MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Is it light or heavy? Weight discrimination of identical objects in Goffin cockatoos (*Cacatua goffiniana*)"

verfasst von / submitted by

Alexandra Stiegler, BSc

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

Master of Science (MSc)

Wien, 2021 / Vienna, 2021

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 878
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Verhaltens-, Neuro- und Kognitionsbiologie
Betreut von / Supervisor:	Mag. Dr. Sabine Tebbich, Privatdoz.

Abstract

Weight perception should be important for species which engage in extractive foraging or tool use. Studies focusing on weight discrimination learning are rare and largely focused on primate species. Perceiving weight differences should be important for some birds, perhaps particularly for parrots since many psittacids are extractive foragers and some even have shown the capacity to use tools. This study tests the ability to discriminate visually identical objects based on weight alone in a tool using parrot, the Goffin's cockatoo (*Cacatua goffiana*). Subjects were given a choice task with two spherical objects, where either the weighted object or the light one was rewarded when placed into a dish. One group was given pre-experience with objects that differed in colour as well as weight before receiving test trials with identical objects, while the other group began directly with the visually identical objects. The results showed that birds in both groups were able to learn to discriminate between visually identical objects based on weight. Nevertheless, birds learned to discriminate between objects faster if they received colour cues alongside the weight cues. After gaining pre-experience with the colour cues, subjects were better at discriminating between the identical objects than the birds that did not receive training before. This suggests that the birds learned about the objects weight even when knowledge about weight was not required for completing the task successfully. Comparing the results of the Goffin's cockatoos with primates the birds seemed to learn much faster, however differences in methodology should be kept in mind.

Zusammenfassung

Die Wahrnehmung von Gewicht ist ein wichtiger Faktor für Spezies die Nahrung extrahieren oder Werkzeuge bauen. Studien die das Lernen von Gewichtsunterscheidung untersuchen sind selten und wurden meist nur an Primaten durchgeführt. Die Wahrnehmung von Gewicht sollte ein wichtiger Faktor für Vogelarten sein, vielleicht besonders für Papageien, weil viele Psittacidae Nahrung extrahieren oder manche Vertreter Werkzeugbau gezeigt haben. Diese Studie untersucht die Fähigkeit einer Vogelspezies visuell idente Objekte anhand ihres Gewichtes zu unterscheiden. Goffin's Kakadus (*Cacatua goffiana*) bekamen die Wahl zwischen zwei kugelförmigen Objekten, wobei entweder das schwere oder das leichte Objekt bei Platzierung in einer Schüssel belohnt wurde. Eine Gruppe von Testsubjekten sammelte Erfahrung mit Objekten, die sich sowohl in Farbe als auch Gewicht

unterschieden, bevor sie Testversuche mit identen Objekten erhielt, während die andere Gruppe sofort mit den visuell identen Objekten begann. Die Ergebnisse zeigten, dass die Vögel in beiden Gruppen die visuell identen Objekte anhand des Gewichts unterscheiden konnten. Die Vögel lernten schneller, wenn zusätzliche Farbhinweise vorhanden waren. Nachdem eine Gruppe Erfahrung mit Farbhinweisen sammeln konnten, war sie besser darin die Objekte zu unterscheiden als die Gruppe ohne Vorerfahrung. Dies lässt vermuten, dass die Vögel über den Gewichtsunterschied lernten obwohl diese Information nicht benötigt wurde, um den Versuch erfolgreich zu beenden. Bei Vergleichen zwischen den Ergebnissen der Goffin's Kakadus mit den Primaten scheinen die Vögel schneller zu lernen Objekte anhand ihres Gewichts zu unterscheiden. Allerdings sollten bei solchen Vergleichen die Differenzen zwischen den Methoden der Versuche berücksichtigt werden.

Acknowledgements

I would like to thank Alice Auersperg for her co-supervision and help carrying out this study and giving me the opportunity to work with the birds at the Goffin Lab Goldegg. I would like to thank Poppy Lambert for her help and advice during the whole project. Furthermore, I am grateful to Theresa Rößler and Roger Mundry for their assistance in statistics. I want to express my gratitude towards my parents for allowing me to follow my dreams. Finally, I would like to thank the amazing birds at the Goffin Lab for their participation during the experiment.

Table of Contents

1. Introduction

In nature, the choice of an object of appropriate weight can influence the efficiency and success of some animal behaviours. For example, when extractive foraging, individuals might pay attention to the proprioceptive feedback elicited by the weight of an object while choosing among nuts and seeds. Nuts or seed kernels can vary in size or be absent from the shell, therefore the weight of the food item might best indicate its nutritional value as there is no visual access to the food source. Experimental research has confirmed the relevance of weight in foraging decisions. Two tufted capuchins (*Cebus paella*) chose the heavier nut from a pair (Visalberghi & Néel, 2003). Black-capped chickadees (*Parus atricapillusa*) and Red-breasted Nuthatches (*Sitta canadensis*) discarded more light seeds than heavy, when presented simultaneously with normal seeds and empty seeds, or normal seeds alongside seeds made artificially heavier (Heinrich *et al*., 1997). When offered the same comparison with peanuts, Mexican Jays (*Aphelocoma ultramarina*) similarly chose more heavy nuts than light (Jablonski *et al*. 2015).

