\varnothing 6\text{cm}$ ). Each Skinner Box was controlled by an adjacent PC and its input/output processing application CognitionLabLight ( M. Steurer ). The schematics can be found in Figure 2.1

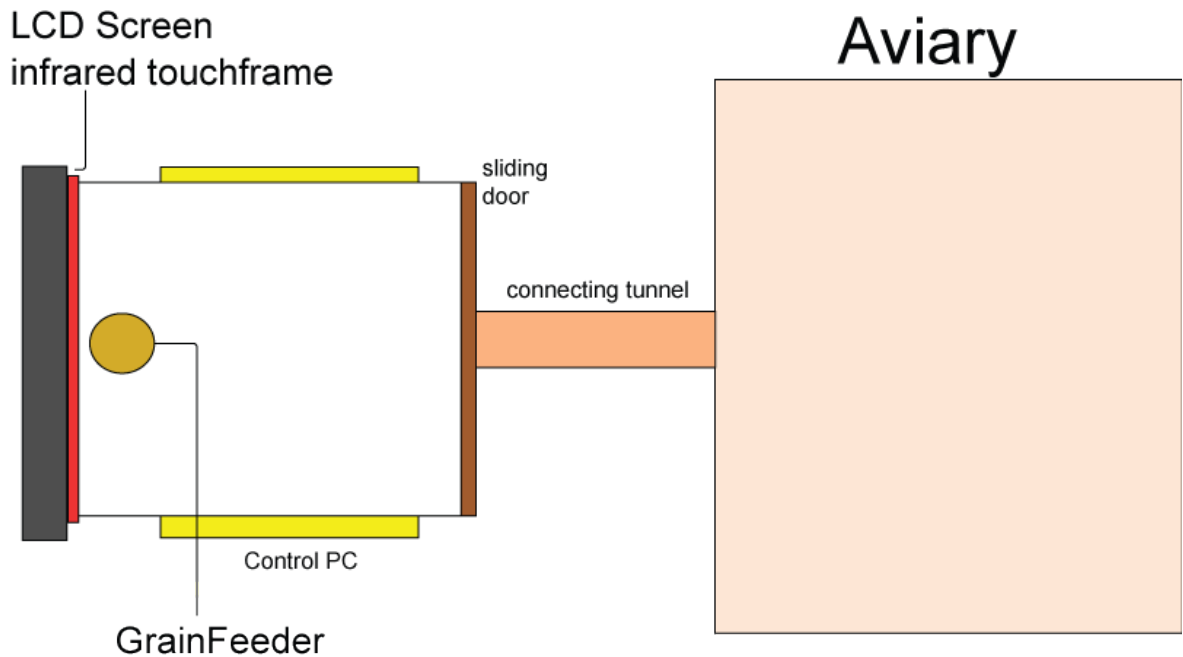


Figure 2.1: Experimental set-up viewed from above.

## 2.3 Stimuli and Procedure

### 2.3.1 General

For comparison with the previous studies [1, 3, 2, 52] I decided to use natural color photographs as stimuli. Images used in the Training, Generalization Test, POR Test and the HSS Test were also used in the work of Grabner [3]. This stimulus continuity provided maximal comparability. All images were taken from free web galleries. Positive pictures (S+) contained one or multiple snails from different viewpoints (Class Snails, S, Fig 2.3) and negative pictures (S-) showed something else (Class No Snails, NS, Fig 2.3). Images with natural background were preferred for both S+ and S- to prevent any learning strategies based on pure background features. None of the S- images contained any snails or parts thereof. All animals, regardless of group membership, were reinforced on presentation of snails or their parts.

Photos were digitally processed to remove parts of a snail where necessary. Backgrounds remained intact to prevent learning from image processing artifacts. All stimuli were 128 x 128 pixels in size with 72 dpi thus producing a 45 x 45 mm picture on the screen. Images in the appendix have been decreased in size, in order to achieve print quality. For all tests a neutral contingency scenario was selected. This is, the stimulus disappeared after 11-20 seconds regardless of the subjects' pecking behavior. There was no reward for pecking and no penalty for withdrawal on such trials.

I used the same procedure as in the series of experiments by Aust&Huber and by Grabner. A visual schematics of the selected standard Go/No-Go procedure is provided Figure 2.2. The procedure started with an Inter Trial Interval (ITI) of 3 seconds when the chamber was not illuminated and no image was shown. The ITI was followed by a 11 to 30 seconds long image presentation phase. The duration of this phase was kept variable in order to keep the birds pecking the whole time and not only at the end of the phase. Pecks emitted throughout stimulus presentation were registered by the

computer, however, only those emitted in the first 10 seconds entered analysis. That is, if pecks were delivered only after the 10th second a response rate of 0 would have been registered for that trial. The presentation phase was followed by a decision phase with different termination criteria. In case of a positive stimulus, reward was offered after 5 successive pecks delivered within no more than 2 seconds. If a negative stimulus was present, the pigeon was required to refrain from pecking for 8 seconds in order for the stimulus to disappear. If the subject pecked during these 8 sec, the counter was reset and the subject had to wait another 8 seconds. There was no food reward on negative stimuli. The reward was the presentation of the next stimulus, which could possibly lead to reward again. Each stimulus, regardless of contingency, was followed by an ITI. At the beginning of every session a starter stimulus was shown. This was a positive stimulus that was not rewarded and not taken into statistical evaluation. By this, it was ensured that the pigeon was attentive and motivated to work.

The subjects were tested 5 days a week with 2-3 sessions a day. Each session consisted of 40 stimuli out of which 20 were negative and 20 positive. This proportion remained the same in all sessions. In tests, 8 test images replaced equal numbers of positive and negative training stimuli. In no case there was a different number of positive and negative images in a session. Chains of unreinforced stimuli longer than three were broken down to prevent demotivation. If a session was not finished on one day it continued the next day where it had previously ended.

One training-”cycle” consisted of 100 positive and 100 negative images, which were randomly divided into five sessions, each containing 20 S+ and 20 S-. One cycle thus consisted of the presentation of all five sets. Training images were used as fillers in all tests. They remained reinforced to ensure motivation in the subjects. The number of sessions for each phase can be found in Table 2.2.

## The Go/No-Go Procedure

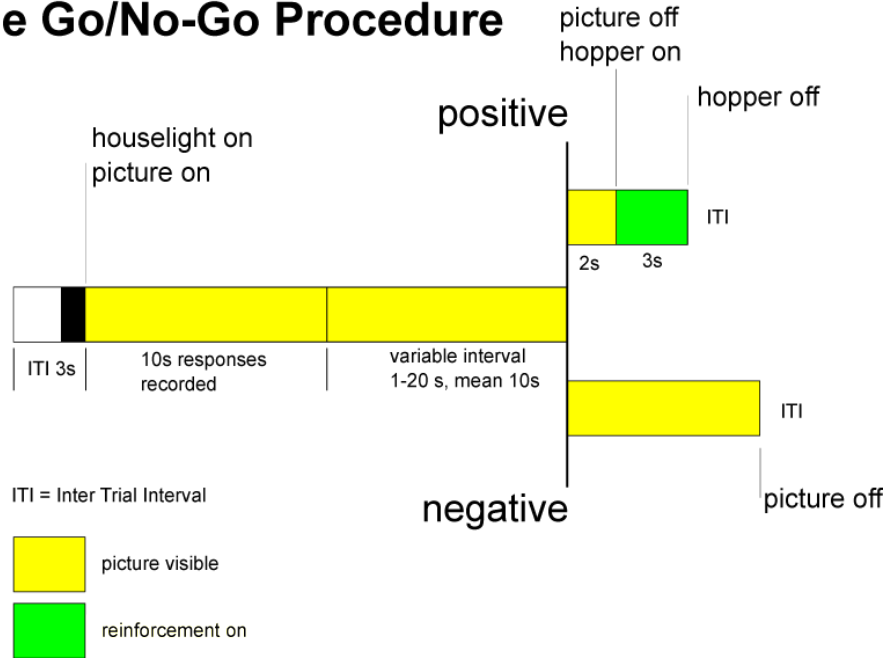


Figure 2.2: The Go/No-Go Procedure. Redrawn from [21]. The Inter Trial Interval (drawn black-and-white) where no stimulus is shown, is followed by a 11-30 sec presentation phase. Dependent of the stimulus class a termination criterion has to be fulfilled. Upon termination (shown as end of yellow line) the stimulus presentation is turned off and positive images get rewarded. Presentation of negative images is never rewarded. In both cases, after the end of the decision phase (positive or negative) the procedure is restarted and a new stimulus is selected.

Table 2.2: Experimental Phases. In sum, 720 images were used in 8 training and test phases.

Phase	Images	Sessions	Duration(Weeks)	Subjects
<b>Training</b>	200	14-20	2 +	6(+8)
<b>Generalization</b>	80	10	2	6(+8)
<b>HSS</b>	120	15	3	6
<b>POR</b>	80	10	2	2(+2)
<b>GrayScale</b>	80	10	2	6
<b>Blur</b>	80	10	2	6
<b>Scramble</b>	80	10	2	6
<b>Ammonites</b>	40	5	1	6
<b>Spirals</b>	40	5	1	6
<b>Sum</b>	720	90	36	8 (+6)

*Note.* Animals in brackets were tested by [3] but were included here for comparison reasons.

### **2.3.2 Training**

Animals from all three groups ( $n = 14$ ) started with the training phase. There was no pre-training as all animals were already experienced with both the procedure and the apparatus. Subjects had to learn the discrimination between Snails (Group FullSnails: Class S+) or parts of snails (Group NoHead and NoShell: Class S+) and images containing something else (all groups: Class S-).

Group Full Snails was trained on full (i.e., intact) snails containing all parts of a snail body (trunk, head and shell). No parts of the snail were removed or digitally processed. Group NoHead was shown only snail trunks plus shells and Group NoShell saw only the trunks plus the heads. Examples can be found in Table 2.3.

### **2.3.3 Generalization Test**

The training phase was followed by a Generalization Test. Its objective was to check for transfer of the learned discrimination to novel stimuli.

I selected a set of 40 positive and 40 negative images that were previously unseen by either of the birds. Each sessions contained a randomi selection of 4 positive and 4 negative novel images. All three groups were shown the same type of images as that in training (Full Snails/ No Head/No Shell, Table 2.3).

### **2.3.4 Picture-Object Recognition Test (POR)**





40 novel images of Seen Snail Parts (SP) were shown to the NoShell and NoHead groups. 40 novel previously Unseen Parts (UP) represented parts complementary to the parts shown in training. Stimulus examples can be found in Table 2.3. Additionally, 40 arbitrary skin patches (SK) were presented. Like in all other test phases, 8 test images were shown as part of one session. During each session 2-3 Shells, 2-3 Heads and 2-3 Skin Patches stimuli were selected and presented. The order of stimulus presentation and the number of stimuli of each type shown within a session were randomized

with the standard Java Math.random() function. All test images were presented once throughout testing.

### 2.3.5 Head vs Shell vs Skin Test (HSS)

Three classes of test stimuli 1) 40 Heads (Class He, Fig. 2.3c), 2) 40 Shells (Class Sh, Fig. 2.3d) and 3) 40 arbitrary Skin Patches (Class SK, Fig. 2.3e) were shown to the Group Full Snails. These were the same images as used during training of other groups. The procedure of stimulus selection was equal to the one described in Sec. 2.3.4

Table 2.3: Overview of the experimental phases in which the individual groups were shown full snails, heads, shells and skin patches

Image Type	Example	Gr. Full Snails	Gr. No Head	Gr. No Shell
<b>Full Snail</b>		Tr, Gen	Never	Never
<b>Head</b>		HSS-Test	POR-Test	Tr, Gen
<b>Shell</b>		HSS - Test	Training	POR-Test
<b>Skin</b>		HSS - Test	POR-Test	POR-Test

*Note.* Tr = training, Gen = Generalization Test, HSS = Head Vs Shell vs Skin Test, POR = Picture Object Recognition Test





(a) Positive training images



(b) Negative training images



(c) Heads



(d) Shells



(e) Skin Patches

Figure 2.3: Examples of stimuli used for Group FullSnail in Training, Generalization, POR and HSS Tests

### 2.3.6 Grayscale Test

The first experiment that included an impoverishment in image information was the Grayscale test. Color was removed from the stimuli by a simple calculation of the average intensity based on 3 primary colors in Gimp (version 2.6). The grayscale filter in Photoshop (all CS versions) uses a conversion ratio of: red=30%, green=59% and blue=11%. I did, however, not want to include any kinds of perceptual adaptation to the human visual system as this could cause error for the pentachromatic birds. Both default Gimp filters ("convert to grayscale" and "desaturate") use approximately the same function as Photoshop. To acquire an even color contribution to the overall luminosity I used a custom Channel Mixer (*Filters* → *Color* → *ChannelMixer.*) with red =  $\frac{1}{3}$ , green =  $\frac{1}{3}$ , blue =  $\frac{1}{3}$ . The test consisted of 40 positive and 40 negative images. All images were gray-scaled versions of randomly chosen training images. Examples are shown in Fig. 2.4a.

