universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Inference by exclusion and alternative strategies in carrion crows (*Corvus corone*)"

verfasst von / submitted by

Theresa Rößler, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2016 / Vienna 2016

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 878
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Verhaltens-, Neuro- und Kognitionsbiologie
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Thomas Bugnyar

ACKNOWLEDGEMENTS

I would like to thank everybody who made this work possible. Special thanks go to Thomas Bugnyar for supervision as well as Ira Federspiel and Mark O´Hara for great and enduring support. Many thanks also to the whole KLF Team, who were always there to help along the way and with whom I enjoyed working very much.

TABLE OF CONTENTS

ABSTRACT 4

ZUSAMMENFASSUNG 5

INTRODUCTION 7

METHODS 11

Study animals and housing 11

Experiment 1 13

Experimental Set-Up 13

Procedure 13

Data collection and analysis 17

Experiment 2 20

Experimental Set-Up 20

Procedure 20

Data collection and analysis 23

RESULTS 24

Experiment 1 24

Experiment 2 31

DISCUSSION 32

Experiment 1 32

Experiment 2 37

General discussion 39

Future ideas 40

REFERENCES 41

LIST OF FIGURES AND TABLES 47

APPENDIX 49

ABSTRACT

Inference by exclusion, hereafter called ‘IBE’, is the ability to base a choice on the exclusion of alternatives. Considered as a ‘higher’ cognitive ability that is typical for humans, IBE has also been investigated in a variety of animal species. However, most studies have used tasks that allow addressing only a binary question, whether or not IBE occurs. Recently two parrot species, Goffin cockatoos and keas (O´Hara et al. 2015; 2016), have been tested with a holistic approach that investigates IBE and possible alternatives underlying the decision making process in IBE tasks simultaneously. The aim of this study is to use this new holistic approach and investigate response strategies of the carrion crow, as a representative of the corvid family, in such an IBE task.

Crows are renowned for solving cognitively demanding tasks and approaching novel situations with curiosity and caution. A previous study showed the ability to infer by exclusion in this species but alternative strategies such as avoidance could not be ruled out with certainty. In the current study the birds learned to associate stimuli with positive and negative contingencies in a training phase through a two-choice procedure presented on a touch screen computer. I then tested whether they could logically follow the line of contingencies when those stimuli were subsequently replaced by novel stimuli. I also conducted a second experiment in which this series of tests was transferred to a more ecologically relevant food finding task.

In the touch screen experiment, high levels of IBE were observed in one individual crow while others seemed to rely mostly on neophilic strategies. Despite of methodological problems in the second experiment, the opposite trend seemed to emerge. I discuss how those seemingly contrasting reactions are compatible and how a species predisposition has to be carefully considered when testing for cognitive abilities such as IBE, e.g. through thorough habituation to novelty.

ZUSAMMENFASSUNG

Die Fähigkeit, logische Schlüsse durch Ausschließen alternativer Möglichkeiten zu ziehen, wird in der Fachliteratur mit dem englischen Begriff „inference by exclusion“ beschrieben und folgend verkürzt „IBE“ genannt. Diese Fähigkeit, die als kognitiv anspruchsvoll und für Menschen typisch gilt, wurde in einer Vielzahl verschiedenster Tierarten untersucht. Meist wurde jedoch vorrangig auf eine binäre Fragestellung fokussiert, also ob IBE genutzt wird oder nicht, als auf alternative Mechanismen einzugehen. Zwei kürzlich veröffentlichte Studien, die IBE in zwei Papageienarten, den Goffin Kakadus und Keas (O´Hara et al. 2015; 2016) untersuchten, widmeten sich einer neuen, holistischen Herangehensweise. Dabei wurde nicht nur behandelt ob IBE auftritt oder nicht, sondern gleichzeitig auch auf andere Mechanismen, die dem Entscheidungsprozess zugrunde liegen könnten, eingegangen. Das Ziel der vorliegenden Studie ist mittels dieser Methodik sowohl IBE als auch alternative Strategien der Aaskrähe zu untersuchen.

Krähen sind dafür bekannt verschiedenste, kognitiv anspruchsvolle Aufgaben zu lösen und neuen Situationen mit Neugierde und Vorsicht zu begegnen. In einer vorangehenden Studie wurde gezeigt, dass Aaskrähen möglicherweise die Fähigkeit IBE besitzen und verwenden, jedoch konnten alternative Strategien nicht mit Sicherheit ausgeschlossen werden. In einem ersten Experiment der vorliegenden Studie lernten die Vögel mit Hilfe eines Touchscreen Computers einzelnen Bildern positive und negative Wertungen zuzuschreiben. Darauffolgend wurden diese Bilder nach und nach durch unbekannte, neue Bilder ersetzt wobei die Wertigkeit der Bilder durch Schlussfolgerungen jeweils ermittelt werden konnte. Zusätzlich wurde die Testserie auf einen zweiten Versuchsaufbau übertragen, in dem es galt, Futter in verschiedenen Substraten zu finden.

Die Resultate des ersten Versuchs zeigten, dass eine Krähe IBE benutzte um die Aufgabenstellung zu lösen, während jedoch Strategien, die auf Neophilie basierten, insgesamt am häufigsten verwendet wurden. Trotz methodischer Probleme im zweiten Experiment deuten die Daten darauf hin, dass die gegensätzliche Strategie, nämlich die Vermeidung des weniger Bekannten, benutzt wurde. Es wird diskutiert wie sich diese beiden Herangehensweisen

vereinen lassen und wie zusätzliche Habituation zu neuen Situationen einen beträchtlichen Einfluss auf die Ergebnisse haben könnte. Darauf basierend lies sich schließen, dass die Veranlagungen der Versuchsart berücksichtigt werden müssen um höhere kognitive Fähigkeiten, wie IBE, zu testen.

INTRODUCTION

Humans and other animals often encounter situations – either familiar or novel – where they only have access to incomplete or partial information. There are various different ways to deal with this lack of knowledge. Behavioural or cognitive processes range from simpler ones, such as continued exploration or avoidance of such situations, environments, objects or individuals, to cognitively more demanding ones, such as 'inferential reasoning', which is the ability to associate a visible and an imagined event (Premack 1995). Such an example is object permanence, where an individual knows that an object does not cease to exist when it is out of sight, e.g. covered or hidden (Piaget 1952). Another way of dealing with partial information is 'inference by exclusion' (hereafter referred to as 'IBE'), where an individual derives information by logically excluding potential alternatives (Call 2006). Many terms are used for this ability including 'inferential reasoning by exclusion', 'reasoning by exclusion', 'learning by exclusion', 'fast mapping' or 'emergent matching' (see Schloegl et al. 2009a for review). As it is thought to be cognitively demanding, IBE has received increasing interest in the field of comparative cognition and investigated in a variety of non-human mammalian species (chimpanzees, *Pan troglodytes* (Premack & Premack 1994; Hill et al. 2011; Call 2004; 2006; Bräuer et al. 2006; Beran & Washburn 2002; Beran 2010), bonobos, *Pan paniscus* (Hill et al. 2011; Call 2006; Bräuer et al. 2006), gorillas, *Gorilla gorilla* (Hill et al. 2011; Call 2006), orang-utans, *Pongo abelii* and *Pongo pygmaeus* (Hill et al. 2011; Call 2006; Marsh & MacDonald 2012), siamangs, *Symphalangus syndactylus* (Hill et al. 2011), one spider monkey, *Ateles geoffroyi* (Hill et al. 2011), olive baboons, *Papio hamadryas anubis* (Schmitt & Fischer 2009; Petit et al. 2015), tufted capuchin monkeys, *Cebus apella* (Sabbatini & Visalberghi 2008; Paukner et al. 2009; Marsh et al. 2015; Heimbauer et al. 2012), lion-tailed macaques, *Macaca silenus* (Marsh et al. 2015), rhesus macaques, *Macaca mulatta* (Petit et al. 2015), Tonkean macaques, *Macaca tonkeana* (Petit et al. 2015), one Hamadryas Baboon, *Papio hamadryas* (Marsh et al. 2015), squirrel monkeys, *Saimiri sciureus* (Marsh et al. 2015), black and brown lemurs, *Eulemur macaco* and *Eulemur fulvus* (Maille & Roeder 2012), Californian sea lions, *Zalophus californianus* (Kastak & Schusterman 2002), bottlenose dolphins, *Tursiops truncatus* (Herman et

al. 1984), dwarf goats, *Capra aegagrus hircus* (Nawroth et al. 2014), sheep, *Ovis orientalis aries* (Nawroth et al. 2014), swines, *Sus scrofa scrofa* and *Sus scrofa domestica* (Albiach-Serrano et al. 2012), Asian elephants, *Elephas maximus* (Plotnik et al. 2014) and dogs, *Canis familiaris* (Kaminski et al. 2004; Erdőhegyi et al. 2007; Aust et al. 2008; Bräuer et al. 2006; Zaine et al. 2014)), as well as avian species (greater anis, *Crotophaga major* (Riehl et al. 2015), pigeons, *Columba livia* (Tomonaga 1993; Clement & Zentrall 2003; Aust et al. 2008), African grey parrots, *Psittacus erithacus* (Schloegl et al. 2012; Pepperberg et al. 2013; Mikolasch et al. 2011), Goffin cockatoos, *Cacatua goffiniana* (O'Hara et al. 2015), keas, *Nestor notabilis* (O'Hara et al. 2016; Schloegl et al. 2009b), common ravens, *Corvus corax* (Schloegl et al. 2009b), jackdaws, *Corvus monedula* (Schloegl 2011), Eurasian jays, *Garrulus glandarius* (Shaw et al. 2013), Clark's nutcrackers, *Nucifraga columbiana* (Tornick & Gibson 2013), New Caledonian Crows, *Corvus moneduloides* (Jelbert et al. 2015), and carrion crows, *Corvus corone* (Mikolasch et al. 2012)).

Although the experimental set-ups in the above cited studies vary greatly, the same questions are asked in all of them: Does the subject choose correctly, if the only information provided is about the absence or falsity of something else? And if it does, then why/how? While the *IF* question has been addressed in a straightforward manner, the underlying mechanisms are often speculative, i.e. the *HOW* question needs further investigation. Studies such as the one by Beran & Washburn (2002) highlight the necessity for controlling alternative mechanisms. In a matching-to-sample task, three chimpanzees were able to match an unfamiliar comparison to an unfamiliar sample when given the choice between the unfamiliar and a familiar but incorrect comparison. However, tested later they did not associate this former unfamiliar comparison with the former unfamiliar sample, which suggests that subjects were simply avoiding the familiar but incorrect comparison rather than inferring and learning about the new match.

In order to control for alternative mechanisms, Aust et al. (2008) developed a sophisticated method to investigate IBE using a standardized two-choice computer test. So far, pigeons, dogs and humans (children as well as adults) have been tested with this method. First, subjects were taught to associate four rewarded pictures with positive contingencies (S+) and four with negative contingencies (S-). Subsequently, S- stimuli were paired with novel, unfamiliar stimuli

(S'). One out of six pigeons, three out of six dogs, five out of six students and all children chose S' above S-. The authors suggest that choosing the correct stimulus (S') could be a result of inferring by exclusion, avoiding the negatively associated stimulus or neophilic tendencies. To control for the latter explanations, they further tested all individuals, who showed a preference for S' in the first test, in a following second test, where S' was paired with another novel, unfamiliar stimulus (S''). The authors argued that if S' was chosen by reasoning by exclusion in Test 1, it would be logical to choose S' again in this subsequent test. All remaining dogs, students, and six out of eight children did so. The pigeon and two children showed a preference for S'' instead. Although pigeons are reported to show neophobic tendencies (Clement & Zentrall 2003) and therefore, choosing based on neophilia seemed unlikely, a differentiation between a neophilic and avoidance strategy could not be made. Also, it is important to note that test trials were not rewarded – regardless of the chosen stimulus. This might have led to a violation of expectation in Test 1 and thus possibly searching for alternative strategies in Test 2. Addressing these methodical problems, O'Hara and colleagues (2015) adopted and revised the methods, using 4 tests in total, to not only test the ability to infer by exclusion, but to also gain insight into the use of alternative strategies. They did so by, additionally to the above mentioned test trials, pairing the positive familiar stimulus (S+) with another novel stimulus (S-1) and then present this former novel stimulus (S-1) with another novel one, which should be attributed to positive feedback as S-1 can be excluded. The authors argued that if all tests are solved correctly by the subjects, novelty preference and one-trial-learning can be excluded as interpretation and therefore only IBE remains as a possible strategy. Additionally, to overcome the possible explanation of violation of expectation as described above, all correct choices were rewarded. Using this experimental set-up, it was shown that Goffin cockatoos (O'Hara et al. 2015) as well as keas (O'Hara et al. 2016) use IBE amongst other strategies to solve the task. With the present study, I aim to investigate strategies used by a member of the corvid family, the carrion crow, employing the methods of O'Hara et al. (2015; 2016).

Corvids have shown to be able to solve a variety of challenging cognitive tasks (see Emery 2006 for a review) and show high levels of neophobia as well as explorative behaviour (Miller et al. 2015). As representatives of the corvid family, common ravens (Schloegl et al. 2009b) and

carrion crows (Mikolasch et al. 2012) have been shown to successfully solve a frequently used task to test for IBE in a foraging context. In those experiments, the subjects were required to choose from one of two cups after limited information was given to them. To avoid local enhancement, two opaque outer cups were lifted; the inner cups where either transparent or opaque, thus providing visible information (transparent) or not (opaque). The reward was always hidden under one of those two cups, but the information provided to the individuals differed, depending on the condition they received: 'both' (both cups were transparent), 'baited' (only the baited cup was transparent), 'un-baited' (the empty cup was transparent) and 'control' (no cup transparent). On the group level, crows chose the correct cup in the crucial condition to test for exclusion (un-baited) significantly more often than in the control condition. On an individual level, two out of six chose correctly in this condition significantly more often than predicted by chance. Although the possibility of basing the choice simply on avoidance of the empty cup has to be acknowledged this was the first study to show possible IBE abilities of carrion crows. Both of the successful and a third unsuccessful individual were also part of the present study, which allows for a direct comparison of the methodology. According to these findings, I predict that carrion crows are able to show IBE abilities in the described artificial touch screen task, similar to the tested parrot species as both avian families have shown striking cognitive abilities in the past (review in Güntürkün & Bugnyar 2016).