Tool-using species might benefit from considering the weight of their tools as well. For the skilled technique of nut-cracking, heavier stones or hammers are more desirable because they will crack a shell with fewer strikes (Schrauf *et al*., 2008, Schrauf *et al*., 2012). Commonly, the size of an object is associated with its weight, especially if differently sized objects are made of the same material. In humans, the phenomenon of prediction of an objects weight given its size is well explored (Robinson, 1964). However, the direct assessment of an objects weight is only possible by handling an object to gain proprioceptive feedback. Wild capuchin monkeys (*Cebus libidinosus*) move stones around more if they discover that larger stones are lighter than smaller ones while searching for suitable stones to crack nuts with (Visalberghi *et al*., 2009).

Classical discrimination learning tasks consist of tasks where a subject needs to choose between two or more objects based on one differentiating property until they are reliably able to make the correct choice. Discrimination learning tasks focusing on weight-related proprioceptive information have largely been carried out with primate species. In one early study, three species of monkey were trained to differentiate between two visually identical boxes which differed in weight (Klüver, 1933). The monkeys needed to pull differently weighted boxes towards themselves. All test subjects had the same weight during the task (450g vs. 150g). Long-tailed macaques required a median of 586 trials, spider monkeys a median of 330 trials, and a capuchin monkey 321 trials to learn to consistently differentiate between the objects. For a similar task, chimpanzees (*Pan troglodytes*) were trained to choose the heavier box of two visually identical boxes. The five subjects required a median of 1,100 trials to reach a stable performance (McCulloch, 1941). A weight-based sorting task was carried out with chimpanzees. In this task, the subjects needed to place one object into a tray of the correct colour according to the object's weight. The criterion of placing the object correctly in 45 out of 50 trials over 5 sessions was only reached by six out of seven subjects. The subjects required 895.2 trials on average to reach this level of performance (Povinelli, 2011). A studie with orangutans (*Pongo pygamaeus abelii*), bonobos (*Pan paniscus*) and gorillas (*Gorilla gorilla gorilla*) rewarded the subjects for exchanging either the heavy or light object with the experimenter. During the task twelve objects were placed in front of the test subject, half of the objects heavy, the other half light. Only five of the twelve subjects managed six consecutive correct exchanges, requiring a median of 30 sessions or 331 exchanges to reach this criterion. During another version of the task where the objects differed in colour instead of weight six subjects reached the same criterion in a median of 64 exchanges (Schrauf & Call, 2009).

Little research has been conducted with non-primate species with weight-based discrimination tasks. However, New Caledonian crows (*Corvus moneduloides*) have shown that they are able to discriminate between non-identical heavy and light objects in an artificial setting. They quickly learnt which object type (light or heavy) of a collection of visually different objects was rewarded and were able to transfer this rule onto previously unseen objects which were also visually distinct (Jelbert *et al*., 2019). Both New Caledonian crows and kea (*Nestor notabilis*) were able to learn about the weight of coloured blocks when they were given the opportunity to freely explore a set of novel items (Lambert *et al*., 2017). Following the exploration of the novel items the test subjects were given tool-use tasks where they could choose among the objects to solve the tasks. If the subjects were given the opportunity to explore the objects before participating in the tasks some of them performed above chance-level compared to a control experiment where no prior exploration phase was given. Nevertheless, to be able to make more meaningful comparisons between primate and avian species it is important to also test if and how quickly birds can discriminate between visually identical objects based on weight alone.

This study tested Goffin's cockatoos (*Cacatua goffiana*), a species of parrot endemic to the Tanimbar archipelago of the Moluca region of Indonesia, in a weight-based choice task. As Goffin's cockatoos acquire a regular part of their diet through extractive foraging in the wild, they provide an ideal avian model for this question. For example, they extract the flesh and water of young coconuts and the seeds and flesh of ripe papayas, which first requires biting through the tough outer layer of the fruit (O'Hara *et al*., 2019). In the lab as well in the wild, members of the species have demonstrated skilful tool use and manufacture, even though they are not specialised tool users in the wild. The birds showed the ability to fabricate stick-tools out of three different materials, as well as crafting hook-tools out of bendable materials and most recently it was found that wild individuals use and make sets of up to three tools of different functions to obtain a single goal. (e.g. Auersperg *et al*., 2016; Laumer *et al*., 2017, O'Hara & Mioduszewska *et al*. 2021). To successfully use tools proprioceptive feedback is an important factor (Fragaszy & Mangalam, 2020).