### 2.3.7 Blur Test

I decided to reduce the pictures' local feature content by blurring the images. This was done by means of the Photoshop CS2 Gaussian blur filter with a radius of 10 pixel. Each pixel value is thereby adjusted by a weighted sum of its neighbors. The weight decreases as the distance to the central pixel increases and is computed from a standard gaussian distribution  $G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$  (where x and y are the distances to the mid point and  $\sigma$  is the radius in pixels). Altogether 40 positive and 40 negative randomly selected training images were blurred. Examples are shown in Fig. 2.4b.

### 2.3.8 Scramble Test

For the Scramble Test, 40 positive and 40 negative randomly selected training images were modified. Examples are shown in Fig. 2.4c. I used a filter by Telegraphics

(<http://www.telegraphics.com.au/sw/info/scramble.html> , 3.6.2011) to split each image into blocks of 16x16 pixels . The image was divided into 8x8 equally-sized squares that were randomly repositioned in the image. Reordered parts were pixel squares and not based on biological pieces as it was the case in the experiments by Aust& Huber [1]. The authors of the previous work [25, 19? ] used different types of scrambling. I only used one type, where the background was also scrambled.

### **2.3.9 Test Ammonites**

Images of fossilized ammonites were taken from free image galleries and museum sites. This test included 40 novel images and had no negative counterpart. Examples are shown in Fig. 2.4d. Therefore the proportion of test and training images per session was as follows: 16 positive training images, 16 negative training images, 8 test images (ammonites).

### **2.3.10 Test Spirals**

Based on the results obtained in the previous tests I decided to test performance on artificial spiral-formed shapes on natural backgrounds. A logarithmic spiral was constructed in Python ( $x(t) = r(t) \cos(t)$ ;  $y(t) = r(t) \sin(t)$ ) and saved on a transparent background of 128x128 pixels. The spiral was then digitally added to negative training images (Gen-) in Photoshop (version CS2). In total 40 novel images were created. Examples are shown in Fig. 2.4e.



(a) Gray-scaled Images. Positive left, negative right.



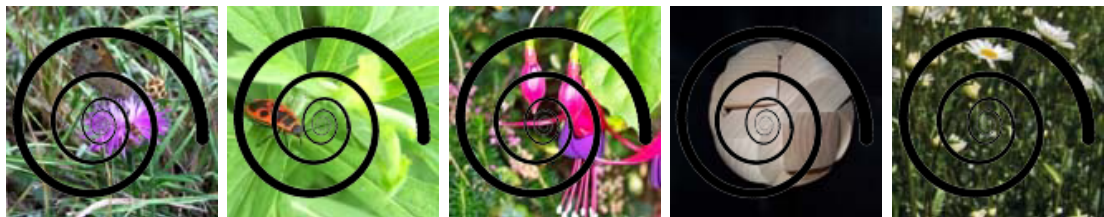
(b) Blurred images. Positive left, negative right.



(c) Scrambled images. Positive left, negative right.



(d) Fossilized ammonites



(e) Spirals

Figure 2.4: Examples of stimuli used for Group FullSnail in Grayscale, Blur, Scramble, Ammonites and Spirals Tests

## 2.4 Data Analysis

For analysis I used a self written suite of statistical analysis programs. The latest version can be found at <http://code.google.com/p/pylabsuite/>. The statistical analysis was based on Mann-Whitney U-tests and on Wilcoxon's matched-pairs signed-ranks tests. The significance level was uniformly set to 5% in contrary to the work of Grabner who used both 5% in training and the POR-test, but 1% in the generalization test.

Rho( $\rho$ ) is a statistical measure introduced by Herrnstein in [53]. When Rho is 1, perfect discrimination is present and all images from the positive class ranked above all members from the negative class (i.e., subjects pecked more on all positive than on all negative exemplars). A Rho of 0.5 is given if both image classes are perfectly intermixed and no difference in the response rates is present. More generally, the rho value represents the chance that a member of the positive image class gets a higher ranking (more pecks) than a member from the negative class. Therefore a rho below 0.5 is interpreted as reversed ranking.

$$\rho = \frac{1 - U}{n1 \times n2} \quad (2.1)$$

Equation 2.1 shows the calculation of Rho, with  $U = U_{\text{statistic}}$  (in Mann Whitney test);  $n1$  = size of group one,  $n2$  = size of group two.

Animals had to remain in training until they reached  $\rho \geq 0.676$  (which is the limit of significance for  $\alpha = 0.05$ ;  $n1=n2=20$ ) in five successive trials. In tests, performance assessments were based on the mean standardized response rate (MSRR), which is calculated as follows: Response rates are normalized separately for each session, based on the average peck rate emitted to training images in that session. Test images are excluded from calculating the average peck rate of a session to avoid self reference. Then the absolute number of pecks emitted to each training and each test picture in a session is divided by the average peck rate of that session. Performance on individual

types of test stimuli is calculated by averaging the standardized response rates emitted throughout testing to all stimuli of that type. The resulting mean is the MSSR for a particular type of stimuli.

$$MSRR = \frac{\sum \frac{RR_i}{\overline{RR_t}}}{n} \quad (2.2)$$

Equation 2.2 shows the calculation of the mean standardized response rate where MSRR = mean standardized response rate,  $RR_i$  = response rate to a test image  $i$ ;  $\overline{RR_t}$  = average number of responses emitted to training stimuli in a particular session and  $n$  = number of stimuli.

# Chapter 3

## Results

### 3.1 Training

Results of the trainings can be found in Table 3.1. It shows average rho-values for each cycle and subject, separately for the three groups. The performance in each session is shown in Fig.3.1. All subjects of Group Full Snails managed to reach the learning criterion during the third or fourth cycle. Animals trained on parts of snails by Grabner [3], took five to thirteen cycles (range Group No Head: 5-8; range Group No Shell: 9-13). All four animals of the No Head Group acquired the discrimination. Two animals from the No Shell Group (Bobbison and Verena) did not learn the task within 20 cycles (i.e., 100 sessions) and were discontinued. The other two animals (Josef and Klara) were transferred to the generalization test. Thus a total of 12 animals were further tested. The average numbers of sessions the animals required to reach the learning criterion can be found in Figure 3.2. Animals of Group Full Snails were significantly faster than animals of both other groups (Mann-Whitney U; both  $p \leq 0.01$ ). Group Full Snails required  $6(\pm 2)$  sessions (range 4-20) which is only a half of the sessions required by Group No Head  $32(\pm 6)$  and less than one fourth of the sessions required by Group No Shell  $77(\pm 27)$ .

Table 3.1: Mean rho-value per cycle for the individuals of all groups. Significant values are given in bold ( $\rho \leq 0.676$ ). The upper table shows the animals tested as part of this work and the lower part those tested by Grabner in[3].

Cycle	Group Full Snails					
	Anatol	Hannes	Lukas	Niki	Olivia	Rhonda
1	0.638	0.608	0.589	0.632	0.599	0.618
2	<b>0.727</b>	<b>0.695</b>	<b>0.733</b>	<b>0.682</b>	<b>0.841</b>	<b>0.71</b>
3	<b>0.828</b>	<b>0.74</b>	<b>0.776</b>	<b>0.82</b>	<b>0.803</b>	0.65
4	<b>0.841</b>	<b>0.858</b>	<b>0.79</b>			<b>0.811</b>
3						
4						
5						
6						
7						
8						
...	...	...	...	...	...	...
20						

Cycle	Group No Head				Group No Shell			
	Daisy	Cord.	Franz	Herb.	Josef	Klara	Bobb.	Verena
1	0.55	0.595	0.613	0.513	0.502	0.509	0.475	0.558
2	<b>0.717</b>	0.579	<b>0.781</b>	0.602	0.627	0.63	0.519	0.564
3	<b>0.677</b>	0.427	0.674	0.626	0.502	0.646	0.593	0.56
4	<b>0.773</b>	0.518	0.483	<b>0.695</b>	0.568	0.612	0.589	0.514
5	<b>0.846</b>	<b>0.733</b>	0.658	0.631	0.561	0.493	0.536	0.572
6		<b>0.762</b>	<b>0.716</b>	<b>0.738</b>	0.648	<b>0.709</b>	0.573	0.611
7			<b>0.735</b>	<b>0.816</b>	<b>0.706</b>	<b>0.706</b>	<b>0.678</b>	0.542
8				<b>0.798</b>	<b>0.732</b>	0.586	0.582	0.659
9					<b>0.743</b>	0.587	0.646	<b>0.679</b>
10						0.633	<b>0.69</b>	0.636
11						0.625	0.652	0.645
12						0.59	<b>0.691</b>	0.647
13						<b>0.733</b>	0.494	<b>0.696</b>
14							0.572	0.667
15							0.637	0.609
16							0.526	0.603
17							0.589	0.585
18							0.672	0.638
19							0.601	<b>0.689</b>
20							0.624	0.624



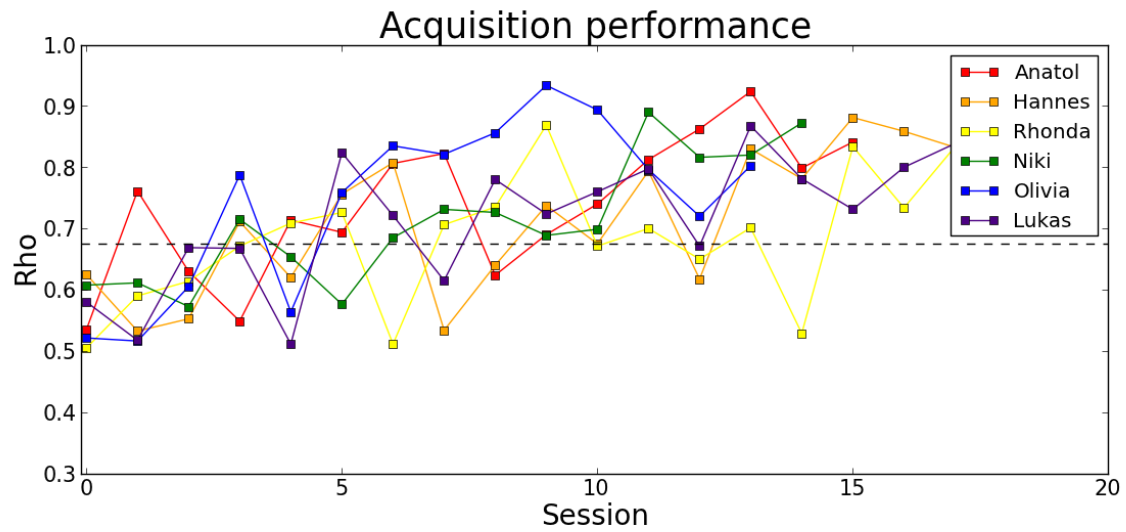


Figure 3.1: Acquisition performance of the individual subjects in Group Full Snails shown as rho-values and separately for each session. The black dashed horizontal line indicates the limit of significance ( $\rho = 0.676$ ). The x axis is drawn with an increment of five, indicating the division into cycles.