Additionally, a second experiment was conducted in which the crucial information given was of auditory nature. Eurasian jays (Shaw & Clayton 2012) and Western scrub-jays, *Aphelocoma californica* (Stulp et al. 2009), limit auditory information while 'caching', i.e. hiding food, when competitors are within hearing distance, suggesting that they are aware of the link between the sound and the caching procedure. In the second experiment, the test series of the first experiment was transferred to an ecologically more relevant food-finding procedure using acoustic cues. Like jays, carrion crows cache food and paying attention to auditory information produced by others might be of advantage for an individual. Therefore I predict that they will use acoustic information to solve the task of the second experiment as well.

Taken together, I aim to investigate not only carrion crows' ability to infer by exclusion but also to focus on potentially applied alternative strategies using an abstract touch screen experiment

as well as an ecologically relevant foraging task. In doing so, both the visual as well as the acoustic domain are investigated in order to broaden our understanding of the underlying mechanisms.

METHODS

Study animals and housing

The experiments were conducted from February 2014 until January 2015 with five carrion crows (*Corvus corone*), two females and three males (Table 1). They were kept in pairs in outdoor aviaries of the size of 23.5 m^2 in the Cumberland game park in Grünau im Almtal. As shown in Figure 1, all aviaries were adjacent to the testing area. The crows were fed twice a day and received water ad libitum. The diet consisted of meat, bread, yoghurt, curd, fruits and herbs and was supplemented with vitamins and minerals for corvids twice a week. The aviaries were equipped with natural vegetation such as bushes, branches, stones and grass (Figure 3). All birds have participated in experiments involving touch screen computers before (Braun et al unpubl. data, Rockenbach et al. unpubl. data, Federspiel et al. in prep.). Experiments were conducted twice a day.

Table 1: Details of subjects and participation in experiments.

	SEX	YEAR OF HATCHING	PAIRED WITH	PARTICIPANT IN
Petra	f	2007	Bärchen	Experiment 1 & 2
Bärchen	m	2008	Petra	Experiment 1 & 2
Walter	m	2011	Bärbel	Experiment 1 & 2
Willi	m	2012	Momo	Experiment 1
Töffel	f	2008	Rüdiger	Experiment 1

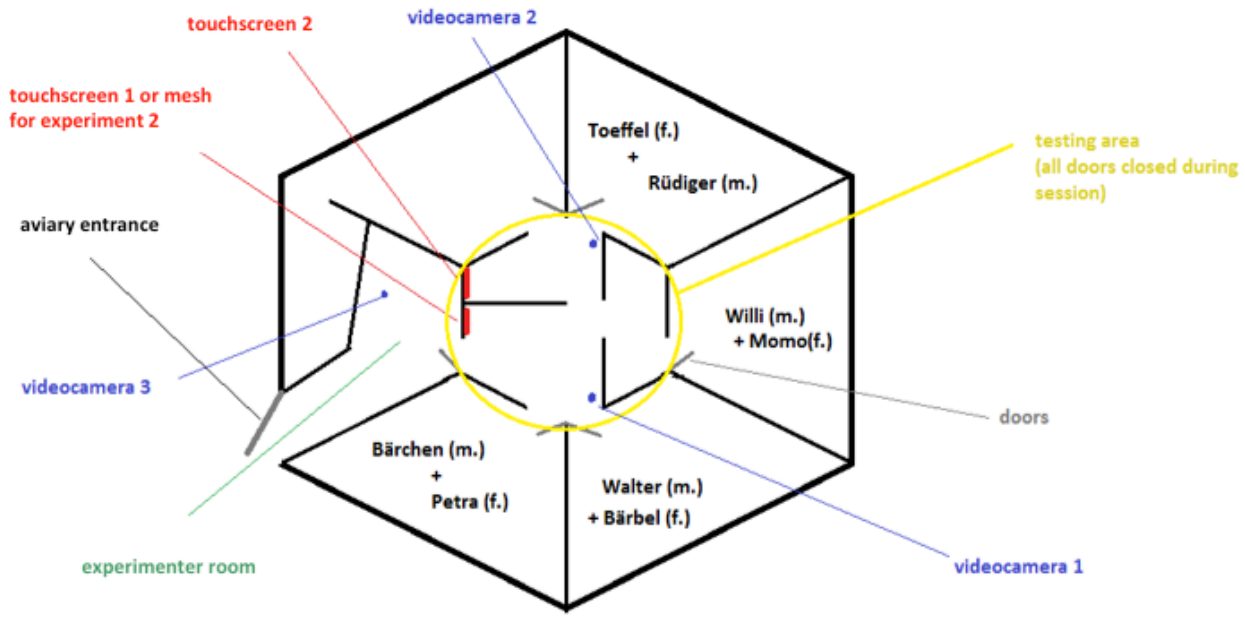

Figure 1: Outline of housing and experimental area: Aviaries were arranged in a circle around an inner testing area, where the experiments were carried out. Four of the five aviaries were used for housing.

All birds were tested individually by calling them into the visually isolated test area in the middle of the aviary complex (Figure 1). In front of the touchscreens (Figure 2a) and choice boards (Figure 2b), perches were installed for the birds to sit on while testing.

a) b)

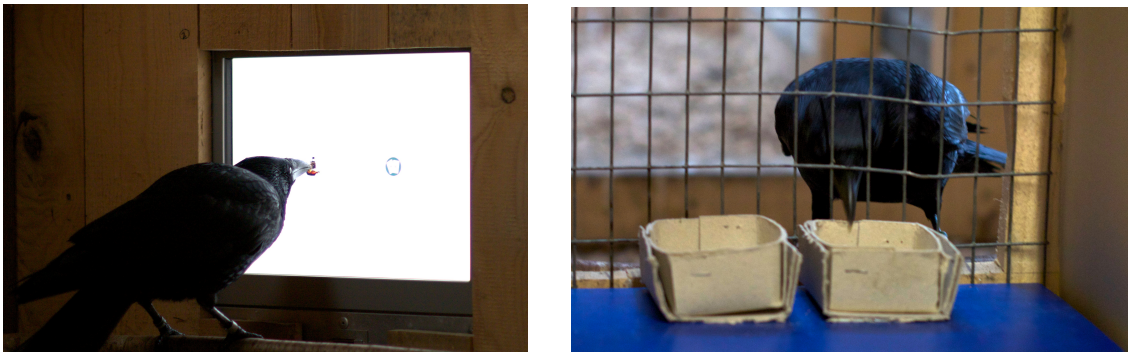

Figure 2: a) A crow choosing a stimulus by pecking at the touch screen in experiment 1. b) A crow making a choice by touching a box in experiment 2.

Experiment 1

EXPERIMENTAL SET-UP

Coloured clip arts (http://www.openclipart.org) (2,47cm x 2, 47 cm; resolution: 72ppi) were used as stimuli (see Table 2 for examples), which were randomly assigned a positive or negative contingency for each individual. White was used as background colour and shown during inter-trial intervals (ITI), which lasted for one second. With the exception of the phase 'Habituation', two stimuli were always shown next to each other in the centre of the screen. Sides (left/right) were assigned randomly for each trial by the software.

PROCEDURE

During the experiment in the testing area, the subject was visually and physically separated from its partner. The experimenter was in a room next to the testing area and thus visually and physically isolated from the crow. The subject was allowed to choose one of the two touch screen computers (Figure 1).

Birds were moved on to the next phase after reaching a learning criterion (see below for the specific criterion of each phase). Within each phase, birds participated in one session of 20 trials twice a day. When a correct choice was made, a small piece of dry dog food was dispensed automatically as a reward after each trial. After completing a session, the bird received a 'higher quality reward', a greave, was reunited with its partner and provided with their regular food. If a bird showed obvious decrease in motivation and stayed away from the experimental set-up for more than 15 minutes, the session was aborted but continued on the next half-day, with the exception of the phase 'Pre-Training', where new sessions were started after terminating one due to lack of motivation and sessions were only counted as completed when a limit of at least 14 trials was reached.

The experiment was conducted in four phases: Habituation, Pre-Training, Training and Test phase. The methods for Training and Test phase were based on the methods employed for testing Goffin cockatoos and keas on IBE (O'Hara et al. 2015; 2016).

Habituation

First, birds were ‘reminded’ of how to work with the touch screen computer and to associate a feedback sound with the reward. Stimuli were displayed at random in different locations at the screen. When birds pecked at a stimulus with their beak, a reward (dry dog food) was dispensed by an automated feeder and a sound (300 Hz, 100 ms) was played back. This sound was then assumed to be perceived by the crows as positively associated with food. This phase concluded once the birds reliably pecked at the stimulus and consumed the food reward without any signs of discomfort with the procedure.

Pre-Training

In this phase, animals learned to choose one of two images. Pictures of circles and triangles of different colours were displayed in pairs. Triangles were always rewarded with dry dog food and a positive feedback sound (same as during Habituation), while pecking at a circle led to a negative feedback sound (150 Hz, 300 ms), a delay of two seconds and a correctional trial, in which the same pair of pictures was shown again in the same position. This phase ended when birds had reached a level of at least 80% correct first choices in three out of six consecutive sessions.

After one bird (Willi) completed three sessions with additional negative feedback of a red screen for one second, I decided to change feedback settings to match the procedure of Training and Test phase of O´Hara et al. (2015; 2016) as to not compromise comparisons.

To familiarize the bird with those feedback procedures, four additional sessions without the red screen as negative feedback were conducted although the criterion was achieved after the first three. All other birds received feedback as described above.

Training

After the Pre-Training, birds learned to associate positive feedback with a rewarded stimulus (S+) and negative feedback with an unrewarded one (S-). S+ and S- were paired in 18 baseline

trials. Additionally, two 'novelty trials' were interspersed, showing an S+ and a novel stimulus (S-*), which led to negative feedback when pecked at. Pairings of S+ and S-* revealed whether the crow remained pecking at the learned positive stimuli and thus, controlled for novelty preferences and reinforced positive stimuli. S+ and S- stayed the same for each bird across all sessions, while novel stimuli were only shown once. This phase ended when birds had reached a level of at least 80% correct first choices in the baseline trials in two consecutive sessions.

Test phase

In this phase, birds were tested for response strategies. Five different types of trials were presented. One session consisted of 16 baseline trials (same as in Training) and four interspersed test trials.

In test trials one stimulus, either with positive or with negative contingencies, was replaced subsequently by novel stimuli. Test 1 consisted of the presentation of the familiar incorrect baseline stimulus (S-) together with a novel stimulus (S+1). S+1 was later paired with another novel stimulus (S-2) in Test 3. In Test 2, S- was replaced by a novel stimulus (S-1), which was later paired with another novel stimulus (S+2) in Test 4. Each test provides information on possible response strategies and controls for alternative explanations (see Table 2).

Test trials were always presented in the order from test trial 1 to 4. The positions of test trials within a session were randomly assigned to one of two possible positions in the trial order (Test 1: trial 4 or 5; Test 2: trial 8 or 9; Test 3: trial 12 or 13; Test 4: trial 16 or 17). In the baseline trials, crows received positive and negative feedback (dog food and/or sounds, correctional trials, delays) as they did in the Training. During test trials, positive feedback was given for correct choices, but no feedback for incorrect choices. The positive feedback was supposed to avoid discouragement and potential searching for other underlying rules by the crow as a result of violation of expectation. No feedback for incorrect choices, on the other hand, should minimize learning effects over the duration of the experiment. S+ and S- stayed the same across all sessions, while S+1 (first novel positive stimulus), S+2 (second novel positive

stimulus), S-1 (first novel negative stimulus), S-2 (second novel negative stimulus) were novel for every session with the exception of repeated sessions.

Table 2: Overview of a test session with example stimuli. Response strategies for choosing the correct stimulus are given for each trial separately. Response strategies in bold and colour are the most likely ones applied if birds chose correctly also in all previous test trials. The abbreviation OTL stands for 'one–trial-learning'

	EXPERIMENT 1			
	positive	negative	response strategies	control for
16 x baseline trials			**association**	
1 x test trial 1			**IBE** **S- avoidance** **novelty preference**	novelty aversion
1x test trial 2			**association** novelty aversion	novelty preference
1x test trial 3			**OTL through IBE** **OTL through S- avoidance** novelty aversion	S- avoidance (without OTL effect)
1x test trial 4 sequence A			**IBE** novelty preference	OTL through S- avoidance
sequence B			**IBE** novelty preference	OTL through S- avoidance

In the first 10 sessions, birds were randomly assigned to group A or group B. The sequences A and B were alternated for every crow after every session. Birds of group A started with sequence A, while group B started with sequence B (see Table 2). Subsequently, those same 10 sessions were repeated for four birds (not Petra, as she showed motivational problems over longer periods of time and continuing with unique sessions was necessary to finish data collection in time). Additionally, every bird was tested for another unique 20 sessions, where only sequence B was presented. Thus, a total of five unique test sessions of sequence A and a total of 25 unique test sessions of sequence B was conducted with the exception of one female crow (Töffel), who unfortunately died for reasons unrelated to testing before starting the additional 20 sessions with sequence B.

DATA COLLECTION AND ANALYSIS

As computer 'VIA Nehemiah' was used with an 'Elo-Touchscreen' (Modell: Carrol Touch) and the software 'CognitionLab Light' Version 1.9 built 56 ((c) 2007-2009, 2010 Michael Steurer). All sessions were recorded with Sony digital video camera recorders (Model No. DCR – SX34 E).

As proposed by O'Hara et al. (2015; 2016), I looked for the most parsimonious explanation for every possible outcome in the 25 unique sessions of sequence B (see also Table 3):

The pattern of correct choices in all test trials was attributed to 'IBE' as all other explanations can be excluded (see Table 2 for controls in each test trial).

Four patterns were ascribed to 'one-trial-learning'. In those patterns feedback for a stimulus in one test trial resulted in correct choices the second time this stimulus was shown. Therefore, although strategies of why the choice was made in the first place might vary, only one test trial was needed by the subject to associate with the contingency of the new stimulus, hence 'one-trial-learning'. This also applies for the pattern 'one-trial-learning 3', where all test trials but Test 4 could be solved by other mechanisms in combination of one-trial-learning. When this was the case no direct feedback was given to the stimuli in Test 4. Therefore if choices were correct also in this test one can concluded that IBE was used – if not one-trial-learning is the most parsimonious explanation for correct choices in previous test trials.

As described by O´Hara and colleagues (2015; 2016), subjects might have developed a rule of avoiding the novel stimulus due to feedback in novelty trials of the Training phase. Therefore an abolishment or reversal of this rule could have played a role in 'one-trial-learning 1' and 'one-trial-learning 2'. However in contrast to those two studies the crows did not need to reach a novelty criterion before proceeding with the next phase. Therefore one cannot conclude that a novelty rule was learned before entering the Test phase. However, if so 'one-trial-learning 1' and 'one-trial-learning 2' should have occurred more often in the earlier test session than in the later ones.