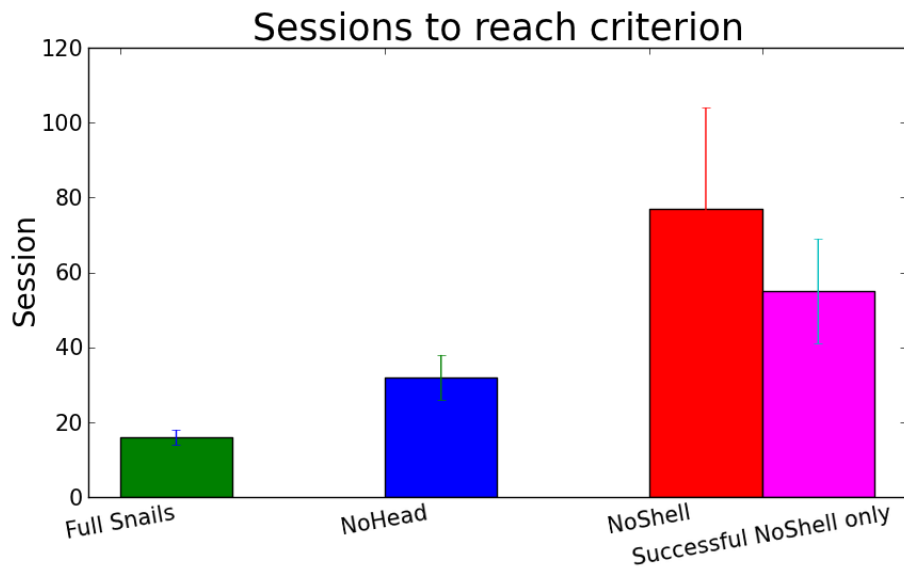


Figure 3.2: Mean ( $\pm$  SD) number of sessions required to reach criterion, shown separately for the three groups. Group Full Snails ( $n = 6$ ) was significantly faster than Group No Head ( $n = 4$ ) and Group No Shell ( $n = 4$ ). Results of Group No Shell are depicted separately for the two birds that mastered the task (magenta bar) and together with the two animals that failed to reach criterial level within 20 sessions and were discontinued (red bar).

## 3.2 Generalization Test

The results of all groups are summarized in Table 3.2 and a visualization of the results for Group Full Snails can be found in Figure 3.3. All animals of Group Full Snails that were successful during training performed significantly in the Generalization Test. The Mann Whitney U Test (Gen+/Gen-) yielded significance in all six animals ( $p \leq 0.001$ ). The threshold was set to 5% and therefore these results point out to a high level of discrimination.

This was not the case in Groups No Head and No Shell where 3 and 1 animal passed the test. Josef (No Shell) and Daisy (No Head) were tested by Grabner. Cordula and Franz (both No Head) were tested as part of this thesis. The No Shell animal Herbert ( $p = 0.1504$ ) and the No Head animal Klara ( $p = 0.0999$ ) tested by Grabner did not perform at a significance level. All other animals managed to discriminate significantly between positive and negative test stimuli.

The peck rates to positive generalization images were higher than peck rates to the negative images in all animals. Rhonda, a subject of Group Full Snails, had the highest MSRR for positive images (1.626) and the lowest value was achieved by Herbert (1.063; Group No Head). The lowest MSRR for negative images was found for Daisy (0.438; Group No Head) and the highest for Josef (0.904; Group No Shell)

Table 3.2: Performance of the individual subjects in the Generalization Test, shown separately for the three groups. The first and the second column give the MSSRs achieved for the novel positive (Gen+) and the novel negative (Gen-) stimuli. The third column shows the results of the Mann Whitney U Test (p-values) comparing the two stimulus types (Gen+/Gen-). Significant results are denoted with a bold font. To avoid confusion the responsible experimenter is also given.

Group	Subject	Gen Test			
		Gen +	Gen -	p	Experimenter
Full Snails	Anatol	1.475	0.845	<b>0.001</b>	Kysel
	Hannes	1.288	0.847	<b>0.002</b>	Kysel
	Lukas	1.262	0.721	<b>0.002</b>	Kysel
	Niki	1.542	0.605	<b>0.0001</b> >	Kysel
	Olivia	1.23	0.751	<b>0.004</b>	Kysel
	Rhonda	1.626	0.689	<b>0.0001</b> >	Kysel
No Shell	Josef	1.489	0.904	<b>0.0092</b>	Grabner
	Klara	1.369	0.444	0.0999	Grabner
No Head	Daisy	1.191	0.438	<b>0.0001</b>	Grabner
	Cordula	1.482	0.638	<b>0.0001</b> >	Grabner
	Franz	1.272	0.702	<b>0.001</b>	Grabner
	Herbert	1.063	0.532	0.1504	Grabner

### 3.3 Picture-Object Recognition Test

Both Groups (No Head and No Shell) pecked more on arbitrary Skin Patches (SK) than on Unseen Parts (UP) of snails (see Fig. 3.4a). However, this difference was not significant in either animal. Two animals (Cordula and Franz) responded more to SK and one animal (Daisy) to UP. By contrast, in the work of Aust and Huber [2] all birds responded more to UP than to SK stimuli, and this difference was significant (see Fig. 3.4b).

The results of the two additional animals tested in the current study confirmed, that Daisy’s responding behavior was indeed different from that of the other animals. The MSRRs for SP of Cordula [1.195(0.854)] and Franz [1.111(0.775)] were similar to each other and less similar to the MSRRs of Daisy [ $\pm 0.798(0.95)$ ]. The MSRRs for both animals were above 1.0 which indicates some recognition of previously seen parts.

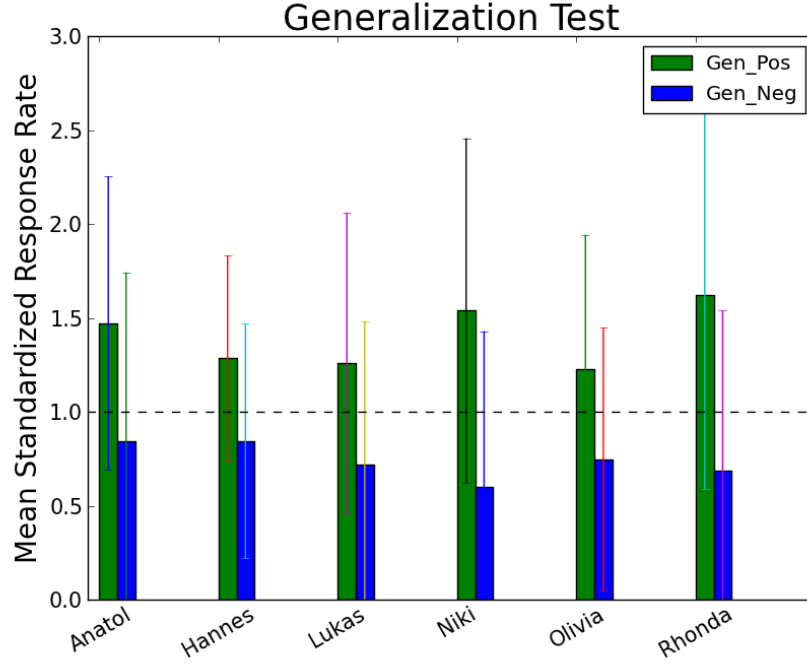


Figure 3.3: Results of the Generalization Test in Group Full Snails shown as MSSR ( $\pm$ SD). The horizontal dashed line indicates the level of the average peck rate (1.0). Response rates above that line indicate a tendency to treat the respective image class as positive rather than as negative (and vice versa for values below the line). Gen\_Pos= Positive Generalization stimuli, Gen\_Neg= Negative Generalization stimuli.

For UP, the MSRR of Cordula [ $\pm 0.885(0.81)$ ] was similar to Franz [ $\pm 0.839(0.74)$ ] and only related to Daisy [ $0.626(0.834)$ ]. There was, however, a difference between the additionally tested birds and Daisy in the ranking of Skin Patches. The MSRRs of Cordula [ $\pm 0.95(0.798)$ ] and Franz [ $\pm 1.014(0.854)$ ] were close to the average peck rate of 1.0 and therefore above the MSRRs towards UP. This is different from the results of Daisy [ $\pm 0.424(0.605)$ ] who responded less to this image class than to the UP stimuli. Both Cordula and Franz showed no difference in the ranking of Unseen Parts and Skin patches. Because the difference between those two classes is the biggest in the MSRRs of Daisy [ $\pm(\Delta 202)$ ] it got a ranking of 3, Franz [ $\pm(\Delta 175) -2$ ] and Cordula [ $\pm(\Delta 65) -1$ ], the Wilcoxon-Test indicated a perfect class intermixture. The resulting value  $T=1.0$  might seem odd, but this is the case for the given data set and group

size. The difference between SP and negative generalization images was significant in all animals. MSRRs emitted to UP was significantly below MSSRs emitted to novel positive images in all No Head animals. The results for all animals are listed in Table 3.3 and a visualization can be found in Figure 3.4a.

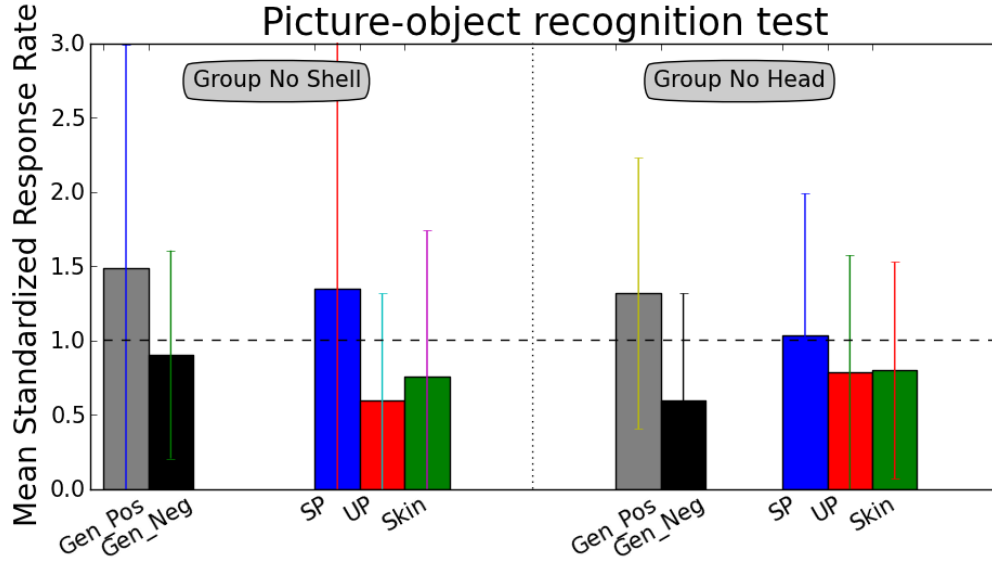
Table 3.3: Results of the POR Test. The top part of the table shows the results of parts (SP, UP, SK) as MSSRs. These are shown separately for the individual birds as well as averaged across subjects (mean( $\pm$ SD)). The middle and bottom part of the table show the results of Mann-Whitney U tests (p-values) for the individual birds and Wilcoxon matched-pairs signed-ranks test (p-values) including performance of all birds. The middle panel compares test parts and novel generalization images. The bottom panel compares the three test parts with each other.

Group	Subject	POR Test					
		SP	SD	UP	SD	SK	SD
No Shell	Josef	1.345	2.711	0.597	0.716	0.755	0.983
No Head	Daisy	0.798	0.905	0.626	0.834	0.424	0.605
	Cordula	1.195	0.854	0.885	0.81	0.95	0.798
	Franz	1.11	0.775	0.839	0.74	1.014	0.854
	<i>Mean(<math>\pm</math>SD)</i>	<i>1.034(0.845)</i>		<i>0.783(0.795)</i>		<i>0.796(0.752)</i>	

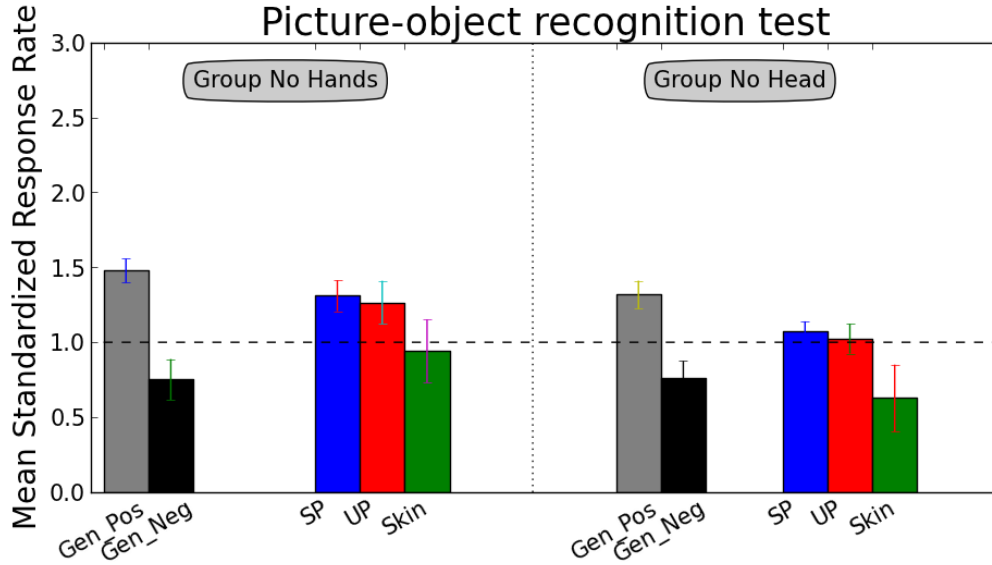
Group	Subject	POR Test					
		Gen+ /SP	Gen- /SP	Gen+ /UP	Gen- /UP	Gen+ /SK	Gen- /SK
No Shell	Josef	0.996	<b>0.0003</b>	0.249	<b>0.026</b>	0.231	<b>0.029</b>
No Head	Daisy	0.1197	<b>0.015</b>	<b>0.015</b>	0.132	<b>0.0007</b>	0.169
	Cordula	0.103	<b>0.036</b>	<b>0.0016</b>	0.16	<b>0.006</b>	0.1
	Franz	0.139	<b>0.033</b>	<b>0.01</b>	0.373	0.154	0.631
	<i>Wilcoxon</i>	<i>0.054</i>	<i>0.054</i>	<i>0.054</i>	<i>0.054</i>	<i>0.054</i>	<i>0.296</i>

Group	Subject	POR Test			
		SP/UP	SP/SK	UP/SK	Experimenter
No Shell	Josef	<b>0.041</b>	0.1218	0.7639	Grabner
No Head	Daisy	0.3372	0.7677	0.1137	Grabner
	Cordula	0.256	0.465	0.672	Kysel
	Franz	0.192	0.644	0.279	Kysel
	<i>Wilcoxon</i>	<i>0.054</i>	<i>0.054</i>	<i>1.0</i>	

*Note.* SP/UP = seen parts to unseen parts, SP/SK = seen parts to skin patches, UP/SK = unseen parts to skin patches; Gen+ = novel positive images; Gen- = novel negative images. Significant results are denoted with a bold font. The cursive font identifies group level statistics.



(a) Groups No Shell and No Head in the snail experiment.



(b) Groups No Hands and No Head in the human experiment.

Figure 3.4: a) Results of the POR-Test in the snail experiment, shown as MSSRs. Means were taken across all birds of each group ( $\pm$ SD), including those tested by Grabner and in the present study. b) Results of the POR-Test in the human experiment, shown as MSSRs. Means were taken across all birds of each group. Green bars indicate performance on arbitrary skin patches. Blue bars indicate performance on Seen Parts (SP) and red bars indicate performance on Unseen Parts (UP). Figure 3.4a is an adaptation and extension of the figures by [3], and Figure 3.4b has been redrawn from [1]. For reasons of comparison, peck rates emitted to the positive and negative transfer stimuli of the generalization test are shown as well in both figures (gray bars). The horizontal dashed line indicates the average response rate (1.0).

### 3.4 Head vs Shell vs Skin Test (HSS)

Figure 3.5 shows the results of Group Full Snails in the HSS-Test. As can be seen, there was a difference between response rates to Shells, Heads and arbitrary Skin Patches. Only Shells were considered positive rather than negative, with the MSSR being  $1.185(\pm 0.948)$  and therefore exceeding the average response level of 1,0. Heads and Skin Patches were considered rather negative with the MSSR being  $0.619(\pm 0.756)$  for Heads and  $0.826(\pm 0.813)$  for Skin Patches.

Table 3.4 shows the MSSRs for the three types of test stimuli (top panel), and the results of Mann-Whitney U-Tests that compared the three test stimulus types of the HSS-Test with the novel positive and negative stimuli of the previous generalization test (middle panel) as well as with each other (bottom panel). Significant values ( $p < 0.05$ ) are denoted with a bold font. Also, the results of Wilcoxon tests including the data of all birds are shown. The values in the table are not corrected for multiple analyses. This is due to the fact that the main question of the test concerned the difference between UP and SK, while all other comparisons only added supplementary information. Yet I still consider all of these side tests very important. A result of 0.021 in the Wilcoxon test equals the highest level of significance for the given group size. It indicates a perfect discrimination where all peck rates from one image class were ranked above the peck rates from another class. Results with better confidence levels would require a bigger sample size.

The middle panel shows that both classes Heads and Skin Patches were treated differently from the positive class yet not differently from the negative class. In contrary, the third tested part of a snail, the Shell was treated differently from the negative and similarly as the positive class. Only one animal, Niki, pecked at Shells (MSRR = 0.879) significantly less ( $p = 0.006$ ) than to positive generalization images (MSRR = 1.542). For all other animals, no significant difference between shells and positive generalization images was found. Accordingly, the comparison between Shells and the

negative class, revealed a significant difference in all animals except Niki. On the group level, there was no significant difference between Shells and novel positive images. The Wilcoxon test  $p=0.059$  also indicated no group difference. Hannes responded even more to Shells ( $MSRR = 1.425$ ) than to novel positive images ( $MSRR= 1.288$ ), but the difference was not significant ( $p = 0.559$ ). The MSRRs that were used in the Wilcoxon test were higher in novel generalization images compared to Shells in all remaining animals. Both heads and skins were judged to be different from positive generalization images on the group level. The only inconsistency in group vs. subject ranking was between Heads and negative generalization images (HE/Gen-). There was no difference between the two classes on the subject level, yet the average response rates indicated discrimination on the group level. This means that slight (non-significant) individual tendencies added up to yield significance on the group level.

In the bottom panel of Table 3.4 we see the differences in MSRR among the stimulus types showing complementary parts of a snail. The only significant difference was found between Shells and Heads. Apart from Niki, all animals pecked significantly more to Shells than to Heads. At least, however, the result of Niki ( $p = 0.052$ ) showed the same trend as was found in the other animals. This is supported by the group level Wilcoxon Test. All animals besides Lukas pecked more to SK than to HE. The Wilcoxon ( $p = 0.036$ ) indicated a difference between the two stimulus types. Yet there was no significant difference between Skin Patches and Heads in any individual bird. The peck rates emitted to SH were higher than those emitted to HE in all animals (Wilcoxon  $p = 0.021$ ). On the individual level, only two birds - Hannes and Lukas - showed a significant difference.

Taken together the results of the HSS test indicate that the animals of Group Full Snails used Shells rather than Heads or Skin as discriminative cues.



Table 3.4: Head vs Shell vs Skin Test results. Top panel: MSRRs achieved on the individual types of test stimuli. Results are shown separately for the individual birds of Group Full Snails and averaged across subjects ( $\pm$ SD). Middle and bottom panel: Results of the Mann Whitney U Tests (p-values) obtained for the HSS-Test. The last row gives the results of Wilcoxon matched-pairs signed ranks tests across all six birds (p-values). Middle pannel: Comparison of performances on the individual types of test stimuli with performance in the Generalization Test. Bottom panel: Comparison of the three test stimulus types of the HSS-Test with each other..

Group	Subject	HSS-Test					
		SH	SD	HE	SD	SK	SD
Full Snails	Anatol	1.254	1.034	0.535	0.734	0.796	0.85
	Hannes	1.425	0.871	0.723	0.738	0.979	0.72
	Lukas	1.037	0.802	0.709	0.708	0.643	0.698
	Niki	0.879	1.023	0.391	0.718	0.759	0.973
	Olivia	1.226	0.806	0.712	0.753	0.919	0.752
	Rhonda	1.29	1.154	0.644	0.883	0.859	0.887
	<i>Mean(<math>\pm</math>SD)</i>	<i>1.185 (0.948)</i>		<i>0.619 (0.756)</i>		<i>0.826 (0.813)</i>	

Group	Subject	HSS-Test					
		Gen+ /SH	Gen- /SH	Gen+ /HE	Gen- /HE	Gen+ /SK	Gen- /SK
Full Snails	Anatol	0.386	<b>0.012</b>	<b>0.0001</b> >	0.055	<b>0.001</b>	0.788
	Hannes	0.559	<b>0.003</b>	<b>0.0001</b> >	0.335	<b>0.036</b>	0.405
	Lukas	0.233	<b>0.034</b>	<b>0.002</b>	0.878	<b>0.001</b>	0.878
	Niki	<b>0.006</b>	0.59	<b>0.0001</b> >	0.052	<b>0.0001</b> >	0.765
	Olivia	0.878	<b>0.004</b>	<b>0.004</b>	0.806	<b>0.08</b>	0.329
	Rhonda	0.223	<b>0.004</b>	<b>0.0001</b> >	0.862	<b>0.001</b>	0.174
	<i>Wilcoxon</i>	<i>0.059</i>	<i>0.021</i>	<i>0.021</i>	<i>0.021</i>	<i>0.021</i>	<i>0.093</i>

Group	Subject	HSS-Test			
		SH/HE	SH/SK	HE/SK	Experimenter
Full Snails	Anatol	<b>0.002</b>	0.07	0.118	Kysel
	Hannes	<b>0.0001</b> >	<b>0.036</b>	0.058	Kysel
	Lukas	<b>0.035</b>	<b>0.016</b>	0.711	Kysel
	Niki	0.052	0.821	0.061	Kysel
	Olivia	<b>0.004</b>	0.057	0.264	Kysel
	Rhonda	<b>0.002</b>	0.096	0.103	Kysel
	<i>Wilcoxon</i>	<i>0.021</i>	<i>0.021</i>	<i>0.036</i>	

*Note.* SH = shell, HE = head, SK = skin, SD = standard deviation, Gen+ = positive generalization stimuli, Gen- = negative generalization stimuli. Significant results are denoted with a bold font. The cursive font identifies group level statistics.

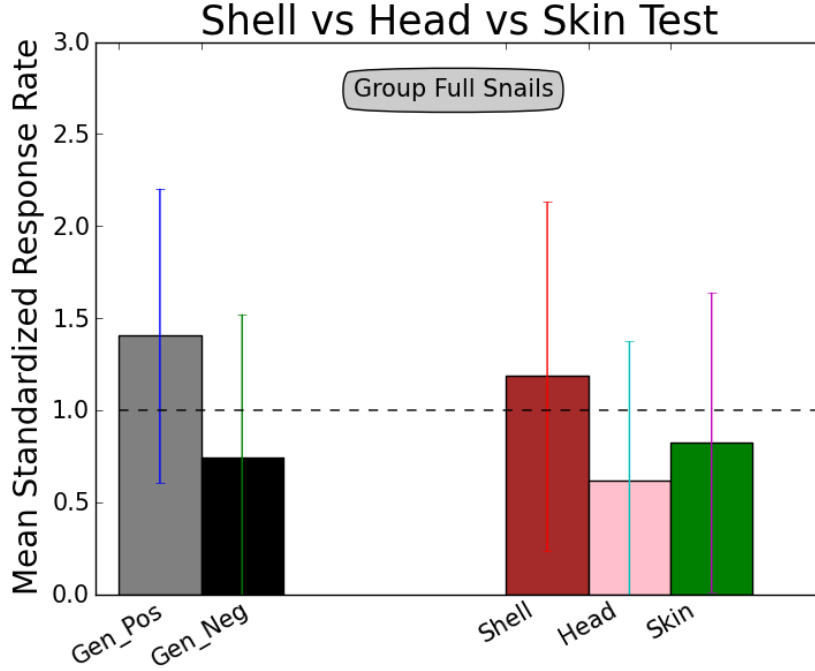


Figure 3.5: Results of the Head vs Shell vs Skin Test as MSSRs, shown separately for the individual test stimulus types. Means were taken across all six subjects. The height of the bar represents the average response rate of all animals in the group ( $\pm$ SD). Shell stimuli are represented by a brown bar so as Head stimuli are represented by a pink bar and Skin stimuli by a green bar. For reasons of comparison, peck rates emitted to the positive and negative transfer stimuli of the generalization test are shown as gray bars. The horizontal dashed line indicates the average response rate (1.0).

### 3.5 GrayScale Test

The results of the grayscale test are summarized in Table 3.5. and illustrated in Figure 3.6 (in comparison with the results of the Scramble, Blur, Ammonite, and Spiral Tests). A reduction in the amount of color information resulted in significant drops in MSSRs in both positive [ $0.659(\pm 0.709)$ ] and negative [ $0.229(\pm 0.449)$ ] images. The results of the Mann Whitney U Test for Niki ( $p = 0.0005$ ), Olivia ( $p \leq 0.0001$ ) and Rhonda ( $p = 0.014$ ) show that the animals were still able to discriminate significantly between positives and negatives when shown in grayscale. As for Anatol ( $p = 0.062$ ) and Hannes ( $p = 0.054$ ), these animals showed a correct tendency at least. Only in Lukas ( $p = 0.296$ ) discrimination performance was strongly impaired.