When the least novel stimulus was chosen in all test trials the pattern was assigned to 'novelty aversion'. In contrast 'novelty preference' describes the pattern when the most novel stimuli were chosen consistently. Additionally, in contrast to O´Hara et al. (2015; 2016) I ascribed one pattern to 'novelty preference & S+ preference'. Here all choices, except in Test 2, were

consistent with the strategy of choosing the most novel stimulus. However, I believe that choosing the familiar and repeatedly rewarded baseline stimulus S+ in combination of a novelty preference might be the most parsimonious explanation for this pattern.

In earlier studies (e.g. Aust et al. 2008) the possibility of solving tasks, designed to test for IBE, by purely avoiding familiar but unrewarded stimuli have been addressed. Here the pattern in which Test 1 is the only test trial where choices were made correctly was ascribed to 'S-avoidance'. Choosing the familiar and repeatedly rewarded stimulus in Test 2 additionally does not stand in contrast to an avoidance strategy and is therefore named 'S-avoidance & S+ preference'. For the remaining patterns the most parsimonious explanations of their occurrence seem to be combinations of stimulus preferences and avoidances, possibly in combination with an initial preference or aversion of novel stimuli (see Table 3 for each pattern), as well as having a 'bad day' when no test was solved correctly.

In sum, the patterns add up to 16 possible outcomes. Therefore, each pattern had a cumulative probability of $p=0.0625$ to occur by chance. By using two-sided binomial tests, I calculated which patterns occurred more frequently than predicted by chance.

To compare the results with Aust et al. (2008), I investigated preferences for positive or negative stimuli for each trial category in test sessions (baseline, Test 1, Test 2, Test 3, Test 4b) separately of all 25 unique sessions of sequence B. Based on a probability of $p=0.5$ for each choice to occur by chance significance levels were calculated using one-tailed binomial tests. Preferences for positive stimuli were significant when they were chosen at least 18 times (≥72%, $p= 0.022$) and for negative stimuli when they were selected 7 times or less out of 25 sessions (≤28%, $p= 0.022$).

In addition I looked at the performance in the combination of Test 1 and Test 3 as well as Test 2 and Test 4 of sequence B separately. To choose correctly in Test 3 information about the positive stimulus (S+1) had to be transferred from Test 1 but no information of Test 2 or Test 4 was needed. When S+1 was chosen in Test 1 positive feedback was given for this stimulus. Therefore the pairing of Test 1 and Test 3 is termed 'reinforced combination' hereafter. In the same manner one had to acquire information about S-1 in Test 2 to solve Test 4 correctly, while no information of Test 1 and Test 3 was needed. However a correct choice in Test 2 did not

result in direct feedback for S-1. Therefore this pair of tests is referred to as 'unenforced combination' hereafter. Exact binomial test (two-sided) was applied to set significance levels at 11 out of 25 sessions (44%, p= 0.037) and one-sided Wilcoxon signed rank test was used to determine the difference of both, post hoc. For comparing the latencies of different conditions I used two-sided Wilcoxon signed rank tests (corrected after Bonferroni).

Additionally, the performances in the first five unique sessions of Test 4b (= Test 4 sequence B) and Test 4a (= Test 4 sequence A) were compared as well as the first set of five unique sessions of Test 4b and second set of the same sessions for each pattern, using two-sided or one-sided Wilcoxon signed rank tests, as appropriate. To detect a possible side bias two-sided binomial test was conducted.

Tests were conducted using the statistical program R (R Core Team 2014) and visualized with the package ggplot 2 (Wickham 2009). α was set at 0.05.

Table 3: Overview of all 16 possible patterns with the most parsimonious explanations. ✓ indicates the correct choice of the positive stimulus (shown on the left side of each pair) while ✗ represents incorrect choices (choosing the negative stimulus, which are shown on the right side of each pair).

BASELINE		NOVELTY TRIAL		
TEST 1	TEST 2	TEST 3	TEST 4	RESPONSE STRATEGY
✓	✓	✓	✓	IBE
✗	✓	✓	✓	one-trial-learning 1
✗	✗	✓	✓	one-trial-learning 2
✓	✓	✓	✗	one-trial-learning 3
✓	✗	✓	✓	one-trial-learning 4
✗	✓	✓	✗	novelty aversion (NA)
✓	✗	✗	✓	novelty preference (NP)
✓	✓	✗	✓	novelty preference & S+ preference
✓	✓	✗	✗	S- avoidance & S+ preference
✓	✗	✗	✗	S- avoidance
✓	✗	✓	✗	S-1 preference and/or S+1 preference
✗	✓	✗	✓	S-1 avoidance and/or S+1 avoidance
✗	✓	✗	✗	NA & S+1 avoidance and/or NA & S-2 preference
✗	✗	✓	✗	NA & S-1 preference
✗	✗	✗	✓	NP & S+1 avoidance
✗	✗	✗	✗	bad day

Experiment 2

EXPERIMENTAL SET-UP

In Experiment 2, crows sat on a perch in front of a wire mesh window (Figure 2). The experimenter sat on the other side of the mesh in full view of the bird. A small table at a height of 40.5 cm was positioned on the experimenter's side of the mesh. A sliding board (37 cm x 20 cm) was placed on the table in front of the mesh. Two cardboard boxes (8 cm x 11 cm x 4 cm) were glued onto the board at a distance of 2.5 cm from each other. These boxes were open on the side facing the experimenter so as to allow sliding in two slightly smaller cardboard boxes (10 cm x 7 cm x 4 cm). This was necessary in order to keep birds from flipping boxes, but still enabled the experimenter to switch boxes easily. Rewarded sides were assigned randomly for each trial. Twenty-two 'silent' substrates, which produced no (or minimal) noise when manipulated, and 22 'noisy substrates', which produced sounds when manipulated, were prepared from different materials (see Figure 3 a, b). To the best of my knowledge, those materials were novel to the individuals.

PROCEDURE

As in Experiment 1, the subject was visually and physically separated from its partner. The experimenter was in the room adjacent to the testing area and in full visibility of the subject (Figure 1). Also, as in the previous experiment, birds participated in four phases of Habituation, Pre-Training, Training and Test phase, in each of which they had to reach a criterion for. Testing schedule, reward system and abortion rules were applied as in Experiment 1. I started with sessions of 20 trials, but later on in the experiment switched to 10 trials per session. This was done, because sessions had to be aborted multiple times as attention and motivation seemed to drop quickly after a few trials. The experimenter wore a baseball cap during trials, lowered her view so that eyes were not visible to the bird and looked at the middle part of the sliding board at her end, while birds were choosing, in order to avoid cueing by gaze.

Habituation

In order for the subjects to get used to the materials used in the experiment, 22 ‘silent’ and 22 ‘noisy’ substrates were put into the small cardboard boxes. For eight days, each box was picked up separately by the experimenter twice a day and the containing substrates were shown to the birds form the outside of the aviaries. Noisy substrates were manipulated during presentation, because it was necessary for the crows to associate each manipulation of a noisy substrate with its noise when manipulated in order to be able to infer by exclusion using acoustic cues later on. Silent substrates were not manipulated. Boxes were left in front of each aviary for another five minutes after presentation and then manipulation started at the next aviary (Figure 3 a,b). Additionally, the boxes where left in front of each aviary for a total of 20 hours without manipulation. Empty identical boxes were also placed into the aviaries in order to allow habituation. To prevent olfactory cues, small pieces of dry dog food were put into every substrate for two days before testing.

a)

b)

Figure 3 a),b): Habituation to substrates in boxes in front of an aviary.

Pre-Training

In this phase, crows learned to choose one of two empty boxes by touching it with their beak (Figure 2b). One piece of dry dog food was shown to them and put into one of the boxes in full view of the bird. Subsequently, the sliding board was pushed towards the mesh. In this position, the bird was able to touch a box of choice. If the correct choice was made, the crow was rewarded with the dog food. If the incorrect choice was made, the chosen box was shown again to the bird for it to see that there was no reward inside and then removed. As in Experiment 1, each trial was repeated until the correct choice was made (correctional trials), and the phase ended upon reaching a learning criterion of at least 80% correct first choices in three out of six consecutive sessions.

Training

Next, the birds were further acquainted with parts of the procedure used in the Test phase. One box was filled with ‘noisy’ substrate, while the other was filled with ‘silent’ substrate. The reward was shown to the crows and subsequently hidden out of view of the crows in the ‘noisy’ substrate. The sliding board was put on the table and slid towards the mesh to allow the crows to choose. When the correct choice was made the crows were rewarded with dog food. If the incorrect choice was made, the boxes were removed. As in Experiment 1, each trial was repeated until the correct choice was made (correctional trials), and the phase ended when a level of at least 80% correct first choices in two consecutive sessions was reached.

Test phase

During the Test phase the procedure was identical to the Training, with one exception: There were no correctional trials for test trials. As in Experiment 1, four different types of test trials were conducted to test for response strategies (Table 4). Baseline substrates stayed the same for each crow, while test substrates changed every session. The plan was to conduct 10 sessions per bird; unfortunately, only one crow finished eight sessions, while the others stopped

participating earlier (when entering the Test phase or at one session and eight trials, respectively).

Table 4: Overview of test trials in Experiment 2. Response strategies for choosing the correct substrate are given for each trial separately. Response strategies in bold and colour are the most likely ones applied if birds chose correctly also in all previous test trials. The abbreviation OTL stands for 'one–trial-learning'. n = 'noisy' substrate; s = 'silent' substrate; n-b = baseline 'noisy' substrate (as presented in Training sessions); s-b = baseline 'silent' substrate (as presented in Training sessions); n-1 and n-2 = novel 'noisy' substrates; s-1 and s-2 = novel 'silent' substrates; superscripted: visual = due to visual cue; acoustic =due to acoustic cue

	EXPERIMENT 2			
	rewarded	not rewarded	response strategies	control for
16 x baseline trials	n-b	s-b	**association** visual/acoustic	
1 x test trial 1	n-1	s-b	**IBE** visual/acoustic **s-b avoidance** visual/acoustic **novelty preference** **association** acoustic	novelty aversion
1x test trial 2	s-b	n-b	**IBE** acoustic	s-b avoidance associations visual/acoustic
1x test trial 3	n-1	s-1	**IBE** acoustic **OTL through IBE** visual/acoustic OTL through s-b avoidance visual/acoustic novelty aversion association acoustic	s-b avoidance (without OTL effect) novelty preference
1x test trial 4	s-2	n-1	**IBE** acoustic novelty preference	OTL through s-b avoidance

DATA COLLECTION AND ANALYSIS

All sessions were coded live and videotaped with Sony Handycam (Model No. DCR – SX85). For the one crow (Petra), who finished eight test sessions, I investigated her preferences for the rewarded or non-rewarded substrate for each trial category in those test sessions (baseline, Test 1, Test 2, Test 3, Test 4). Based on a probability of p=0.5 for each choice to occur by chance, significance levels were calculated using one-sided binomial tests. Preferences for rewarded stimuli were significant when they were chosen at least seven times (≥87.5%, p=0.035) and for negative stimuli when they were selected one time or less (≤12.5%, p=0.035) out of eight sessions.

RESULTS

Experiment 1

PRE-TRAINING AND TRAINING

In the Pre-Training, four birds reached the criterion of at least 80% in three out of six consecutive sessions with an average of 8.48 sessions (+/- 3.36 SE). One individual did not reach the criterion (Petra). As she was able to solve this task in a previous experiment I went on with the next phase.

In the Training, all individuals but one reached the criterion of at least 80% in two consecutive sessions in two to nine sessions with an average of 4.09 sessions (+/- 1.29 SE) (Figure 4). One bird (Willi) was first to reach criterion in Pre-Training while being last in Training, whereas another bird (Töffel) was last in Pre-Training and first in Training. Due to an error in data management, one crow (Petra) started Test phase although she reached 80% in session 5 and only 78% in the consecutive session. Nevertheless, she scored at least 80% in all test sessions but session 17. For this reason, she was not excluded from further analyses. All other birds stayed above a performance of at least 80% during all test sessions.

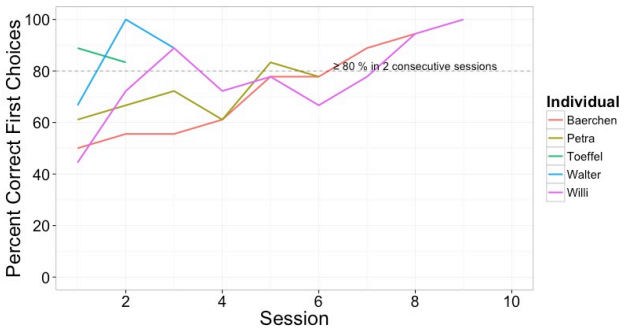

Figure 4: Performance in Training. Coloured lines indicate individual learning curves. The dashed line represents performance of 80% correct first choices. The length of individual lines indicate the amount of sessions needed to reach the Training criterion.

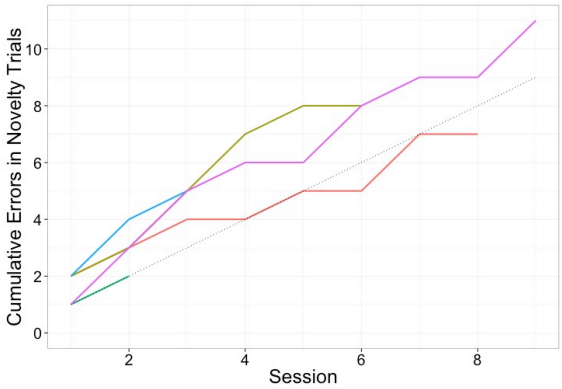

Figure 5: Cumulative errors in novelty trials in Training sessions. The dotted line indicates the slope of choosing the novel stimulus one out of two possible times per session (chance level). Horizontal lines indicate correct choice of S+. The length of individual lines is restricted to reaching Training criterion. The steeper the slope, the more novel stimuli were chosen.

All individuals chose the novel stimulus at least once in their first session. Three birds (Walter, Petra and Willi) tended to choose novelty over the course of the Training phase while one individual's slope (Bärchen) flattens during the course of the sessions, meaning that novelty preference decreased (Figure 5).