The MSSR of Olivia was almost as high for positive gray-scaled images (1.187) as for the original (colored) generalization images (1.23). She showed no significant difference (Mann-Whitney U;  $p = 0.985$ ) between these two image classes. All other animals responded significantly less to Grayscaled images than to their originals ( $p \leq 0.0001$ ). The MSSR obtained for negative grayscaled images [ $0.229 \pm 0.449$ ] differed significantly from that obtained to the negative originals [ $0.743 \pm 0.778$ ]. The level of confidence was below 5% in all animals, and below 1% in Anatol, Hannes and Lukas. Because the drop in the MSSR occurred uniformly across animals and for both positive and negative stimuli, all three tests (Table 3.5: GS+/GS-, GS+/Gen+ and GS-/Gen-) were significant on group level (Wilcoxon U = 0.021).

Table 3.5: Results of the Grayscale test. Left panel: mean standardized response rate on positive (GS+) and negative (GS-) gray-scale images, shown separately for the individual birds, and with means ( $\pm$ SD) being taken across birds. Right panel: Results of Mann Whitney U Tests and Wilcoxon Tests (p-values) comparing positive and negative grayscale images (GS+/GS-), and grayscale images with positive and negative transfer stimuli from the generalization test (GS+/Gen+ and GS-/Gen-).

Subject	Grayscale Test						
	GS+	SD	GS-	SD	GS+/GS-	GS+/Gen+	GS-/Gen-
Anatol	0.416	0.618	0.166	0.473	0.062	<b>0.0001&gt;</b>	<b>0.0001&gt;</b>
Hannes	0.675	0.802	0.214	0.364	0.054	<b>0.0001&gt;</b>	<b>0.0001&gt;</b>
Lukas	0.436	0.631	0.191	0.404	0.296	<b>0.0001&gt;</b>	<b>0.001</b>
Niki	0.763	0.669	0.287	0.507	0.0005	<b>0.0001&gt;</b>	<b>0.026</b>
Olivia	1.176	0.811	0.342	0.478	0.0001>	0.985	<b>0.011</b>
Rhonda	0.487	0.719	0.175	0.47	0.014	<b>0.0001&gt;</b>	<b>0.037</b>
Mean( $\pm$ SD)	0.659(0.708)		0.229(0.449)				
Wilcoxon					0.021	0.021	0.021

*Note.* SD = standard deviation, Gen+ = positive generalization stimuli, Gen- = negative generalization stimuli. Significant results are denoted with a bold font. The cursive font identifies group level statistics.

### 3.6 Blur Test

The results of the Blur test are summarized in Table 3.6 and illustrated in Figure 3.6. With the exception of Rhonda (Bl+/Gen+;  $p \leq 0.0001$ ), there was no significant difference in MSSR between novel positive generalization stimuli and the blurred stimuli. On the contrary, Anatol, Lukas, Niki and Hannes showed even increased pecking at blurred versions of previously seen training images compared to the positive generalization stimuli. On group level, there was no difference between the Gen+ and the Bl+ class (Wilcoxon  $p=0.979$ ). Discrimination between positive blurred images [MSSR 1.279( $\pm 0.947$ )] and negative blurred images [MSSR 0.506( $\pm 0.743$ )] was present in all animals as well as on group level. In some cases, there was a significant drop in peck rates emitted to blurred non-snail images compared to novel negative stimuli (Bl-/Gen-). The drop was significant for Anatol ( $p=0.013$ ), Olivia ( $p=0.037$ ) and Rhonda ( $p=0.021$ ). On group level, the drop was also significant (Wilcoxon  $p = 0.036$ ).

Table 3.6: Results of the Blur test. Left panel: mean standardized response rates on positive (Bl+) and negative (Bl-) blurred images, shown separately for the individual birds, and with means ( $\pm$ SD) being taken across birds. Right panel: Results of Mann Whitney U Tests and Wilcoxon Tests (p-values) comparing positive and negative blurred images (Bl+/Bl-), and blurred images with positive and negative transfer stimuli from the generalization test (Bl+/Gen+ and Bl-/Gen-)

Subject	Blur Test						
	Bl+	SD	Bl-	SD	Bl+/Bl-	Bl+/Gen+	Bl-/Gen-
Anatol	1.725	1.154	0.494	0.807	<b>0.0001</b> >	0.577	<b>0.013</b>
Hannes	1.303	1.064	0.647	0.861	<b>0.004</b>	0.977	0.053
Lukas	1.355	0.635	0.571	0.647	<b>0.0001</b> >	0.564	0.363
Niki	1.722	1.264	0.673	0.914	<b>0.0001</b> >	0.567	0.651
Olivia	0.98	0.786	0.416	0.634	<b>0.0006</b>	0.206	<b>0.037</b>
Rhonda	0.587	0.776	0.235	0.593	<b>0.0066</b>	<b>0.0001</b> >	<b>0.021</b>
<i>Mean(±SD)</i>	<i>1.279(0.947)</i>		<i>0.506(0.743)</i>				
<i>Wilcoxon</i>					<i>0.021</i>	<i>0.979</i>	<i>0.036</i>

*Note.* SD = standard deviation, Gen+ = positive generalization stimuli, Gen- = negative generalization stimuli. Significant results are denoted with a bold font. The cursive font identifies group level statistics.

### 3.7 Scramble Test

The results of the Scramble test are summarized in Table 3.7 and illustrated in Figure 3.6. With a MSSR of  $[0.469(\pm 0.692)]$  for the positive class and  $[0.106(\pm 0.274)]$  for the negative class, there was a significant drop compared to novel unchanged images. The MSRRs of all six animals were significantly lower for positive scrambled stimuli than for novel positive generalization images ( $p \leq 0.001$ ). The drop in the negative class was also highly significant ( $p \leq 0.0001$ ) in all animals but Lukas. This subject showed a significant drop ( $p = 0.02$ ). Furthermore, there was a difference on both animal and group level (Wilcoxon  $p = 0.021$ ). The recognition of S+ images modified with the blurring filter remained intact in four out of six animals. The p-values (Sc+/Sc-) were as follows: Anatol (0.0034), Lukas (0.0011), Niki (0.0189) and Olivia (0.0073). Hannes ( $p = 0.1101$ ) and Rhonda ( $p = 0.0742$ ) showed a correct though non-significant tendency. On group level a significant difference was given (Wilcoxon  $p = 0.021$ ). As we can see in Figure 3.6, scrambling caused the most radical drop in MSRRs of all depicted stimulus manipulations.

### 3.8 Ammonites and Spirals Tests

Test Ammonites created a cross-modal snail-like image class. While bearing some overall similarity with snails (at least to the human observer), it nevertheless deviates from snails along various features and dimensions (e.g., regarding color, local and global information). As we can see in Table 3.8 and Figure 3.6 ammonites elicited an average peck rate of  $[0.761(\pm 0.884)]$  and were thus considered as negative rather than as positive. As confirmed by U-tests for the individual birds and a Wilcoxon test at group level, there was no significant difference between Ammonites and negative Generalization images. Lukas (MSRR=0.855), Niki (MSRR=0.871) and Olivia (MSRR=0.945) pecked more to Ammonite images. Anatol (MSRR=0.723), Hannes (MSRR=0.693)

and Rhonda (MSRR=0.480) during the presentation of negative Generalization images. Performance of all animals except Olivia differed significantly between positive transfer stimuli used in the Generalization Test and the Ammonites. Olivia treated Ammonites (MSRR=0.945) similarly as the positive class (MSRR=1.23), with a respective Mann Whitney U test failing to reveal a significant difference.

Test Spirals seemed to add global (snail like) information to the negative Generalization images. Namely, the shape of the Shell was approximated by a simple spiral. As we can see from Figure 3.6 this was, however, not sufficient to make these stimuli members of the positive class (snails). The average peck rate was  $[0.844(\pm 1.024)]$  and the individual U-tests as well as the Wilcoxon failed to reveal any significant differences between Spirals and negative Generalization images. Anatol pecked less while presented with Spirals, but the difference was not significant ( $p=0.248$ ). Hannes ( $p=0.158$ ), Lukas ( $p=0.097$ ) and Niki ( $p=0.985$ ) pecked more. Only Olivia showed a significantly higher response rate ( $p = 0.024$ ) to Spirals (MSRR=1.133) than to the negative Generalization stimuli (MSRR=0.751). This particular animal treated Spirals as positives in terms of MSRRs and as revealed by the U-test ( $p = 0.672$ ). On the contrary, Rhonda responded significantly less ( $p = 0.017$ ) to Spirals (MSRR=0.229) than to negative transfer images from the Generalization Test (MSRR=0.689) and the difference to the positive generalization stimuli was highly significant  $p \leq 0.0001$ . The difference between the positive generalization images and Spirals for all other animals (Anatol, Hannes, Lukas and Niki) was significant at least at the 5% level.

Table 3.7: Results of the Scramble test. Left panel: mean standardized response rate on positive (Sc+) and negative (Sc-) scrambled images, shown separately for the individual birds, and with means ( $\pm$ SD) being taken across birds. Right panel: Results of Mann Whitney U Tests and Wilcoxon Tests (p-values) comparing positive and negative scrambled images (Sc+/Sc-), and scrambled images with positive and negative transfer stimuli from the generalization test (Sc+/Gen+ and Sc-/Gen-).

Subject	Scramble Test						
	Sc+	SD	Sc-	SD	Sc+ / Sc-	Sc+ / Gen+	Sc- / Gen-
Anatol	0.515	0.752	0.048	0.227	<b>0.0034</b>	<b>0.0001</b> >	<b>0.0001</b> >
Hannes	0.32	0.656	0.053	0.223	0.1101	<b>0.0001</b> >	<b>0.0001</b> >
Lukas	0.692	0.632	0.279	0.396	<b>0.0011</b>	<b>0.001</b>	<b>0.02</b>
Niki	0.31	0.502	0.056	0.268	<b>0.0189</b>	<b>0.0001</b> >	<b>0.0001</b> >
Olivia	0.669	0.845	0.19	0.487	<b>0.0073</b>	<b>0.001</b>	<b>0.0001</b> >
Rhonda	0.308	0.762	0.009	0.042	0.0742	<b>0.0001</b> >	<b>0.0001</b> >
Mean( $\pm$ SD)	0.469(0.692)		0.106(0.274)				
Wilcoxon					0.021	0.021	0.021

Note. SD = standard deviation, Gen+ = positive generalization stimuli, Gen- = negative generalization stimuli. Significant results are denoted with a bold font. The cursive font identifies group level statistics.

Table 3.8: Results of the Ammonites and Spirals Tests. Left panel: mean standardized response rates to ammonites and spirals, shown separately for the individual birds, and with means ( $\pm$ SD) being taken across birds. Right panel: Results of Mann Whitney U Tests and Wilcoxon Tests (p-values) comparing ammonites and spirals with negative transfer stimuli from the generalization test (Amm/Gen- and Spirals/Gen-).

Subject	Ammonites and Spirals Tests					
	Amm	SD	Spirals	SD	Amm /Gen-	Spirals /Gen-
Anatol	0.723	1.139	0.654	0.817	0.098	0.248
Hannes	0.693	0.898	0.852	1.201	0.112	0.158
Lukas	0.855	0.629	1.055	0.932	0.149	0.097
Niki	0.871	0.953	1.071	1.476	0.482	0.985
Olivia	0.945	0.798	1.133	0.917	0.176	<b>0.024</b>
Rhonda	0.48	0.888	0.299	0.8	0.244	<b>0.017</b>
<i>Mean(±SD)</i>	<i>0.761(0.884)</i>		<i>0.844(1.024)</i>			
<i>Wilcoxon</i>					<i>0.675</i>	<i>0.402</i>

Note. SD = standard deviation, Gen+ = positive generalization stimuli, Gen- = negative generalization stimuli. Significant results are denoted with a bold font. The cursive font identifies group level statistics.

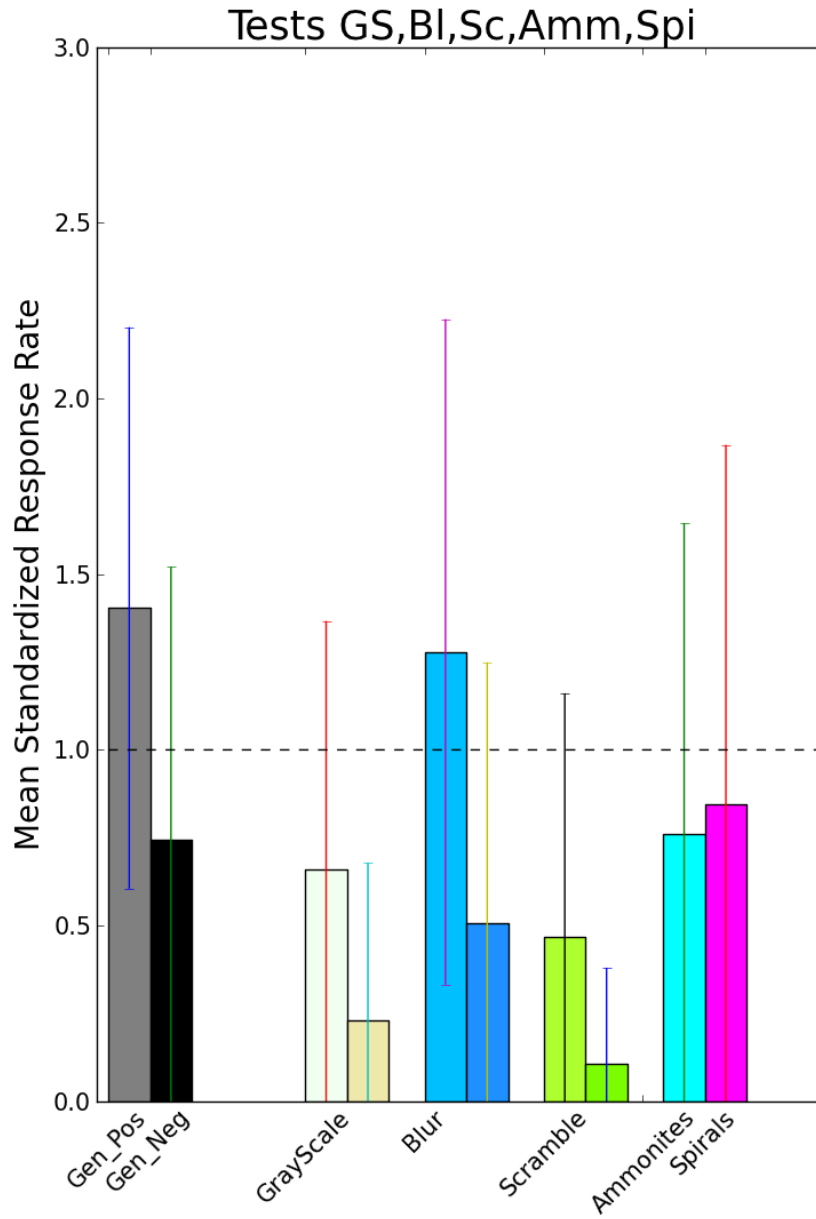


Figure 3.6: Results of Tests Grayscale, Blur, Scramble, Ammonites, and Spirals. The height of the bars represents the average response rates ( $\pm$ SD) of all animals in the group. The horizontal dashed line indicates the average peck rate. Higher peck rates indicate that the respective picture were regarded as positives rather than as negatives (and vice versa for lower peck rates). Gen\_Pos = Positive Generalization Test image, Gen\_Neg = Negative Generalization Test images. Each test has its unique color: Grayscale (white), Blur (blue), Scramble (green), Ammonites (cyan), Spirals (magenta). The left bar of tests GS, Bl and Sc represents the positive stimuli and the right bar the negative.



# Chapter 4

## Discussion

Many studies [4, 39, 40, 41, 42, 43, 44, 45] have tried to assess the relation between objects and the images representing them by using a variety of different designs. The common motivation behind these designs has been the need to prevent the subject's use of more simple strategies than representational insight, namely, purely perceptual ones. Whilst blocking one type of transfer that strengthens the belief in the use of true abstract-like cognitive processes, others can slip the researchers' mind. The results of a study presented by Aust & Huber [1] has yielded support for the view that representational insight is present in pigeons. In a follow up study [2], with pigeons that had never seen real human heads prior to testing it was confirmed that real life experience was indeed the critical factor in transfer from pictures of incomplete human figures to pictures of the complementary parts. To strengthen the case of familiarity dependence of POR, Grabner [3] conducted a study with pictures of snails, i.e., a natural class of objects the pigeons had no real life experience with and which was thus unfamiliar. Thereby this experiment differed from the experiment with human figures, a category that the pigeons were familiar with. Overall, the findings seemed, indeed, to point in this direction. However, the small number of subjects that could eventually be tested (due to insufficient learning of the initial task) and large performance differences among the successful birds prevented any strong conclusions. The present study was aimed

at strengthening, complementing, and extending the findings by Aust & Huber and by Grabner.

A total of fourteen pigeons participated in eight different tests. Thereby, two goals were pursued. First, the sample size of Grabner's experiment was increased by subjecting two further birds to the POR-test with complementary parts. The second main test of this thesis, the Head Vs Shell Vs Skin Test, aimed at the question whether different parts of a snail are equally relevant for categorization. Based on the results by Grabner I expected snail shells to be particularly salient. A series of follow-up tests investigated the perceptual features used for categorization. The color (Grayscale Test), texture (Blur Test) and shape (Scramble Test) of the stimuli were systematically modified. Additionally, artificial snail-like stimuli were created. This was the basis for tests Ammonites and Spirals.

As already noted by Grabner [3], the two animal groups No Head and No Shell differed significantly in the duration of the training. She attributed this difference mainly to the easily recognized perceptual feature of the shell, namely the helix, or the spiral. Should the difference between the snail groups be based purely on the shape of the shell, then the Group Full Snails should take as long as the No Head group. Yet the group that had access to both, heads and shells was significantly faster than any other group. As we can see, it seems that the snail category is defined by a polymorphous rule. Jitsumori [29] pointed out that single categorization cues are neither necessary nor sufficient. The absence of a particular part of the target figure, such as the shell or the head, did not prevent categorization of a picture as snail. Hence, neither part is indispensable for learning the snail/non-snail discrimination. However, if both are present, learning is accelerated. In [54] authors pointed out the contribution of various human parts (heads, hands, clothes, size, texture) to categorization as well as their additive effects, and thereby showed the polymorphous character of the class rule.

The Picture-Object Recognition test was based on the Complementary Information

Procedure by Aust&Huber [1]. No differences in responding to Heads, Shells and patches of Skin were revealed. The only significant result was the difference between Seen Parts and Unseen Parts by the No Shell animal Josef. In other experiments where the animals were familiar with the tested concept such a result did not occur. Tests with animals trained on familiar humans by [1] showed that the subjects treated SP and UP indifferently. Interestingly, to strengthen the hypothesis of a familiarity effect we would expect the same significant difference between SP and UP such as like Josef's in all snail animals. I would therefore assume that both SP and UP shared a set of category defining features and responding to them was therefore similar. On the other hand, Josef was the only successful animal trained on snail bodies devoid of the Shell. This particular body part has hardly any features common with the body of the snail or the Skin whereas the Head does. The animals trained on snail bodies and snail shells had seen the texture of the snail Head and Skin before testing and therefore showed evidence of transfer on test. Both previously unseen Heads and patches of snail Skin shared some of their features with the trunk of the snail that was seen as part of the training. It is therefore plausible that these pictures elicited higher response rates than those emitted to the negative stimuli.

The Head vs Shell vs Skin Test tested the relative importance of different parts of a snail figure. Based on the results, we see that only the Shell was treated in the same way as novel full snail images. The set of visual features provided by a Shell is thus sufficient for categorizing an image as member of the snail category. Both heads and skin patches were not treated differently from the negative image class and therefore seem to lack many of the category defining features. The results of this test extend the findings of the POR test and the difference between Josef and the other animals. Responding to Shells was significantly different from responding to Heads (SH:HE) and a respective tendency was found for also for Skin (SH:SK). An animal trained on snail bodies including heads will lack the ability to classify shells as these differ strongly

from the rest of the snail in terms of perceptual features. Besides the immixture of the negative stimuli with Heads and Skin Patches, the latter showed no difference among each other. This explains why I found no difference in the response rates of the NoHead animals. Interestingly all animals from Full Snails group pecked less to heads than to non-snail (NS) images. None of the images was seen before but there cannot be ruled out any kind of similarity to the learning set.

Three additional tests modified the informational content of the images to assess the effects of these manipulations on categorization ability. The first test reduced color information (Grayscale Test). In a set of experiments, Aust & Huber [50] presented pigeons of two groups with grayscaled images. The group trained on humans as positives retained its discrimination ability, while the group trained on non-humans as positives failed. My Grayscale Test yields similarity to the human positive group as subjects were trained on snail positive images. Five out of six animals tested in the present study were able to discriminate snails from non-snails in grayscale images at a significant or close to significant level. Because my study had no snail negative group I cannot make a full connection to the study by Aust and Huber. The use of color cues is not only bound to the degree of familiarity, that is the main difference between the two studies, but also depends on the type of stimuli used. Watanabe showed that pigeons used color cues while categorizing good vs bad drawings [55] and not while categorizing paintings by Monet and Picasso [56]. Color, as well as other perceptual cues, is only used if it is a good category predictor. Should it not be the case, it is ignored [31] and other cues take precedence. Of course, the natural category snail is not only based on color but also on other visual features. Color of both the snail skin and the shell can vary depending on lighting conditions or based on species differences. I therefore assumed that while presented with grayscale stimuli the subjects must have used other visual cues for categorization and therefore retained their categorization ability.

Pigeons' preference for local features over global configuration is a widely discussed

topic. The results from the tests Blur and Scramble do not point to a specific direction. In the work by Aust & Huber [54], the authors reduced the amount of local features present in a stimulus class. They discarded most of the local cues including shading, lightning, color and texture. Their subjects did not treat human silhouettes differently from a human-absent image. An opposite result was found in the Blur Test presented as part of this work. Biederman [20] claimed that "[...] the silhouette of an object is sufficient for successful classification, as long as it clearly depicts the object's parts." My results, even through using a slightly different method, are in line with the theory. The influential papers by Biederman [20] and Cerella [19], having opposite opinions on the importance of local vs global features, both used scrambled *drawings*. On the contrary, Watanabe [51] used photographs that were pictorial representations of real objects instead of cartoons. In this scenario the recognition rate of scrambled stimuli dropped significantly. I would not share the author's conclusion that "[...] pigeons can see photographs as representations of real objects, as proved by bidirectional transfer of discrimination from real objects to their photographs [...]" because a result as obtained by Watanabe can be achieved also on the associative level of POR and requires no true abstract understanding. On the other hand, his findings highlighting the failure of pigeons to form a connection with the cartoons' real counterpart while presented with simple line drawings are of relevance for this study. This could explain why recognition of Cerella's drawings remained intact with scrambled stimuli. Real photographs, used in other experiments, suffer from scrambling. This conclusion poses a problem to my connection to the study by Beiderman. Because of the differences to both studies by Aust & Huber [54] and Biederman [20] I cannot conclude an effect of familiarity or importance of local features. An blur study on a "familiar" group, preferably humans, should be conducted. The difference between the results of a "familiar" an a "unfamiliar" group would resolve the importance of local features. The test that examined the use of global features was the Scrambling Test. The type of scrambling used in the cur-

rent experiment was denoted as fragmentation by Matsukawa [49]. In his experiment, pigeons were presented with scrambled faces of cartoon characters. At a scrambling degree of 6x6 the recognition rate has fallen significantly. The author does not explain the recognition rate between positive and negative stimuli and only specifies it as "somewhat different". As explained above the difference between photographs and drawings causes a problem with reference. The same type of scrambling was used by Aust & Huber [50] who modified color pictures containing humans as stimuli. Their subjects were able to recognize humans from non-humans even in strongly scrambled images. In the current experiment, scrambled stimuli were recognized at a significant level, despite strong drops in discrimination performance. Only four out of six animals retained their discrimination ability. In particular, peck rates to scrambled negative stimuli were extremely low. One of the possible explanations for the difference between the current study and the studies by Aust & Huber is the degree of familiarity with the concept. By using familiar, semi-familiar and novel stimuli, the authors showed that the more the subject is familiar with the category the lower the effect of scrambling. However, at the 6x6 scrambling degree the performance was strongly impaired in both experiments making any comparisons futile. The explanation, why some of the animals were able to categorize correctly, can be the use of local textural cues that remained intact. Even through the categorization ability in both Blur and Scramble Tests was impaired in some pigeons, others retained their ability. The category is not dependent on any single local or global cue but rather by a combination of both. Experiments by Goto [57], Fremouw [23, 24], Cavoto and Cook [58], Aust and Huber [50, 25] support the hypothesis that pigeons can use both local and global cues and switch between these two if required. Picture recognition is not based on a single perceptual cue but rather on a set of distinct membership criteria.