RESPONSE STRATEGIES

During testing (unless explicitly stated otherwise results are shown for the 25 unique sessions of sequence B (see Table 2)), 'novelty preference & S+ preference' (M=18%, SE=5.51) seemed to be the response strategy used significantly above chance on a group level (including four individuals) (Figure 6).

Other response strategies used include 'S- avoidance' (M=12%,SE=3.83), 'S- avoidance & S+ preference' (M=12%, SE=1.63), 'one-trial-learning 3' (M=10%,SE=4.16) and 'S-1 avoidance and/or S+1 avoidance' (M=10%,SE=2.58). One bird (Töffel) was excluded from group analysis as she only conducted five unique sessions in group B.

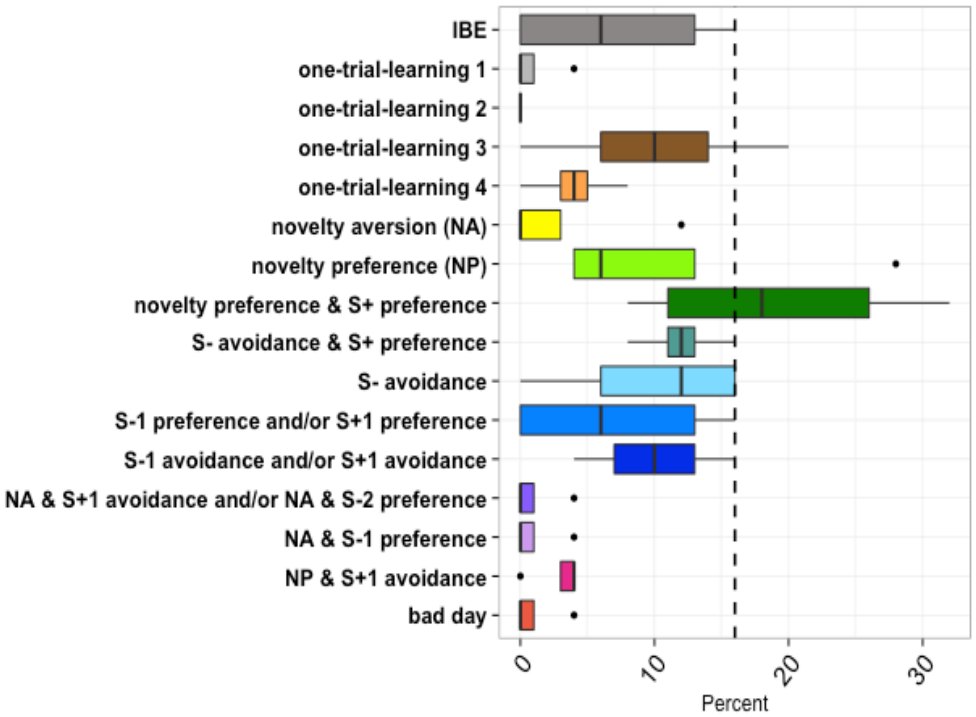

Figure 6: Response strategies on the group level for four individuals. Boxes include the first to the third quartile, whiskers represent 95% confidence interval, dots show outliners, vertical lines are medians. Values to the right of the dashed line are significantly exceeding chance level.

On an individual level, ‘novelty preference & S+ preference’ was used significantly more often than predicted by chance (binomial test: $p<0.05$) by two individuals (Petra and Willi) as well as ‘one-trial-learning 3’ (Willi) and ‘novelty preference’ (Walter) (Figure 7). Tendencies (binomial test: $p=0.067$) for other response strategies include ‘IBE’ (Willi), ‘S- avoidance & S+ preference’ (Walter), ‘S- avoidance’ (Walter and Petra) and ‘S-1 preference and/or S+1 preference’ (Walter) and ‘S-1 avoidance and/or S+1 avoidance’ (Bärchen). Additionally, one individual (Töffel) used ‘IBE’ significantly more often than predicted by chance (binomial test: $p=0.034$; however, she participated in fewer sessions than the others (n=5)).

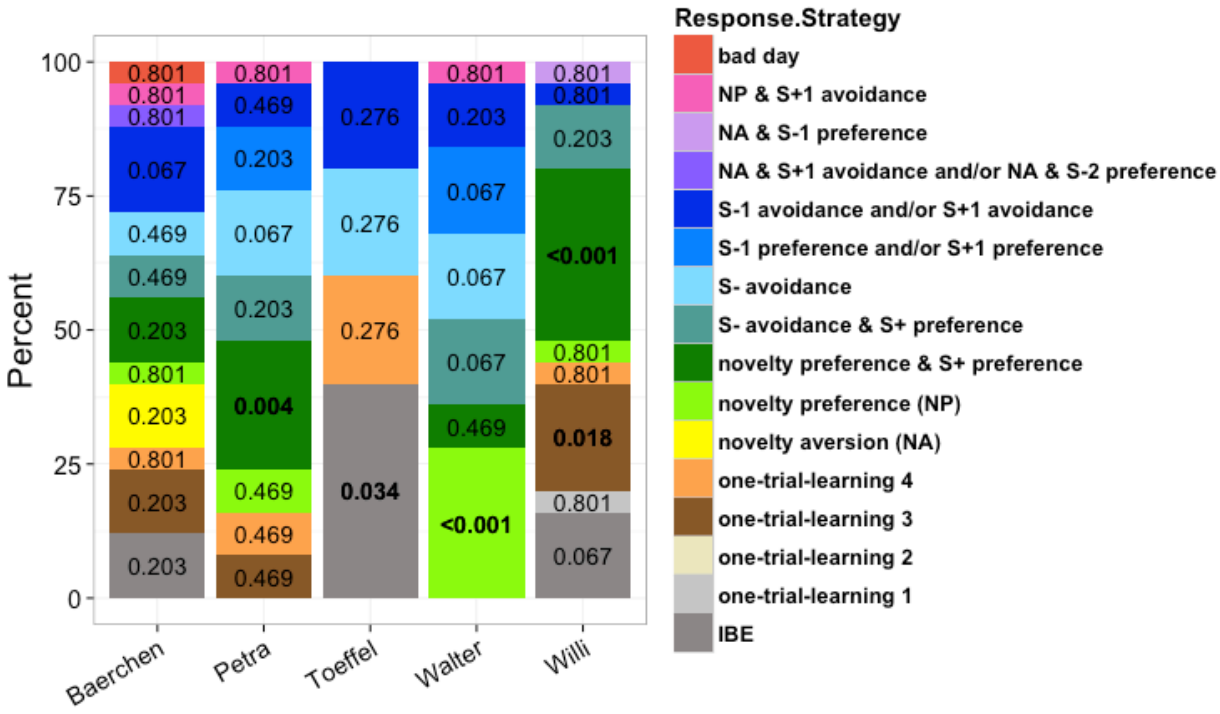

Figure 7: Response strategies per individual. Differently coloured boxes show the percentage for the occurrence of every pattern. Numbers in boxes represent p-values for the cumulative probability of each pattern and bolded numbers have a p-value smaller than 0.05 (Bärchen, Petra, Walter, Willi n=25; Töffel n=5).

Figure 8 shows response strategies per bird over the course of the experiment. One bird (Willi) used three different patterns at least four times: 'IBE', 'one-trial-learning 3' and 'novelty preference & S+ preference'. Two-sided binomial testing showed that those three together occurred significantly more often than by chance (for a probability of 3/16; $p < 0.001$). No learning effects were apparent.

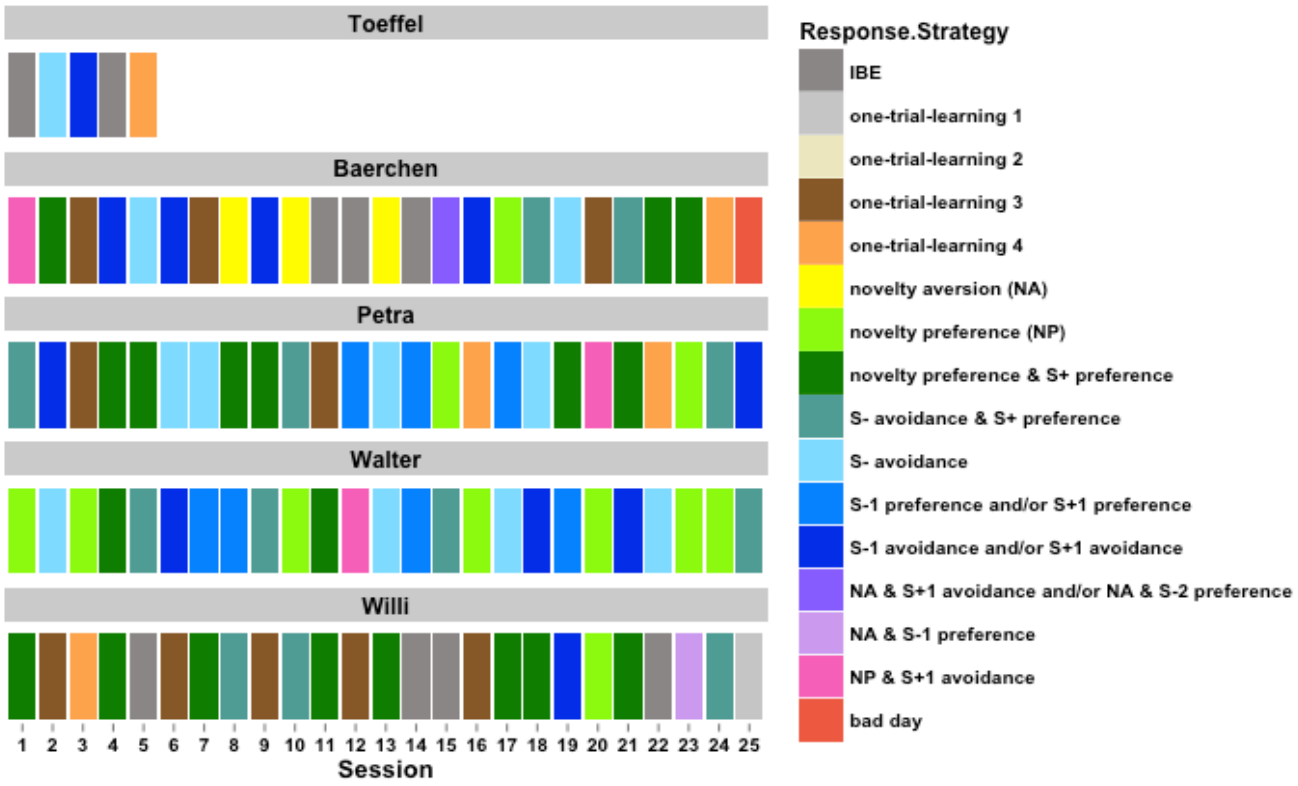

Figure 8: Response strategies of each session per bird.

SEPARATE CONDITIONS

Looking at the average percentage of correct first choices for each separate condition, one can see that all four individuals reached more than 72% in the baseline condition during testing (Töffel was excluded due to the smaller number of trials she had participated in. Therefore different significance levels would apply for her.) (Figure 9). In Test 1, three birds (Petra, Walter, Willi) and in Test 2, two birds (Bärchen and Willi) chose the positive stimulus significantly more often than predicted by chance (≥72%). In Test 3, two birds (Petra and Walter) showed a preference for the negative stimulus (≤28%).

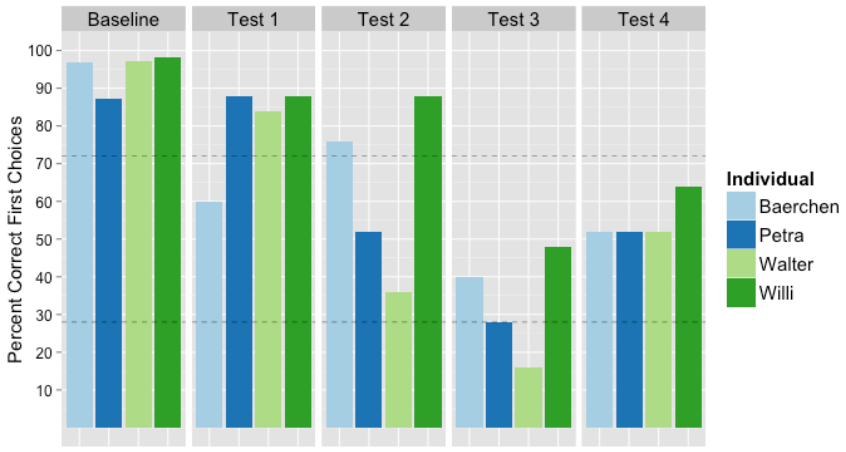

Figure 9: Percentage of correct first choices in each condition. Coloured boxes represent individuals. Values above the upper dashed horizontal line indicate a significant preference for positive stimuli (≥72%), while values below the lower dashed line represent a preference for negative stimuli (≤28%) significantly above chance level for each individual and test trial.

As seen in Figure 10, it took birds different amounts of time to choose a stimulus in the different conditions. I found that latencies were longer in Test 3 (Wilcoxon: $p < 0.001$) and Test 4 (Wilcoxon: $p = 0.003$) compared to the baseline trials, while there was no significant difference between baseline and Test 1 (Wilcoxon: $p = 0.304$) or Test 2 (Wilcoxon: $p = 1$).

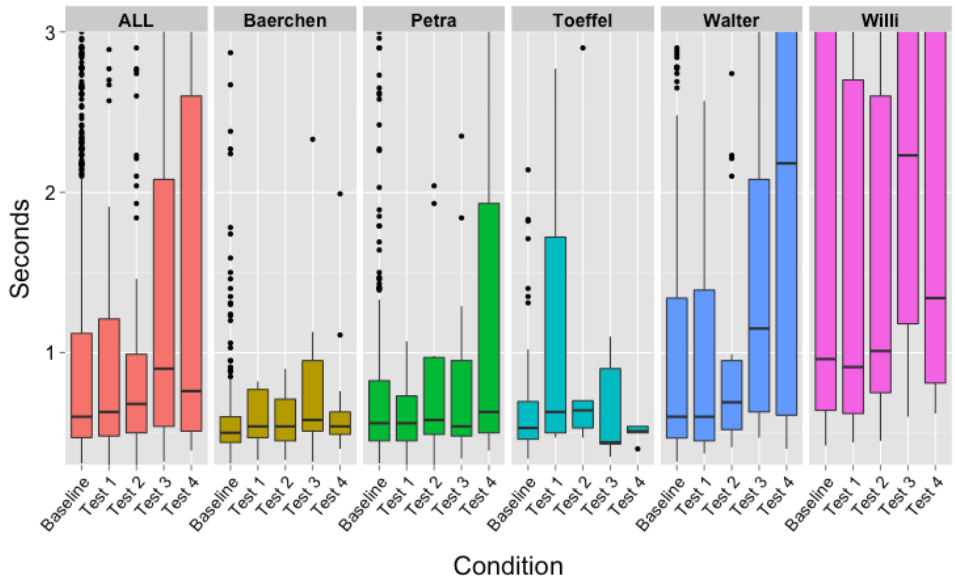

Figure 10: Response latencies are shown for each condition and all birds taken together (= ALL) as well as every individual separately. Boxes include the first to the third quartile, whiskers represent 95% confidence interval, dots show outliners, horizontal lines are medians.