The study by Aust & Huber [54] showed that pictures of animals wearing clothes were categorized as belonging to the human category, presumably because many of

the category-defining features (including clothes) were present. Their results showed that the inclusion principle of pigeons is more open than that of human subjects. On the contrary, the pseudo-snail category Ammonites presented in this work was treated differently like negative images. The snail category seems thus to be more narrow and less inclusive than the human category. Following the argumentation of Medin [59], there might be stimulus types that are in possession of a lot of category defining features yet their number is not high enough to categorize the stimulus as belonging to a particular category (as was the case with ammonites). However, one has to be careful with the interpretation of these results as they are closely bound to image selection. The backgrounds on which the ammonites were presented were different from the ones presented with snails. The body of the ammonites was missing and from the previous tests we know that the non-shell parts also contributed to categorization. In addition, many of the images only contained the contour of the fossilized ammonite or the shell was cut-open. In summary, all these differences may have resulted in ammonites being categorized as non-snail. The same may apply to the other pseudo-snail category, namely Spirals. These helix-like approximations were treated similarly as Ammonites, namely, as non-snails. This is, despite possessing some snail-like features they were not included in the snail category.

The four tests presented above investigated the contribution of head, shell, skin patches, texture, color and global composition to pigeons' categorization of snails. None of the tested features yielded performance as good as found for unmodified full snails. The highest response rates were recorded for Shells in the HSS Test and inversely (their removal caused the most drastic decrease) for color in the Grayscale and global composition in the Scrambling Test. Heads, Skin Patches or the presence of local features (removed in the Blur Test) resulted in no snail positive categorization. Yet their contribution to the overall categorization ability is indisputable. According to Aust & Huber [54], all parts of a human figure are relevant for categorization but their

relative contributions vary. Clothing, heads and torsos were good predictors of the presence of a human just like the Shell was a good predictor of snails in the current study.

This study aimed at exploring the reasons for the striking differences in performance found between the two animals tested in the POR Test by Grabner [3]. I was able to show that the results of Daisy were an outlier and that no other animal showed a difference between response rates towards Unseen Parts and Skin Patches. This supports the claim that familiarity with a category is a prerequisite for POR at a level beyond mere feature discrimination, as was shown by Aust & Huber [1] by means of a Complementary Information Procedure. Parts of objects that are not present in the world experienced by the subjects cannot be recognized in pictures as belonging together. Still, to further strengthen the the notion of category familiarity more than four animals will be required. Based on the results of the HSS-Test it can be concluded that the Shell of a snail is more important for categorization than the Head. There were differences in performance between the NoHead and the NoShell groups due to the parts of a snail that were present in (or missing from) their training stimuli. The findings of the Grayscale, Blur, Scramble, Spirals and Ammonites tests confirm the assumption about the use of multiple types of cues for object categorization. The ability of pigeons to flexibly use different perceptual cues makes it hard to unequivocally demonstrate the ability of true representational insight. Still I believe that methods such as the CIP may contribute to solving this problem.

More research is needed in the field of Picture-Object recognition in pigeons. The present study makes a contribution to the topic by assessing the role of category familiarity and the influence of a variety of perceptual features on categorization. Nevertheless, I believe that more animals are required to make strong claims. Also a study on the effects of scrambling and blurring of familiar versus unfamiliar concepts would shed further light on the pigeons' perception of the world.



# Acknowledgements

First of all, I would like to thank my supervisor Dr. Ulrike Aust who introduced me into the field of Cognitive Biology. She offered me an opportunity to work with the pigeons and enabled me to study their mesmerizing perceptual abilities. Next, I would like to thank Prof. Ludwig Huber whose lectures deepened my knowledge in all subfields of Cognitive Biology. Also, he lent me many interesting books that I enjoyed reading. Not to be forgotten are the people from the Department of Cognitive Biology who helped me with evaluation, statistical methods or with the creation of the certain something present in the work, namely Gesche Westphal Fitch, Mauricio Martins, Markus Böckle and Piera Filippi.

Thank you all!

# References

- [1] U. Aust and L. Huber, “Picture-object recognition in pigeons: Evidence of representational insight in a visual categorization task using a complementary information procedure,” *Animal Behavior Processes, Vol. 32, No. 2, 190195*, 2006.
- [2] U. Aust and L. Huber, “Representational insight in pigeons: comparing subjects with and without real-life experience,” *Animal Cognition, 13:207-218*, 2010.
- [3] V. Grabner, “Influence of familiarity with a category on picture-object recognition in pigeons (columba livia),” 2010.
- [4] R. J. Herrnstein and D. H. Loveland, “Complex visual concept in the pigeon,” *Science, 146, 549-551*, 1964.
- [5] R. G. Weisman and M. L. Spetch, “Determining when birds perceive correspondence between pictures and objects: A critique,” *Comparative Cognition & Behavior Reviews, Volume 5, pp 117-131*, 2008.
- [6] T. G. Barraclough, P. H. Harvey, and S. Nee, “Sexual selection and taxonomic diversity in passerine birds,” *Proc. R. Soc. Lond. B, 259, pp 211-215*, 1995.
- [7] M. D. Eaton, “Human vision fails to distinguish widespread sexual dichromatism among sexually ‘monochromatic’ birds,” *Proc. of the National Academy of Science (USA), 102, 10942-10946*, 2005.
- [8] A. Ödeen and O. Hästad, “Complex distribution of avian color vision systems

- revealed by sequencing the sws1 opsin from total dna,” *Molecular Biology and Evolution* 20(6):855-861, 2003.
- [9] D. Osorio and M. Vorobyev, “Photoreceptor spectral sensitivities in terrestrial animals: adaptations for luminance and colour vision,” *Proceedings of the Royal Society B*, 272, 1745-1752, 2005.
- [10] D. Osorio and M. Vorobyev, “Tetrachromacy, oil droplets and bird plumage colours,” *Journal of Computational Physiology A*, 183: 621-633, 1998.
- [11] J. D. Delius, J. Emmerton, W. Hörster, R. Jäger, and J. Ostheim, “Picture-object recognition in pigeons,” *In: Picture perception in animals, East Sussex Psychology Press*, pp. 1-36, 2000.
- [12] R. F. Oliveira, “Considerations on the use of video playbacks as visual stimuli: the lisbon workshop consensus,” *Acta Ethol*, 3:61-65, 2000.
- [13] R. D’Eath, “Can video images imitate real stimuli in animal behaviour experiments,” *Biol Rev* 73:267-292, 1998.
- [14] J. Zeil, “Depth cues, behavioural context, and natural illumination: some potential limitations of video playback techniques,” *Acta Ethol* 3:39-48, 2000.
- [15] D. Bovet and J. Vauclair, “Picture recognition in animals and humans,” *Behavioural Brain Research*, 109, 143-165, 1999.
- [16] M. Tomasello, J. Call, and A. Gluckman, “Comprehension of novel communicative signs by apes and human children,” *Child development*, 68, 1067-1080, 1997.
- [17] R. Miller, “Cross-cultural research in the perception of pictorial materials,” *Psychological Bulletin*, 80, 135-150., 1973.
- [18] M. Herskovits, “Art and value,” *In: Redfield R, Herskovits M, Ekholm GF, editors. Aspects of Primitive Arts. New York: Museum of Primitive Art*, 1959.

- [19] J. Cerella, "Pigeons and perceptrons," *Pattern Recognition*, 19, 431-438, 1986.
- [20] I. Biederman, "Recognition-by-components: A theory of human image understanding," *Psychological Review*, 94, 115-147, 1987.
- [21] R. G. Cook, *Avian Visual Cognition*. Comparative Cognition Press, 2001.
- [22] C. P. Shimp, W. T. Herbranson, and T. Fremouw, "Avian visual attention in science and culture," *In R. G. Cook (Ed.), Avian visual cognition*, 2001.
- [23] T. Fremouw, W. T. Herbranson, and C. P. Shimp, "Dynamic shifts of pigeon local/global attention," *Anim. Cogn*, 5, pp 233-243, 2002.
- [24] T. Fremouw, W. T. Herbranson, and C. P. Shimp, "Priming of attention to local or global levels of visual analysis," *Journal of Experimental Psychology: Animal Behavior Processes*, 24(3), pp 278-290, 1988.
- [25] U. Aust and L. Huber, "Elemental versus configural perception in a people-present /people-absent discrimination task by pigeons," *Learning & Behavior*, 31 (3), 213-224, 2003.
- [26] S. Watanabe, "How do pigeons see pictures? recognition of the real world from its 2-d representation," *In: Picture perception in animals, East Sussex Psychology Press*, pp. 71-90, 2000.
- [27] R. J. Herrnstein, "Acquisition, generalization, and discrimination reversal of a natural concept," *Journal of Experimental Psychology: Animal Behavior Processes*, 5, 116-129., 1979.
- [28] R. J. Herrnstein, "Levels of stimulus control: A functional approach," *Cognition*, 37, 133-166, 1990.
- [29] M. Jitsumori, "Category discrimination of artificial polymorphous stimuli based on feature learning," *Journal of Experimental Psychology: Animal Behavior Pro-*

cesses, 19, 244-254, 1993.

- [30] M. Jitsumori, "A prototype effect and categorization of artificial polymorphous stimuli in pigeons," *Journal of Experimental Psychology: Animal Behavior Processes*, 22, 405-419, 1996.
- [31] L. Huber, "Visual categorization in pigeons," In R. G. Cook (Ed.), *Avian visual cognition*, 2001.
- [32] L. Huber and R. Lenz, "Categorization of prototypical stimulus classes by pigeons," *The quarterly journal of experimental psychology*, 49B (2), 111-133, 1996.
- [33] M. D'Amato and P. Van Sant, "The person concept in monkeys (cebus apella)," *Journal of Experimental Psychology: Animal Behavior Processes*, 5, 14, 43-55, 1988.
- [34] J. Fagot, J. Martin-Malivel, and D. Dépy, "What is the evidence for an equivalence between objects and picture in birds and nonhuman primates," In: *Picture perception in animals*, East Sussex Psychology Press, pp. 295-320, 2000.
- [35] J. S. DeLoache, "Early symbolic understanding and use," In: D. Medin (Ed.), *The Psychology of learning and motivation*, Academic Press, New York, vol.33, pp. 65-114, 1995.
- [36] R. G. Cook, D. G. Levison, S. Gillett, and A. P. Blaisdell, "Capacity and limits of associative memory in pigeons," *Psychonomic Bulletin and Review*, 12, 350-358, 2005.
- [37] J. Fagot and R. G. Cook, "Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition," *Proceedings of the National Academy of Science of the United States of America*, vol. 103, no. 46, pp. 17564- 17567, 2006.

- [38] J. Fagot, *Picture perception in animals*. East Sussex, Psychology Press, 2000.
- [39] J. D. Delius, "Categorical discrimination of objects and pictures by pigeons," *Animal Learning & Behavior*, 20, 301-311, 1992.
- [40] S. Watanabe, "Object-picture equivalence in the pigeon: An analysis with natural concept and pseudoconcept discriminations," *Behavioral Processes*, 30, 225-232, 1993.
- [41] P. A. Cabe, "Transfer of discrimination from solid objects to pictures by pigeons: a test of theoretical models of pictorial perception," *Perception & Psychophysics*, 19, 545-550., 1976.
- [42] T. Shimizu, "Conspecific recognition in pigeons (*columba livia*) using dynamic video images," *Behaviour*, 135, 43-53, 1998.
- [43] M. Vandenheede and M. F. Bouissou, "Fear reactions of domestic sheep confronted with either a human or a human-like model," *Behavioural Processes*, 34, 81-92, 1995.
- [44] D. L. Clark and G. W. Uetz, "Video image recognition by the jumping spider, *maevia inclemens* (araneae: salticidae)," *Animal Behaviour*, 40, 884-89, 1990.
- [45] E. Winner and G. Ettlinger, "Do chimpanzees recognize photographs as representations of objects?," *Neuropsychologia*, 18, 413-420, 1979.
- [46] R. K. Davenport, C. M. Rogers, and I. S. Russell, "Cross-modal perception in apes: Altered visual cues and delay," *Neuropsychology*, 13, pp. 229-35, 1975.
- [47] S. Watanabe, "An instance of viewpoint consistency in pigeon object recognition," *Behavioral Processes*, 39, 257-261, 1997.
- [48] D. M. Wilkie, R. J. Willson, and S. Kardal, "Pigeons discriminate pictures of a geographic location," *Animal Learning & Behavior*, 17 (2), 163-171, 1989.