FURTHER RESULTS

One individual (Willi) was correct in the 'unenforced combination' (Test 2 and Test 4) more often than predicted by chance (binomial test: p= 0.001) (Figure 11). One-sided Wilcoxon test showed an overall better performance for all birds in the 'unenforced combination' (p= 0.049) than in the 'reinforced combination' (Test 1 and Test 3).

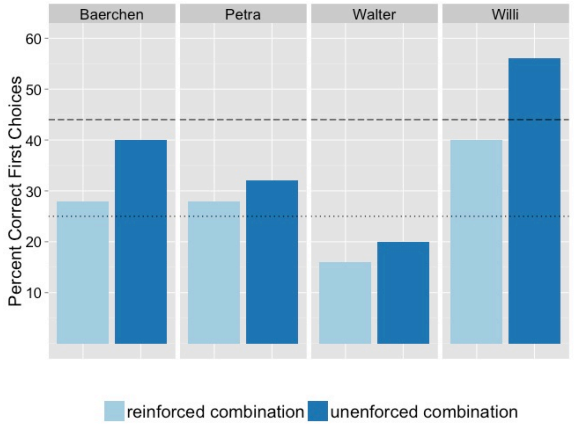

Figure 11: Percentage of correct choices in 'reinforced' (Test 1 and 3) and 'unenforced' (Test 2 and 4) combinations. Coloured boxes represent different combination pairs. Values ≥ 44% are significantly above chance level.

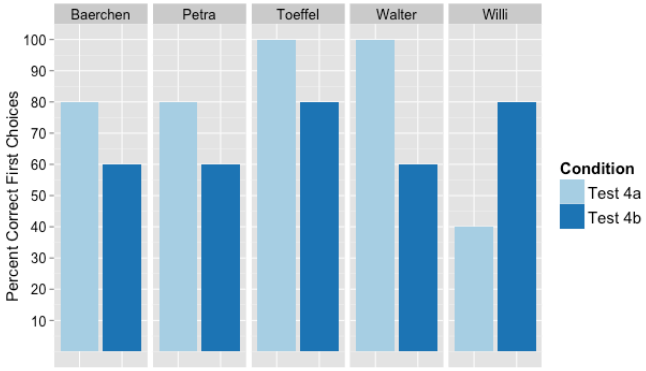

Figure 12: Percentage of correct first choices in the first five unique sessions comparing Test 4a (= light blue) and Test 4b (= dark blue).

Comparing the performance of Test 4a (= Test 4 in sequence A) and Test 4b (= Test 4 sequence B) in the first five unique sessions of each individual revealed a tendency of four birds to perform better in Test 4a, while data for one bird (Willi) showed the opposite trend (Figure 12). Two-sided Wilcoxon test showed no significant difference between the overall performance of Test 4a and Test 4b (p= 0.49). However, after excluding the outlier (the data point for the bird which exhibited an opposite trend), one-sided Wilcoxon test revealed a significantly better performance in Test 4a (p=0.044). Walter, Bärchen and Petra were assigned to group A (starting with Test 4a), while Willi and Töffel received Test 4b first (group B).

Pairwise Wilcoxon tests (two-sided) revealed no significant differences between the first set of five unique sessions and the second set (repetition) in any response strategy.

Figure 13 shows the distribution of consistent and inconsistent choices in the first and in repeated sessions of sequence A and B (both correct: 44.87 %; both incorrect: 19.87%; only first correct: 22.44%; only second correct: 12.82%).

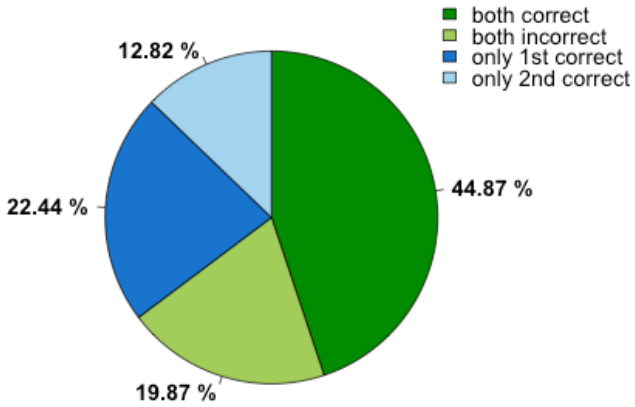

Figure 13: Comparison of choice variance in test trails of set 1 (first unique sessions) and set 2 (repeated sessions).

In one bird (Bärchen), a side preference was found (binomial test: p= 0.008). In all conditions where no significant preference for correct or incorrect stimuli was found (Test 1, 3, 4; see Figure 9), 27 out of 37 errors were made on the right side (as opposed to the left). Due to the automated assignment to the left and right side by the software, positive stimuli were shown in 51.4% of the cases on the right in all trials for this individual. Even if previous feedback increased chance level to this percentage, a significant side preference was found (binomial test: p= 0.009).

Experiment 2

PRE-TRAINING AND TRAINING

In the Pre-Training, two out of three birds reached the criterion after less than six sessions (Petra and Bärchen), while one bird (Walter) needed 40 sessions for a minimum of 80% correct first choices in three out of six consecutive sessions.

Two birds participated in the Training, while the third (Bärchen) refused to touch boxes with substrates in them. Both crows reached the criterion within two sessions, with one (Petra) even achieving 100% correct first choices in both.

SEPARATE CONDITIONS

In Test phase one bird (Walter) only participated for one session and eight trials while the other (Petra) completed eight sessions. Due to the small sample size, no statistical analysis was possible for Walter. Although Petra had reached criterion in the Training, she had an average success rate of 65% correct first choices in all baseline trials during test sessions and only exceeded 80% in two out of eight sessions. In Test 1 she reached 75% while being correct in 37.5% of the Test 2 trials and 50% in both Test 3 and Test 4 (Figure 14).

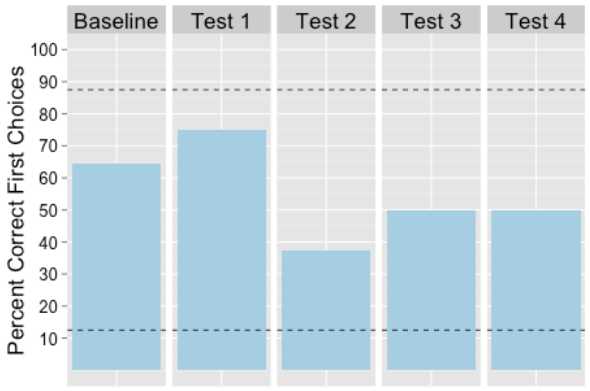

Figure 14: Percentage of correct first choices in each condition for Petra. The dashed horizontal lines indicate significant levels (above ≥ 87.5%; below ≥ 12,5%)

DISCUSSION

Experiment 1

With the current study, I have shown, that carrion crows use a variety of strategies when confronted with an artificial touch screen task, including IBE. Three out of five birds choose according to 'IBE' at some point in the experiment (Figure 8). On an individual level, one bird (Töffel) showed 'IBE' significantly more often than predicted by chance (Figure 7). Although, the small sample size has to be kept in mind, her preference for that pattern was prevalent in the first five sessions already (Figure 8).

One crow (Willi) showed significant effects for the two patterns 'one–trial-learning 3' (= incorrect in Test 4) and 'novelty preference & S+ preference' (= incorrect in Test 3) and a tendency for 'IBE' (= all trials correct) (Figure 7). This can either be explained by having three different response strategies simultaneously or that he might have experienced a trade-off between those patterns. For 'IBE', subjects were required to solve the 'reinforced' and the 'unenforced combination' concurrently. Willi was incorrect in either Test 3 or Test 4 but rarely in both (only in three out of 25 unique sessions as can be seen in Figure 8 in the response strategy 'S- avoidance & S+ preference'). This leads me to the assumption that following both lines of thought simultaneously, as required for 'IBE', might have been at the limit of his cognitive load. The related trials ('reinforced' and 'unenforced combination') being interspersed might have added to the level of complexity. Hence, parallel information storing cannot be excluded as a confounding factor in this study. However, the bird in question showed a better performance in sequence B compared to sequence A (see Figure 12). In sequence A Test 4 was paired with a negative stimulus the bird had just seen in Test 3. Thus, if sequence played an influential role, sequence A should have been easier to solve than sequence B. This was the case for all birds but Willi.

All other crows exhibited 'IBE' below significant levels. Also, in contrast to Goffin cockatoos and keas, carrion crows showed no significant pattern of 'IBE' on a group level (Figure 6).

These results stand in contrast to a previous study by Mikolasch and colleagues (2012), which found IBE in carrion crows on a group level in the food-finding task. Three birds participated in both the current and Mikolasch et al.'s (2012) study, which gives me the opportunity to compare the individual results directly. One bird (Töffel) showed significant outcomes in both studies, which supports her individual ability to reason by exclusion. Another (Bärchen) showed significant results in the study by Mikolasch and colleagues (2012), but not in the current study. This leaves us with two possible explanations: Firstly, it is possible that he has the ability to infer by exclusion, but that this was overshadowed by the complexity and/or nature of this experiment while present in the food-finding study. No significant preference for any response strategy could be observed in the current study (Figure 7). However in conditions where no significant result was observable, a side preference (for the right side) seemed to emerge. Except for baseline trials and Test 2 trials, where it was possible to stay with the learned and repeatedly rewarded S+, he seemed to have developed an alternative strategy of always choosing the right side.

Secondly, it is possible that he used other mechanisms such as avoidance of the empty cup in the food-finding task, which would have led to no IBE being observed in the first place. If this were true, one would expect him to show avoidance strategies in the current study as well. However, neither did his data show a pattern, which would hint at avoidance, nor was he strictly avoiding S- in Test 1. Nevertheless, this could also be due to the differences in set-ups.

The third bird (Petra) failed to show IBE in both experiments.

From this, one might argue that there are individual differences as have been identified in various other contexts (e.g. neophobia, exploration, aggression or sociability; see e.g. O'Hara et al. in prep) in the ability to reason by exclusion in carrion crows and that the compared studies differ in complexity and ecological relevance to facilitate its use.

Even though procedures differ, I would also like to compare the results to the previous touch screen experiment by Aust et al. (2008). One out of six pigeons, three out of six dogs and most of the human subjects showed a preference for the novel, but correct stimulus in Test 1 (similar to Test 1 in the current experiment). When analysed the same way, data for three out of four

crows revealed the same preference significantly above chance level (see Figure 9). Of the three individuals choosing that way in Test 1, two of the crows of the current study seemed to have a preference for the novel, but incorrect stimulus in Test 3 (which is comparable to Test 2 in the previous study), similar to one pigeon and two of the children. The authors argued that this pattern could be due to either novelty preference, avoidance of the familiar but incorrect stimulus, avoidance of the unrewarded but correct stimulus of Test 1 or stimulus preferences. Due to the modified procedure, avoidance of an unrewarded, but correct stimulus (hence violation of expectations) can be excluded as an explanation for the crows' choices. When looking at the individual level, those birds, who showed this preference, showed high levels of 'novelty preference' (Walter $p<0.001$) or the pattern 'novelty preference & S+ preference' (Petra $p=0.004$) (Figure 7). This leads me to the conclusion that a choice based on neophilic tendencies is most likely to explain the results in this comparison. However, two birds also showed tendencies ($p=0.067$) for patterns of 'S- avoidance' (Walter and Petra) and 'S- avoidance & S+ preference' (Walter) meaning that an additional avoidance strategy might have also played a role.

If stimulus preference had a strong overshadowing influence on the choices made, one would expect birds to choose the same stimulus in both sets (repetition sessions) even when it was not rewarded in the first set. This only occurred in 19.87% of test trials (Figure 13). However, this cannot exclude an initial stimulus preference, which can be overruled by learning. Surprisingly, crows chose different stimuli in the second set, although they were rewarded in the first set in 22.44% of all test trials. Two explanations come to mind why this could have happened. Firstly, the correct stimuli were chosen at random in set one without a preference for a stimulus and no one-trial-learning occurred as a result of those choices. Secondly, exploratory behaviour could have led to investigate feedback to both of the presented stimuli when presented repeatedly. Which processes played the major role cannot be disentangled by analysing the data.

I thus conclude that neither learning nor strict stimulus preferences exclusively explain differences in response between repeated stimulus presentations and therefore explorative behaviour might have played an influential role.

As mentioned earlier, both Goffin cockatoos (O'Hara et al. 2015) and keas (O'Hara et al. 2016) showed 'IBE' in the same task. Both species chose 'IBE' significantly more often than predicted by chance on both the group and individual level. Therefore, one can assume that methods are appropriate for at least some avian species. Similar to Psittacines, members of the corvid family also have relatively large brains, and therefore one might expect them to be able to infer by exclusion similarly well. This could only be shown on an individual level in the current study. It is, however, possible that parrots have ecological predispositions for recognizing and remembering more complex and colourful stimuli, as for fruit and nut eating species, determining and distinguishing the state and ripeness of fruit is more important than for scavengers who do not necessarily rely on fruit for their diet. Feeding mainly on meat and carrion, carrion crows are a caching species, and thus locating a food source plays a much more important role to them than the state it is in. This might have put them at a disadvantage compared to parrots in the current study. However, this could have only played a minor role in the weaker result in comparison. For one, crows showed the ability to differentiate between different stimuli as quickly as keas and Goffin cockatoos in Training. Also, remembering many different stimuli was not necessary for choosing correctly over time in the Test phase as only baseline stimuli were shown repeatedly over sessions.