- [49] A. Matsukawa, S. Inoue, and M. Jitsumori, "Pigeon's recognition of cartoons: effects of fragmentation, scrambling and deletion of elements.," *Behavioral Processes*, 65, pp 25-34, 2004.
- [50] U. Aust and L. Huber, "The role of item- and category-specific information in the discrimination of people versus nonpeople images by pigeons," *Animal Learning & Behavior*, 29 (2), 107-119, 2001.
- [51] S. Watanabe, "Discrimination of cartoons and photographs in pigeons: effects of scrambling of elements," *Behavioral Processes, Volume 53*, pp 3-9, 2001.
- [52] N. F. Troje, L. Huber, and U. Aust, "Categorical learning in pigeons: The role of texture and shape in complex static stimuli," ., 1999.
- [53] R. J. Herrnstein, D. H. Loveland, and C. Cable, "Natural concepts in pigeons," *Journal of Experimental Psychology: Animal Behavior Processes*, 2, 285-311., 1976.
- [54] U. Aust and L. Huber, "Target-defining features in a 'people-present / people-absent' discrimination task by pigeons," *Animal Learning & Behavior*, 30 (2), 165-176, 2002.
- [55] S. Watanabe, "Pigeons can discriminate "good" and "bad" paintings by children," *Animal Cognition*, 13, 75-85, DOI 10.1007/s10071-009-0246-8, 2010.
- [56] S. Watanabe, J. Sakamoto, and M. Wakita, "Pigeons' discrimination of paintings by monet and picasso," *Journal of Experimental Analysis of Behavior*, 63, pp 165-174, 1995.
- [57] K. Goto, A. J. Wills, and S. E. G. Lea, "Global-feature classification can be acquired more rapidly than local-feature classification in both humans an pigeons," *Anim Cogn*, 7, pp 190-113, 2004.

- [58] K. K. Cavoto and R. G. Cook, "Cognitive precedence for local information in hierarchical stimulus processing by pigeons," *Journal of Experimental Psychology: Animal Behavior Processes*, 27(1), pp 3-16, 2001.
- [59] D. L. Medin, "Concepts and conceptual structure.," *American Psychologist*, 44, 1469-1481, 1989.



# Appendix A

## Appendix

### A.1 Declaration

Herewith I ensure that I wrote this Master's Thesis independently using only the indicated sources and means for writing. For individual parts inferred from other publications, literally or according to the sense, references have been made within the text. The same applies to pictures, tables and figures.

Hiermit erkläre ich, dass ich diese Masterarbeit selbstständig und nur unter Verwendung angegebener Quellen und Hilfsmittel verfasst habe. Für einzelne Teile, die im Sinne oder Wortlaut aus anderen Publikationen entnommen wurden, sind im Text dementsprechende Referenzen angegeben. Dasselbe gilt auch für Bilder, Tabellen und Grafiken.

## A.2 Stimuli

THE STIMULI APPENDIX IS NOT PART OF THIS VERSION

## A.3 Summary

The ability to perceive images and understand their abstract character was thought to be common for many species. Since the study by Herrnstein and Loveland [4] researchers have been investigating the pigeon’s ability to categorize images based on specific features or perceptual rules. It is known that pigeons readily categorize instances of humans, bodies of waters, trees etc. and that they transfer this rule to previously unseen images. However, this successful generalization to novel stimuli does not imply their ability to perceive the abstract relationship between the image and the real world. I examined this ability that is called *picture-object recognition* (POR) at the level of representational insight.

One of the methods used for examining POR is the *Complementary Information Procedure* (CIP) by Aust & Huber [1], which disentangles true representational insight and simpler mechanisms like the recognition of simple 2D-features. Their study used pictures of humans, a category that the subjects were familiar with. Indeed, the results suggested POR at an abstract level. A control condition showed that this transfer was not based on simple textural cues. A follow-up study by Grabner [47] showed a lack of representational insight in the absence of real life experience by testing pigeons with pictures of snails, an unfamiliar natural class.

As part of this thesis I tested two additional animals using the CIP and thereby strengthened the notion of a familiarity effect. Subjects were trained to discriminate between images containing one or multiple parts of a snail (S+) and pictures showing something else (S-). Then they were tested with pictures of the previously missing parts as well as with pictures of arbitrary patches of snail skin (i.e., non-representative stimuli). The pigeons treated Unseen Parts (UP) like non-representative skin patches (SK), while parts that had also been present in the training stimuli (Seen Parts; SP) yielded significantly higher response rates. Together, the results indicated a lack of representational insight, as would indeed be predicted for an unfamiliar category. A second line

of research examined the relative importance of various perceptual features of snails for categorization. Therefore, animals trained on complete snails were subsequently tested with pictures of isolated parts. Only the shell was regarded as belonging to the snail category, while the head and arbitrary non-representative skin patches were considered as non-snail. A set of follow-up tests that varied the pictures' informational content revealed the importance of the global image configuration. Scrambling, a method that reduces the amount of global cues, caused the strongest drop in performance. By contrast, blurred images were treated similarly as the unmodified originals. Removal of color led to significant performance decrements regarding discrimination between positives and negatives. The inclusiveness of the snail category was tested with pictures of Ammonites and Spirals. Actually, both types of stimuli were classified as non-snails.

In summary, the unfamiliar snail category seems to represent a collection of otherwise meaningless visual features that can be combined into a polymorphous response rule. However, the correspondence with objects in the real world does not seem to be understood. Thereby, the results of this thesis support the claim that pigeons are able to recognize the abstract relationship between objects and pictures, however, only if they are familiar with the tested category.

## A.4 Zusammenfassung

Im Allgemeinen wird davon ausgegangen, dass alle Spezies in der Lage sind, Bilder wahrzunehmen und ihre abstrakte Bedeutung zu erfassen. Seit der Studie von Herrnstein und Loveland [4] wird erforscht, nach welchen spezifischen Regeln Tauben Bilder kategorisieren. Es ist bekannt, dass Tauben Menschen, Bäume und Wasserkörpern kategorisieren und die verwendeten Regeln auf neue Objekte übertragen können. Diese erfolgreiche Generalisierung sagt jedoch nichts über die Fähigkeit aus, den Zusammenhang zwischen dem Bild und der realen Welt zu erkennen. Diese Fähigkeit wird *Bild-Objekt-Erkennung* (Picture-object recognition, POR) genannt und ist Thema meiner vorliegenden Masterarbeit.

Eine der Methoden, um POR zu untersuchen, ist die *Complementary Information Procedure* (CIP) von Aust & Huber [1]. Die Versuchstiere werden dabei daran gehindert, einfachere wahrnehmungsbezogene Strategien anzuwenden, wie das Erkennen einfacher 2D-Merkmale. Die Studie von Aust & Huber verwendete Bilder von Menschen. Die Versuchstiere waren mit dem Konzept Mensch vertraut. Die Ergebnisse stützten die Annahme, dass POR an einer abstrakten Ebene stattfand. Die Kontrollkondition zeigte, dass keine einfachen texturellen Merkmale benutzt worden sind. Eine Folgestudie von Grabner [3] zeigte einen Mangel an Einsicht in den Repräsentationscharakter von Bildern einer unbekannten natürlichen Klasse.

Im Rahmen meiner Masterarbeit konnte ich zwei zusätzliche Tiere mittels CIP testen, was die Annahme eines Bekanntheitseffektes unterstützte. Die Versuchstiere wurden darauf trainiert, zwischen Fotos von Schnecken (S+) und Bildern, die etwas anderes zeigten (S-), zu unterscheiden. Danach wurden sie mit zuvor nicht gesehenen Schneckenanteilen und mit nicht-repräsentativen Hautausschnitten trainiert. Die Tiere behandelten zuvor nicht gesehene Teile (Unseen Parts, UP) ähnlich wie nicht-repräsentative Hautausschnitte (Skin Patches, SK) wobei Teile die auch im Training zu sehen waren (Seen Parts, SP) signifikant höhere Pickraten aufwiesen. Die Ergebnisse zeigten auf eine

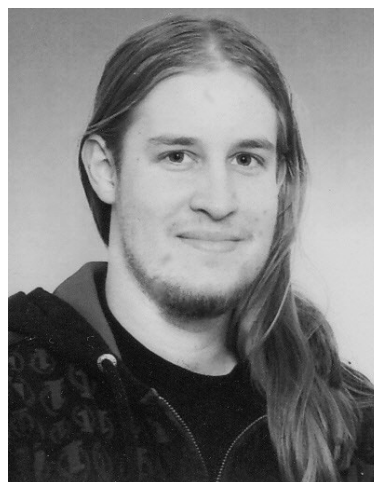
mangelnde Fähigkeit hin, die komplementären Teile der Schnecke zu erkennen, genau wie man von einer "nicht vertrauten" Klasse erwarten würde. Im zweiten Experiment habe ich die relative Wichtigkeit der einzelnen Schneckenmerkmale die zur Klassifizierung beitragen untersucht. Deswegen wurden Tauben mit Bildern von vollständigen Schnecken trainiert und danach mit Bildern von isolierten Schneckenteilen getestet. Nur das Schneckenhaus wurde als zur Schneckenkategorie gehörig eingestuft. Der Kopf und beliebige nicht-repräsentative Hautausschnitte wurden eher als Nicht-Schnecken klassifiziert. Mehrere Folgetests, die den Informationsgehalt der Bilder systematisch variierten, zeigten die Bedeutung der globalen Bildkonfiguration. Scrambling, eine Methode, welche die globale Konfiguration zerstört, führte zu den stärksten Leistungsabfällen. Demgegenüber unterschieden sich die Antwortraten bezüglich verschwommener Bilder, in denen lokale Merkmale zerstört waren, nicht von den unmodifizierten Originalen. Das Entfernen von Farbinformation beeinträchtigte die Unterscheidung zwischen positiven und negativen Bildern. Die Grenzen der Kategorie Schnecke wurden mittels Bildern von Ammoniten und Spiralen untersucht. In der Tat wurden beide Bildtypen als Nicht-Schnecken klassifiziert.

Die unbekannte Kategorie der Schnecken, scheint als eine Sammlung perzeptueller Merkmale wahrgenommen werden, die in eine polymorphe Klassenregel integriert werden können, ansonsten aber bedeutungslos sind. Die Verbindung zwischen den Bildern und der realen Welt bleibt dabei unverstanden. Die Ergebnisse der vorliegenden Arbeit unterstützen Annahme, dass Tauben die abstrakte Beziehung zwischen Bildern und echten Objekten wahrnehmen können, aber nur wenn sie mit der Klasse "vertraut" sind.

# Cirriculum Vitae

## Personal Data:

Name: KYSEL Martin  
Date of Birth: 27<sup>th</sup> July 1988  
Address Chrobakova 6  
84102 Bratislava  
SVK  
Email Martin.Kysel@vegi.sk  
Nationality Slovak



## Education:

1999 – 2006      Gymnasium Bilikova 24. Bilingual high school

2006 – 2009      Undergraduate Studies: Media Informatics, BSc  
Department of Informatics, University of Vienna, Vienna, AUT

Bachelor Thesis entitled: *"Wiideo Chat - Wii-Remote Controlled SIP-based Video Chat"*  
Referee: Univ. Prof. Helmut Hlavacs

2009 - 2011      Graduate Studies: Joint Degree MEI:CogSci, MSc  
Department for Philosophy, University of Vienna, Vienna, AUT

Master Thesis:  
Department for Cognitive Biology, University of Vienna,  
Vienna, AUT  
Thesis entitled: *„Picture-object recognition in pigeons: The use of visual features for classifying pictures of an unfamiliar natural category“*  
Referee: Dr. Ulrike Aust