Another reason for a weaker 'IBE' performance of crows compared to keas and Goffin cockatoos could be the lack of a novelty criterion. In both previous studies, individuals had to reach not only a training criterion but also a novelty criterion (two out of two correct first choices in novelty trials in two consecutive sessions) (O'Hara et al. 2015; 2016). Both studies showed a trend for learning discrimination of two stimuli faster than to withhold from selecting the novel stimulus. None of the crows would have reached said novelty criterion at the point of transition to the Test phase, as all of them where still choosing novel stimuli at least once in any two consecutive sessions during Training (Figure 5). Therefore, one might argue that a predisposition for neophilic tendencies was still present in the test subjects and thus might have overshadowed the ability to reason by exclusion. This was supported by significant levels of the pattern 'novelty preference' (Walter $p < 0.001$) and 'novelty preference & S+ preference'

(Petra p=0.004; Willi p<0.001) (Figure 7). Still, the later pattern was also very frequently observed in both Goffin cockatoos and keas, being assigned to 'S-2 preference'.

I believe attributing this pattern to a kind of novelty preference is as valid as ascribing it to a preference for a stimulus sequence as three of the test trials could be solved by a preference for the most novel stimulus. Only in Test 2, a different strategy had to be applied, but this might be the most parsimonious strategy of choosing the repeatedly rewarded baseline stimulus. Novelty preference might also be a valid explanation as to why the birds should choose incorrectly in Test 2 for picking the most novel stimulus over the reinforced stimulus. Indeed, Petra and Walter did not show a preference for S+ in Test 2 trials (Figure 9), although this choice was repeatedly rewarded in baseline trials.

Interestingly, all birds were more often correct in both test trials of the 'unenforced combination' (Test 2 and 4) than in the 'reinforced' one (Test 1 and 3) (Figure 11). One bird (Willi) significantly exceeded chance level in the 'unenforced combination', while none did so in the 'reinforced combination'. One might argue that this shows a preference for novel stimuli. Choice based on novelty would yield incorrect results in Test 2 and Test 3. In Test 2, however, the positive and rewarded stimulus from baseline trials was used, which most likely increased the affinity to it even if novelty preferences existed. This would explain why latencies, which one might argue could be seen as 'thinking time', are not significantly higher in Test 2 than in baseline trials. Test 3, on the other hand, can only be solved correctly when IBE or one-trial-learning had occurred in Test 1.

The results from the separate test trials support this argument as two birds (Petra and Walter) showed a significant preference for the novel, but negative stimulus in Test 3, even when the correct stimulus was rewarded in Test 1 (Figure 9). I therefore conclude that novelty preference is consistent in all analysis for those two birds, which supports the argument that explorative behaviour has played an influential role. However this does not exclude other strategies playing a role in the decision-making process.

If avoidance was used to solve Test 1, no information would be present to solve Test 3 and 4, except the bird would use one-trial-learning. Therefore, the possibility of choosing correctly

would be reduced to a 50/50 chance. Two of the crows with high levels of novelty preference also showed strong tendencies, even though not significant (p=0.067), towards 'S- avoidance' (Petra und Walter) and 'S- avoidance & S+ preference' (Walter) (Figure 7). This leads to the conclusion that a combination of both strategies might have occurred as a rule of thumb, i.e. "avoid negative and if you don't know what to choose next try the most novel stimulus".
In sum, I conclude that in the current study carrion crows were not sufficiently habituated to novelty in Training and therefore neophilic tendencies were still exhibited in Test phase, possibly masking IBE abilities.

Experiment 2

With the current study, the potential for carrion crows to find a hidden reward in a semi-natural setting based on acoustic characteristics of objects was demonstrated for the first time, although other underlying mechanism cannot be ruled out with certainty.
Unfortunately, I encountered methodical problems in the second experiment. First of all, the sample size of birds working in the testing area and with the boxes there was limited to three. This could have potentially been avoided by choosing a testing area that was more comfortable for the birds, for example in front of their home aviaries while the partner is in a different aviary. However, I decided against this, because it was not possible to visually separate this area from other birds. I expected the same four birds, who had worked at the touch screen already, to also participate in the second experiment as it was conducted in the same testing area.
Sample size was further reduced as one individual (Bärchen) stopped participating when boxes were filled with substrates. Additional habituation and smaller boxes might have helped at overcoming neophobia, which seemed to be the reason for not participating.
Furthermore, the birds' motivation seemed to decrease massively over the course of the sessions. Although they completed 20 trials in a row at the touch screen, at the second experiment, there were longer breaks in between trials, and I could observe the birds getting

impatient and turning their attention to something else early on. For this reason, the amount of trials was decreased to 10 later on.

In the Pre-Training, two birds (Petra and Bärchen) reached criterion in less than six sessions while it took the third bird (Walter) 40 sessions to reach criterion. One possible explanation would be the pre-experience of the two crows in the food-finding task (Mikolasch et al. 2012). Another reason for the long time needed by this one bird could have been loss of motivation, which was observed for him many times during the Pre-Training sessions.

In contrast to Experiment 1, both birds reached the training criterion in the minimum amount of sessions. Although one cannot exclude that the birds picked up on other (unintended) cues given or had a preference for the noisy substrates, this seems to show that they used acoustic information to find the hidden reward. Alternatively neophobic tendencies could have guided the crows' choices. For reasons mentioned above, 'noisy' substrates where manipulated during habituation, while 'silent' substrates were not. Therefore birds might have experienced less habituation to the 'silent' substrates, which could have resulted in an avoidance of the less familiar.

Only one crow (Petra) participated in enough test sessions to be analysed although due to only eight completed test session statistical analysis required almost perfect scores. Therefore she did not significantly exceed chance level in any of the conditions (Figure 14). Apart from Test 1, where a trend for correct choices could be observed, all choices seemed random. Therefore, one might argue that either motivational levels were influencing her performance or the set-up was not appropriate for this particular crow. It is possible that although no acoustic associations were formed during habituation, the bird would attribute some sound to 'silent' substrates. In this case, hearing no sound would mean that the reward was equally likely in one or the other substrate or in no substrate at all, and therefore, choices would have been made at random.

In Test 1, choice could have been based on pure association with the 'noisy' substrate. However, this should then also be true for Test 3, which was not the case. Alternatively novelty preference would explain high levels of correct choices in Test 1 but this theory is not consistent with the results of Test 3 and Test 4 (as no preference for the novel substrate was observed). This leaves us with the possibility of avoidance of the 'silent' substrate, which was never

rewarded in baseline trials. Indeed a weak trend in Test 2 for the noisy substrate could be observed, which fits the theory of avoidance (see Table 4 and Figure 14). However this could also be due to a positive association towards the 'noisy' baseline substrate alone.
In total I have to acknowledge that, without increasing the number of sessions as well as sample size of subjects, interpretations of choices in Test phase remain speculative. However, neophobic tendencies possibly led the birds to not participate and might have been in part responsible for ceiling scores in Training. Future studies should address problems as mentioned above to motivate birds to participate.

General discussion

In the current study I showed that carrion crows exhibit a novelty preference as a frequently used response strategy to the experimental set-up of experiment 1, whereas some of the methodical difficulties in experiment 2 might have resulted from the birds' neophobic tendencies. One might conclude from this that a well-known artificial set-up encourages carrion crows to explore responses to unfamiliar stimuli. Neophobic responses can be triggered by the violation of expectancies but given that all birds have been tested in touch screen experiments with changing images before, the presentation of an unfamiliar stimulus can be expected by the individuals. They have learned over time that no harm comes from new stimuli and costs of explorative behaviour a very low, especially in test trials. The set-up of experiment 2, in contrast, was newly introduced to the crows in this form. Although I aimed to sufficiently habituate the birds, this might not have been enough for some individuals. Additionally, as discussed above, one cannot exclude the possibility that the subjects chose the 'noisy' substrate in Training solely because they were more familiar to their acoustic characteristics.
Ambivalent responses to novelty have been reported repeatedly for members of the corvid family (Greenberg 2003). They are generally known to be neophobic while at the same time flexible in new environments and innovative problem-solvers (Miller et al. 2015). In the case of the carrion crow, the species inhabits not only the rural areas all over Europe but also urban areas (Richner 1989). Exploiting new environments is possibly facilitated by the right balance of neophilia and neophobia – being curious but careful at the same time.

Studies like Mikolasch et al. (2012) show that simpler mechanisms, such as local enhancement, can overshadow other more cognitively demanding skills. I hypothesize that this might have been the case in the current study and that more sufficient habituation needs to be applied to overcome initial propensities (e.g. implementing a novelty criterion in Training). Authors such as Greenberg (2003) argue that neophobia and neophilia are attributes closely linked to the ecology of species. Therefore I believe both experiments, taken together, highlight the importance of taking a species predisposition into account when testing for higher cognitive abilities, such as IBE.

Future ideas

For testing the hypothesis that explorative behaviour might have detained the exhibition of IBE, an interesting follow-up study would be to test an additional set of crows with the described novelty criterion and to investigate whether this changes their response strategies.
Furthermore, I believe the testing paradigm developed by O'Hara et al. (2015; 2016) has great potential for comparative research on IBE and alternative mechanisms. For instance, it allows a variation of complexity by changing sequences or test trial positions. Tests can also be conducted at different developmental stages to get insight into when the ability to reason by exclusion emerges. I have recently been part of such a study on juvenile parrots (see appendix for preliminary results).
Touch screen approaches have many advantages, including no possibility of cueing by the experimenter and being less error-prone. On the other hand, such approaches require special equipment and quite extensive habituation; they are also less ecologically relevant.
However, possibilities of the described testing paradigm are not restricted to a touch screen set-up. For example, Smirnova et al. (2015) showed that using cards with printed symbols can be an approach to conduct experiments with a great variety of visual stimuli in carrion crows. Therefore, I think it might be possible to transfer the described testing paradigm to a set-up which can be used for experiments without touch screens and would thus be more feasible for a wider variety of research facilities. Also, this might increase ecological relevance as the location of food needs to be inferred, even though still artificial stimuli are used.

REFERENCES

Albiach-Serrano, A.; Bräuer, J.; Cacchione, T.; Zickert, N.; Amici, F. (2012): The effect of domestication and ontogeny in swine cognition (Sus scrofa scrofa and S. s. domestica). In: *Applied Animal Behaviour Science* 141 (1-2), S. 25–35. DOI: 10.1016/j.applanim.2012.07.005.

Aust, U.; Range, F.; Steurer, M.; Huber, L. (2008): Inferential reasoning by exclusion in pigeons, dogs, and humans. In: *Anim Cogn* 11 (4), S. 587–597. DOI: 10.1007/s10071-008-0149-0.

Beran, M. J. (2010): Use of exclusion by a chimpanzee (Pan troglodytes) during speech perception and auditory–visual matching-to-sample. In: *Behavioural Processes* 83 (3), S. 287–291. DOI: 10.1016/j.beproc.2010.01.009.

Beran, M. J.; Washburn, D. A. (2002): Chimpanzee responding during matching to sample: control by exclusion. In: *J Exp Anal Behav* (78), S. 497–508.

Bräuer, J.; Kaminski, J.; Riedel, J.; Call, J.; Tomasello, M. (2006): Making Inferences About the Location of Hidden Food: Social Dog, Causal Ape. In: *Journal of Comparative Psychology* 120 (1), S. 38–47. DOI: 10.1037/0735-7036.120.1.38.

Call, J. (2004): Inferences About the Location of Food in the Great Apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, and Pongo pygmaeus). In: *Journal of Comparative Psychology* 118 (2), S. 232–241. DOI: 10.1037/0735-7036.118.2.232.

Call, J. (2006): Inferences by exclusion in the great apes: the effect of age and species. In: *Anim Cogn* 9 (4), S. 393–403.

Clement, T. S.; Zentrall, T. R. (2003): Choice based on exclusion in pigeons. In: *Psychonomic Bulletin & Review* (10 (4)), S. 959–964.

Emery, N. J. (2006): Cognitive ornithology: the evolution of avian intelligence. In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 361 (1465), S. 23–43. DOI: 10.1098/rstb.2005.1736.

Erdőhegyi, Á.; Topál, J.; Virányi, Z.; Miklósi, Á. (2007): Dog-logic: inferential reasoning in a two-way choice task and its restricted use. In: *Animal Behaviour* 74 (4), S. 725–737. DOI: 10.1016/j.anbehav.2007.03.004.

Greenberg, R. (2003): The role of neophobia and neophilia in the development of innovative behaviour of birds. In: Simon M. Reader & Kevin N. Laland (Hg.): Animal innovation. Oxford, New York: Oxford University Press, S. 175–196.

Güntürkün, O.; Bugnyar, T. (2016): Cognition without Cortex. In: *Trends in Cognitive Sciences* 20 (4), S. 291–303. DOI: 10.1016/j.tics.2016.02.001.

Heimbauer, L. A.; Antworth, R. L.; Owren, M. J. (2012): Capuchin monkeys (Cebus apella) use positive, but not negative, auditory cues to infer food location. In: *Anim Cogn* 15 (1), S. 45–55. DOI: 10.1007/s10071-011-0430-5.

Herman, L. M.; Richards, D. G.; Wolz, J. P. (1984): Comprehension of sentences by bottlenosed dolphins. In: *Cognition* 16 (2), S. 129–219. DOI: 10.1016/0010-0277(84)90003-9.

Hill, A.; Collier-Baker, E.; Suddendorf, T. (2011): Inferential reasoning by exclusion in great apes, lesser apes, and spider monkeys. In: *Journal of Comparative Psychology* 125 (1), S. 91–103. DOI: 10.1037/a0020867.

Jelbert, S. A.; Taylor, A. H.; Gray, R. D. (2015): Reasoning by exclusion in New Caledonian crows (Corvus moneduloides) cannot be explained by avoidance of empty containers. In: *Journal of Comparative Psychology* 129 (3), S. 283–290. DOI: 10.1037/a0039313.

Kaminski, J.; Call, J.; Fischer, J. (2004): Word Learning in a Domestic Dog: Evidence for "Fast Mapping". In: *Science* 304 (5677), S. 1682–1683. DOI: 10.1126/science.1097859.

Kastak, C. R.; Schusterman, R. J. (2002): Sea lions and equivalence: expanding classes by exclusion. In: *J Exp Anal Behav* 78 (3), S. 449–465.

Maille, A.; Roeder, J. J. (2012): Inferences about the location of food in lemurs (Eulemur macaco and Eulemur fulvus): a comparison with apes and monkeys. In: *Anim Cogn* 15 (6), S. 1075–1083. DOI: 10.1007/s10071-012-0531-9.

Marsh, H. L.; MacDonald, S. E. (2012): Information seeking by orangutans: a generalized search strategy? In: *Anim Cogn* 15 (3), S. 293–304. DOI: 10.1007/s10071-011-0453-y.

Marsh, H.L.; Vining, A. Q.; Levendoski, E.K.; Judge, P.G. (2015): Inference by exclusion in lion-tailed macaques (Macaca silenus), a hamadryas baboon (Papio hamadryas), capuchins (Sapajus apella), and squirrel monkeys (Saimiri sciureus). In: *Journal of Comparative Psychology* 129 (3), S. 256–267. DOI: 10.1037/a0039316.

Mikolasch, S.; Kotrschal, K.; Schloegl, C. (2011): African grey parrots (Psittacus erithacus) use inference by exclusion to find hidden food. In: *Biology Letters* 7 (6), S. 875–877. DOI: 10.1098/rsbl.2011.0500.

Mikolasch, S.; Kotrschal, K.; Schloegl, C. (2012): Is caching the key to exclusion in corvids? The case of carrion crows (Corvus corone corone). In: *Anim Cogn* 15 (1), S. 73–82. DOI: 10.1007/s10071-011-0434-1.

Miller, R.; Bugnyar, T.; Pölzl, K.; Schwab, C. (2015): Differences in exploration behaviour in common ravens and carrion crows during development and across social context. In: *Behav Ecol Sociobiol. DOI:* 10.1007/s00265-015-1935-8.

Nawroth, C.; von Borell, E.; Langbein, J.; McElligott, A. (2014): Exclusion Performance in Dwarf Goats (Capra aegagrus hircus) and Sheep (Ovis orientalis aries). In: *PLoS ONE* 9 (4), S. e93534. DOI: 10.1371/journal.pone.0093534.

O'Hara, M.; Auersperg, A. M. I.; Bugnyar, T.; Huber, L.; Boraud, T. (2015): Inference by Exclusion in Goffin Cockatoos (Cacatua goffini). In: *PLoS ONE* 10 (8), S. e0134894. DOI: 10.1371/journal.pone.0134894.

O'Hara, M.; Schwing, R.; Federspiel, I.; Gajdon, G. K.; Huber, L. (2016): Reasoning by exclusion in the kea (Nestor notabilis). In: *Anim Cogn. DOI:* 10.1007/s10071-016-0998-x.

Paukner, A.; Huntsberry, M. E.; Suomi, S. J. (2009): Tufted capuchin monkeys (Cebus apella) spontaneously use visual but not acoustic information to find hidden food items. In: *Journal of Comparative Psychology* 123 (1), S. 26–33. DOI: 10.1037/a0013128.

Pepperberg, I. M.; Koepke, A.; Livingston, P.; Girard, M.; Hartsfield, L. A. (2013): Reasoning by inference: Further studies on exclusion in grey parrots (Psittacus erithacus). In: *Journal of Comparative Psychology* 127 (3), S. 272–281. DOI: 10.1037/a0031641.

Petit, O.; Dufour, V.; Herrenschmidt, M.; De Marco, A.; Sterck, E. H.; Call, J. (2015): Inferences about food location in three cercopithecine species: an insight into the socioecological cognition of primates. In: *Anim Cogn* 18 (4), S. 821–830. DOI: 10.1007/s10071-015-0848-2.

Piaget, J. (1952): The origins of intelligence in children. (Original work published in 1936) - so in Zucca. New York: International Universities Press.

Plotnik, J. M.; Shaw, R. C.; Brubaker, D. L.; Tiller, L. N.; Clayton, N. S. (2014): Thinking with their trunks: elephants use smell but not sound to locate food and exclude nonrewarding alternatives. In: *Animal Behaviour* 88, S. 91–98. DOI: 10.1016/j.anbehav.2013.11.011.

Premack, D. (1995): Cause/induced motion: intention/spontaneous motion. In: Jean-Pierre Changeux & Jean Chavaillon (Hg.): Origins of the human brain. Oxford, New York: Clarendon Press; Oxford University Press (A Fyssen Foundation symposium), S. 286–308.

Premack, D.; Premack, A. J. (1994b): Levels of causal understanding in chimpanzees and children. In: *Cognition* 50 (1-3), S. 347–362. DOI: 10.1016/0010-0277(94)90035-3.

R Core Team (2014): R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. Online verfügbar unter http://www.R-project.org/.

Richner, H. (1989): Habitat-Specific Growth and Fitness in Carrion Crows (Corvus corone corone). In: *Journal of Animal Ecology* 58 (2), S. 427–440.

Riehl, C.; Strong, M. J.; Edwards, S.V. (2015): Inferential reasoning and egg rejection in a cooperatively breeding cuckoo. In: *Anim Cogn* 18 (1), S. 75–82. DOI: 10.1007/s10071-014-0778-4.

Sabbatini, G.; Visalberghi, E. (2008): Inferences about the location of food in capuchin monkeys (Cebus apella) in two sensory modalities. In: *Journal of Comparative Psychology* 122 (2), S. 156–166. DOI: 10.1037/0735-7036.122.2.156.

Schloegl, C.; Bugnyar, T.; Aust, U. (2009a): Exclusion performances in non-human animals: from pigeons to chimpanzees and back again. In: *Rational animals, irrational humans. Keio University Press, Tokyo*, S. 217–234.

Schloegl, C.; Schmidt, J.; Boeckle, M.; Weiss, B. M.; Kotrschal, K. (2012): Grey parrots use inferential reasoning based on acoustic cues alone. In: *Proceedings of the Royal Society B: Biological Sciences* 279 (1745), S. 4135–4142. DOI: 10.1098/rspb.2012.1292.

Schloegl, C. (2011): What you see is what you get—Reloaded: Can jackdaws (Corvus monedula) find hidden food through exclusion? In: *Journal of Comparative Psychology* 125 (2), S. 162–174. DOI: 10.1037/a0023045.

Schloegl, C.; Dierks, A.; Gajdon, G. K.; Huber, L.; Kotrschal, K.; Bugnyar, T.; Iwaniuk, A. (2009b): What You See Is What You Get? Exclusion Performances in Ravens and Keas. In: *PLoS ONE* 4 (8), S. e6368. DOI: 10.1371/journal.pone.0006368.

Schmitt, V.; Fischer, J. (2009): Inferential reasoning and modality dependent discrimination learning in olive baboons (Papio hamadryas anubis). In: *Journal of Comparative Psychology* 123 (3), S. 316–325. DOI: 10.1037/a0016218.

Shaw, R. C.; Clayton, N. S. (2012): Careful cachers and prying pilferers: Eurasian jays (Garrulus glandarius) limit auditory information available to competitors. In: *Proceedings of the Royal Society B: Biological Sciences* 280 (1752), S. 20122238.

Shaw, R. C.; Plotnik, J. M.; Clayton, N. S. (2013): Exclusion in corvids: The performance of food-caching Eurasian jays (Garrulus glandarius). In: *Journal of Comparative Psychology* 127 (4), S. 428–435. DOI: 10.1037/a0032010.

Smirnova, A.; Zorina, Z.; Obozova, T.; Wasserman, E. (2015): Crows Spontaneously Exhibit Analogical Reasoning. In: *Current Biology* 25 (2), S. 256–260. DOI: 10.1016/j.cub.2014.11.063.

Stulp, G.; Emery, N. J.; Verhulst, S.; Clayton, N. S. (2009): Western scrub-jays conceal auditory information when competitors can hear but cannot see. In: *Biology Letters* 5 (5), S. 583–585. DOI: 10.1098/rsbl.2009.0330.

Tomonaga, M. (1993): Tests for control by exclusion and negative stimulus relations of arbitrary matching to sample in a "symmetry-emergent" chimpanzee. In: *J Exp Anal Behav* (59), S. 215–229.

Tornick, J. K.; Gibson, B. M. (2013): Tests of inferential reasoning by exclusion in Clark's nutcrackers (Nucifraga columbiana). In: *Anim Cogn* 16 (4), S. 583–597. DOI: 10.1007/s10071-013-0595-1.

Wickham, H. (2009): Ggplot2. Elegant graphics for data analysis. Dordrecht, New York: Springer (Use R!).

Zaine, I.; Domeniconi, C.; Costa, A. R. A. (2014): Exclusion performance in visual simple discrimination in dogs (Canis familiaris). In: *Psychology & Neuroscience* 7 (2), S. 199–206. DOI: 10.3922/j.psns.2014.014.

LIST OF FIGURES AND TABLES

Figure 1: Outline of housing and experimental area: Aviaries were arranged in a circle around an inner testing area, where the experiments were carried out. Four of the five aviaries were used for housing. *12*

Figure 2: a) A crow choosing a stimulus by pecking at the touch screen in experiment 1. b) A crow making a choice by touching a box in experiment 2. ..... *12*

Figure 3 a),b): Habituation to substrates in boxes in front of an aviary. *21*

Figure 4: Performance in Training. Coloured lines indicate individual learning curves. The dashed line represents performance of 80% correct first choices. The length of individual lines indicate the amount of sessions needed to reach the Training criterion. *24*

Figure 5: Cumulative errors in novelty trials in Training sessions. The dotted line indicates the slope of choosing the novel stimulus one out of two possible times per session (chance level). Horizontal lines indicate correct choice of S+. The length of individual lines is restricted to reaching Training criterion. The steeper the slope, the more novel stimuli were chosen. *24*

Figure 6: Response strategies on the group level for four individuals. Boxes include the first to the third quartile, whiskers represent 95% confidence interval, dots show outliners, vertical lines are medians. Values to the right of the dashed line are significantly exceeding chance level. *25*

Figure 7: Response strategies per individual. Differently coloured boxes show the percentage for the occurrence of every pattern. Numbers in boxes represent p-values for the cumulative probability of each pattern and bolded numbers have a p-value smaller than 0.05 (Bärchen, Petra, Walter, Willi n=25; Töffel n=5). *26*

Figure 8: Response strategies of each session per bird. *27*

Figure 9: Percentage of correct first choices in each condition. Coloured boxes represent individuals. Values above the upper dashed horizontal line indicate a significant preference for positive stimuli (≥72%), while values below the lower dashed line represent a preference for negative stimuli (≤28%) significantly above chance level for each individual and test trial. *28*

Figure 10: Response latencies are shown for each condition and all birds taken together (=ALL) as well as every individual separately. Boxes include the first to the third quartile, whiskers represent 95% confidence interval, dots show outliners, horizontal lines are medians. *28*

Figure 12: Percentage of correct first choices in the first five unique sessions comparing Test 4a (= light blue) and Test 4b (= dark blue). *29*

Figure 13: Comparison of choice variance in test trails of set 1 (first unique sessions) and set 2 (repeated sessions) *30*

Figure 14: Percentage of correct first choices in each condition for Petra. The dashed horizontal lines indicate significant levels (above ≥ 87.5%; below ≥ 12,5%).....31

Table 1: Details of subjects and participation in experiments. 11

Table 2: Overview of a test session with example stimuli. Response strategies for choosing the correct stimulus are given for each trial separately. Response strategies in bold and colour are the most likely ones applied if birds chose correctly also in all previous test trials. The abbreviation OTL stands for 'one–trial-learning'.....16

Table 3: Overview of all 16 possible patterns with the most parsimonious explanations. ✓ *indicates the correct choice of the positive stimulus (shown on the left side of each pair) while* ✗ *represents incorrect choices (choosing the negative stimulus, which are shown on the right side of each pair).* 19

Table 4: Overview of test trials in Experiment 2. Response strategies for choosing the correct substrate are given for each trial separately. Response strategies in bold and colour are the most likely ones applied if birds chose correctly also in all previous test trials. The abbreviation OTL stands for 'one–trial-learning'. 23

APPENDIX

After finishing the data collection for the described study I was invited to conduct a follow-up study with juvenile parrots (one and two years of age) of three different species: African grey parrots (*Psittacus erithacus*), blue-throated macaws (*Ara glaucogularis*) and great green macaws (*Ara ambiguus*). The following part shows preliminary results in comparison to carrion crows.

All birds of each species reached training criterion in two to ten sessions. Great green macaws were fastest to reach criterion and significantly faster than blue-throated macaws (pairwise two-sided Mann-Whitney test, corrected after Bonferroni: p=0.034) (Figure A1).

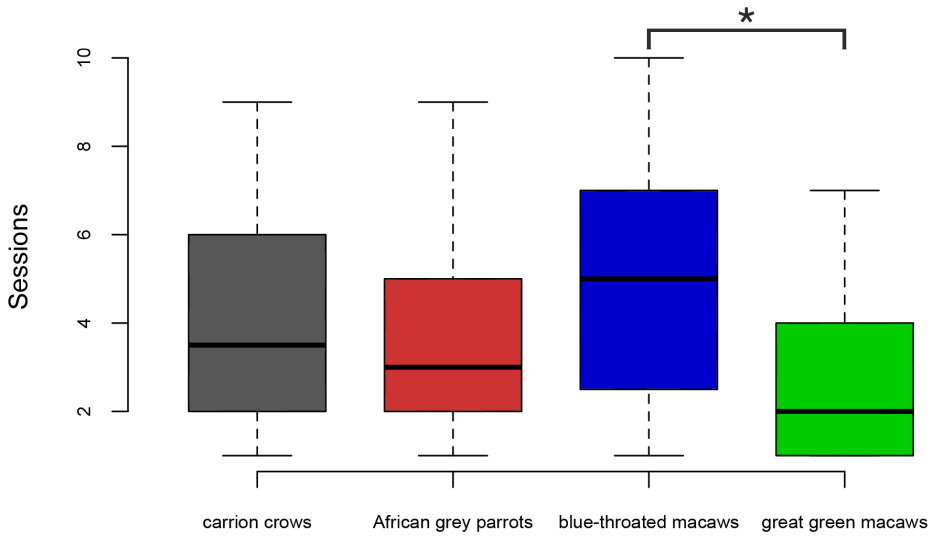

*Figure A1: Sessions needed to reach training criterion. Coloured boxes show results for each species. Boxes include the first to the third quartile, whiskers represent 95% confidence interval, dots show outliners, horizontal lines are medians. * ≤ p=0.05*

For the juvenile parrots a second criterion in Training was implement like it was used for Goffin cockatoos and keas (O'Hara 2015; 2016). This criterion was reached when novel stimuli were rejected in two out of two times in novelty trials in two consecutive sessions. All birds of each species reached this criterion in 3 to 17 sessions. Great green macaws were significantly faster to reach this criterion than both blue-throated macaws (pairwise two-sided Mann-Whitney test: p=0.004) and African grey parrots (pairwise two-sided Mann-Whitney test: p=0.021) (Figue A2). P-values were corrected after Bonferroni.

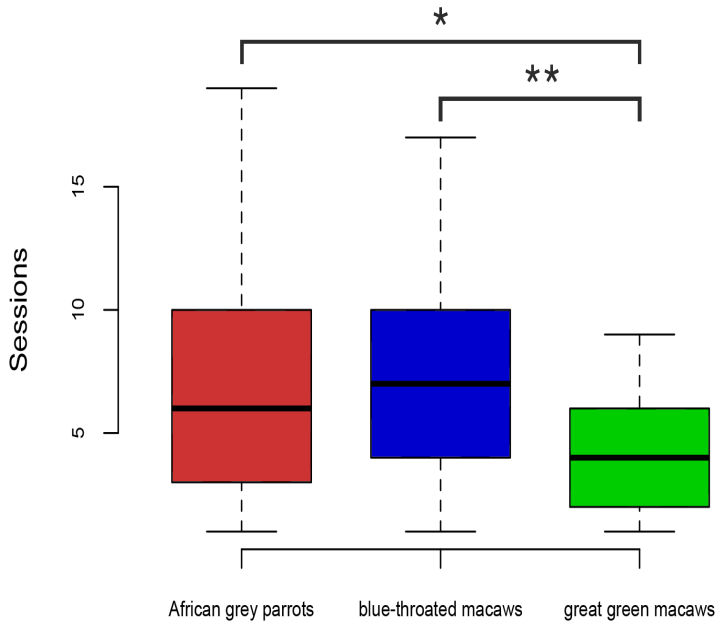

*Figure A2: Sessions needed to reach novelty criterion. Coloured boxes show results for each species. Boxes include the first to the third quartile, whiskers represent 95% confidence interval, dots show outliners, horizontal lines are medians. ** ≤ p=0.01 > * ≤ p=0.05*

Birds tested with trial sequence as described above for the 25 unique sessions in carrion crows (sequence B) were here attributed to 'Group B'. For juvenile parrots an additional 'Group A' was added, in which the order of Test 2 and 3 were switched. This resulted in a sequence where connected test trials were not interspersed and shown after a shorter period of time. In contrast to the crows parrots received 30 test sessions. By using two-sided binomial tests, I calculated which patterns occurred more frequently than predicted by chance.

On an individual level one carrion crow and two great green macaws showed the pattern for 'IBE' more often than predicted by chance. All three were assigned to 'Group B' (Figure A3). With the exception of one carrion crow (Willi) 'one-trial-learning' was only used in 'Group A'. In great green macaws 'one-trial-learning 3' (Luna) and 'one-trial-learning 4' (Luna and Enya) could be observed. Two blue-throated macaws (Charley and Long John) exhibited 'one-trial-learning 1'. No African grey parrot used 'IBE' or 'one-trial-learning' more often than predicted by chance. Other response strategies exhibited by at least three birds include 'novelty preference', 'novelty preference & S+ preference' and 'S- avoidance & S+ preference'.

Additionally two birds each used of the following patterns: ‘S- avoidance’, ‘S-1 avoidance and/or S+1 avoidance’ and ‘NA & S+1 avoidance and/or NA & S-2 preference’.

One African grey parrot (Jack) used ‘novelty aversion’ to solve the task.

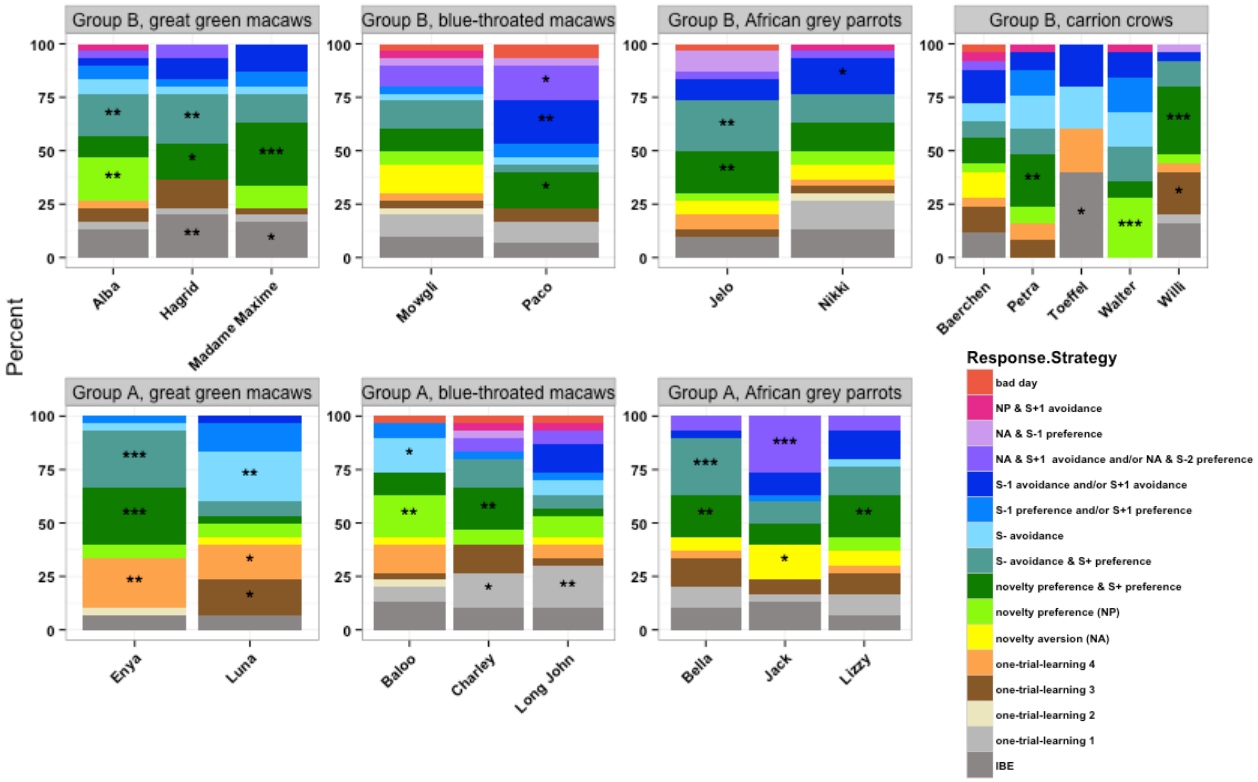

*Figure A3: Response strategies for individuals of each species and each group. Differently coloured boxes show the percentage for the occurrence of every pattern. *** ≤ p=0.001 > ** ≤ p=0.01 > * ≤ p=0.05*

On the group level ‘novelty preference & S+ preference’ was used more often than predicted by chance by African grey parrots, carrion crows and great green macaws (Figure A4). Great green macaws also used ‘S- avoidance & S+ preference’ significantly above chance.

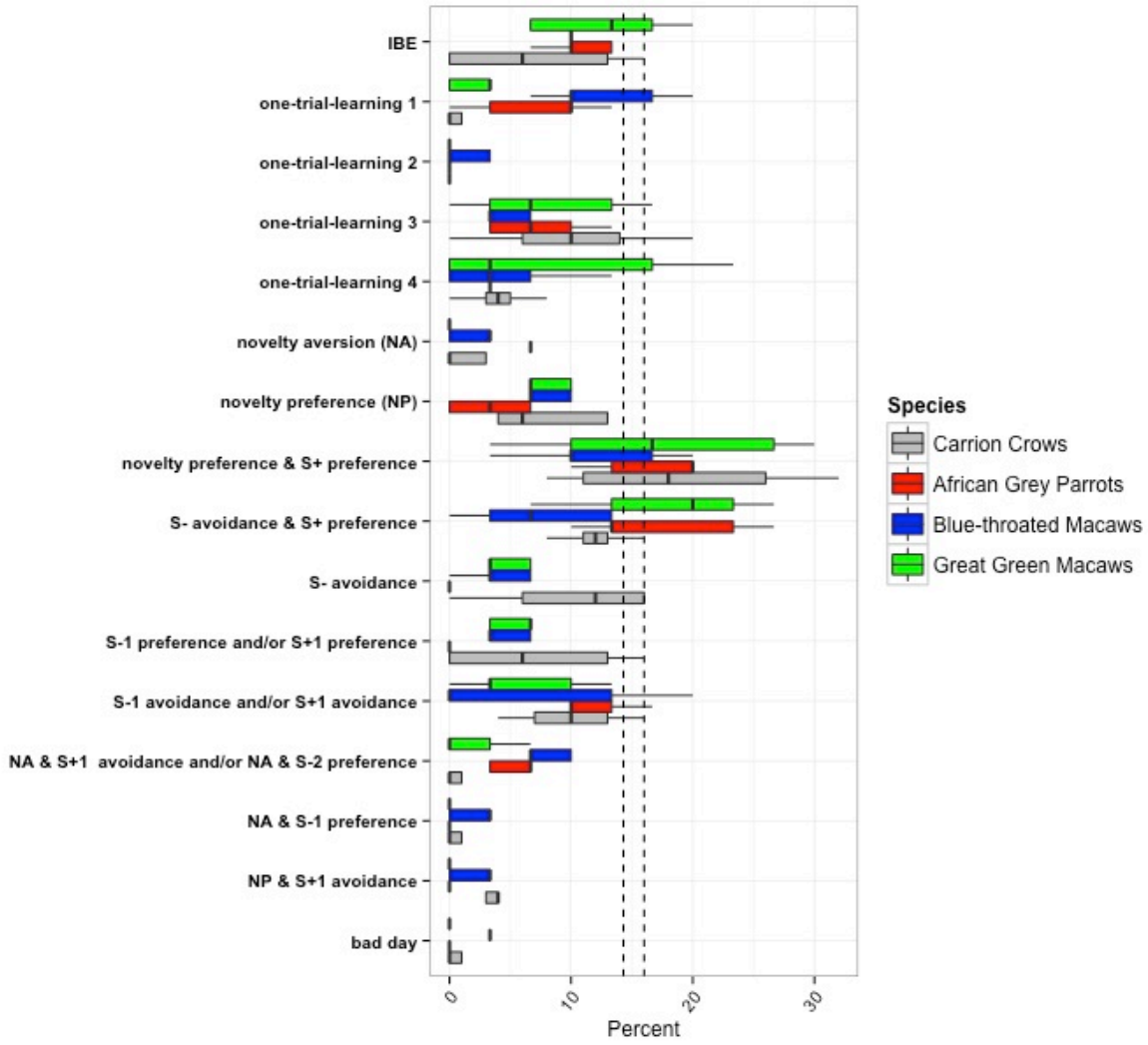

Figure A4: Response strategies at group level for each species. Boxes include the first to the third quartile, whiskers represent 95% confidence interval, dots show outliners, vertical lines are medians. Values right to the dashed lines are significantly exceeding chance level (right dashed line: significance level for crows; left dashed line: significance level for parrots)

Looking at the average percentage of correct first choices for each separate condition in test sessions, one can observe species as well as individual difference in the subjects' performance (Figure A5). While all of the great green macaws and all but one carrion crow exceeded chance level significantly in Test 1, only two African grey parrot and two blue-throated macaw did so. On the other hand all five African grey parrots and four blue-throated macaws performed significantly better than chance in Test 2, while only two birds of the other species did.
Two great green macaws, two carrion crows and one blue-throated macaw showed a preference for the negative stimulus in Test 3. While none of the carrion crows had a significant

preference for Test 4, two great green macaws, two blue-throated macaws and one African grey parrot chose the correct stimulus significantly more often than predicted by chance.

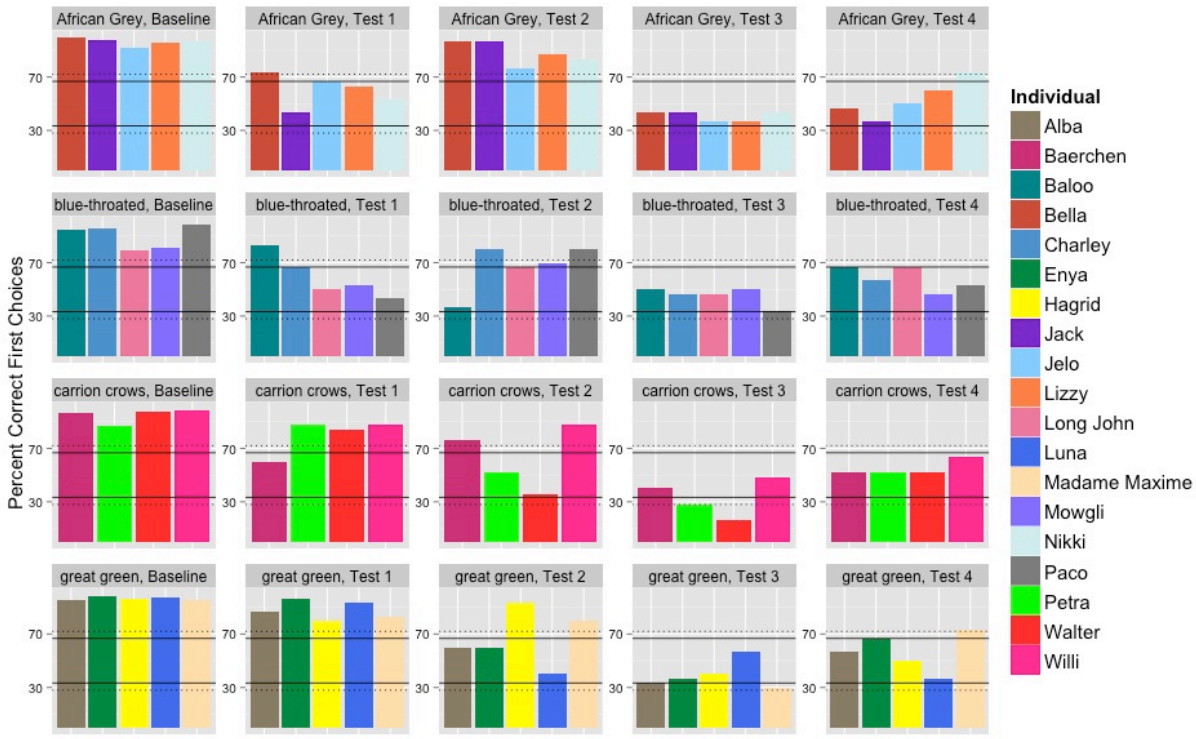

Figure A5: Performance of correct first choice in all conditions during Test phase for each individual of every species. Coloured boxes represent individuals. Significance levels were calculated using one-sided binomial tests. Values above the upper horizontal line (dotted line for crows) indicate a significant preference for positive stimuli, while values below the lower horizontal line (dotted line for crows) represent a preference for negative stimuli significantly above chance level.