In this study, the birds were rewarded for offering either the heavy or the light object of a pair, according to a rule they were trained by. Our first prediction was that subjects would be able to discriminate objects, which are visually identical, given only a difference in weight. This prediction is based on studies carried out with primates, which are extractive-foraging and tool-skilled like our chosen model species (Povinelli, 2011). Additionally, we were interested in learning how the Goffin's would perform, when subjects had the combined cues of weight and colour to differentiate between objects. One group was given differently coloured weighted objects during a training phase before being tested with the visually identical objects. We predicted that the group which received an additional the colour cue would learn faster to reliably discriminate the objects than the group only ever given visually identical objects.. Finally, we predicted if birds that received the colour training first would be better at the task with the identical objects than the birds which did not have this experience.

2. Material and Methods

2.1. Subjects

16 captive-born and hand-reared Goffin's cockatoos were tested. 13 of them were adults, aged between 10 and 13 years, and three were sub- adults, each two years of age. There were eight male and five female adults, as well as one male and two female sub-adults. All birds were housed together in an aviary which is situated at the Goffin

Lab, Goldegg. The indoors area had a floor space of 45m² and a height between three to six meters. The outdoors area had a floor space of 150m² and a height between three to five meters. Within the indoor as well as the outdoor aviary there were multiple perches. For enrichment the animals were always provided with parrot toys and cardboard. Water for drinking and bathing, as well as food was provided ad libitum. The food consisted of seed, various protein sources, minerals and a variety of fresh fruits. For each experiment the birds are called into the testing room and participate voluntarily.

2.3.1 Ethical note

All animals were bought with CITES certificates and were registered at the district's administrative animal welfare bureau (Bezirkshauptmannschaft St. Pölten, Austria). These housing conditions comply with the Austrian Federal Act on the Protection of Animals. The study was approved by the Ethics and Animal Welfare Committee of the University of Veterinary Medicine Vienna (ETK-145/09/2020). All birds were marked with coloured leg bands for identification.

2.3.2 Experimental History

All subjects have previously been involved in various physical problem-solving tasks. However, they have no previous experience with weight discrimination tasks. All adult birds were previously involved in several studies featuring a collapsible platform and a ball which was heavy enough to collapse the platform. Furthermore, the birds showed that they are able to manufacture shapes of a rewarded template, they possess the ability to fit objects with different shapes into the correct keyholes. They are also able to select the correct tool after being presented with an apparatus which was occluded before the tools were presented for selection (Laumer et al, 2016; 2021; Habl & Auersperg., 2017; Beinhauer et al, 2019).

2.3.3 Group Assignment

The birds were randomly assigned into one of two groups, the test-only group and the pre-experience group, with eight individuals each. To randomly assign subjects into groups the website random.org, specifically the list randomizer, was used. For the randomization, the birds were sorted into males and females and equal numbers of males and females assigned to the test groups and the control group, to exclude differences in performance dependent on age and sex. Afterwards their names were inscribed into the excel list in the order issued by the randomizer (Table 1).

The birds were assigned into two further categories in the test-only group and four further categories in the pre-experience training group (see Tab. 1). Within the two groups, half of the birds were rewarded for the heavy object and the other half was rewarded for the light object. The pre-experience group was first presented with objects which additionally differed in colour and the colour-weight combinations of these objects remained constant during all sessions. The test-only group was only presented with the visually identical objects during all trials. The colours of objects were counterbalanced with weight in the pre-experience group and age and sex was controlled for (as much as possible) with regards to the rewarded weight across both groups.

Table 1. Randomized list of the birds sorted into groups based on weight and colour of the rewarded object during the experiment. The pre-experience group was sorted into different colour-weight combinations. The test group was only sorted into weight groups. Fa = female adult (five individuals), Fs = female sub-adult (two individuals), Ma = male adult (eight individuals), Ms = male sub-adult (one individual)

pre-experience group				test-only group	
heavy object		light object		heavy object	light object
purple	blue	purple	blue		
Figaro (Ma)	Kiwi (Ma)	Titus (Ms)	Zozo (Ma)	Pipin (Ma)	Konrad (Ma)
Fini (Fa)	Irene (Fs)	Olympia (Fa)	Muki (Ma)	Dolittle (Ma)	Muppet (Ma)
				Moneypenny (Fa)	Heidi (Fa)
				Jane (Fs)	Mayday (Fa)

2.2. Materials and apparatus

The objects were made of Fimo soft®, a baked modelling clay, and shaped into small smooth balls of identical size. The heavy object contained a 20g spherical lead fishing weight inside, while the light object contained a cotton ball of the same shape and size of the fishing weight. After baking the objects all spheres were 2cm in diameter, but the heavy objects weighed 23g and the light ones 4g. The objects for the pre-experience group were purple and light blue. The identical objects for the testing phase were grey. Since the objects were crafted by hand, minor differences (not perceptible to a human subject) in shape were possible. Seven grey balls were randomly switched between each trial to make the such possibly tiny differences between each pair of objects inconsistent.

Two transparent plastic dishes were also used. These dishes were approximately 10cm in diameter. A third dish was used for the placement of the rewarded object. This dish was coloured black, 9.5cm in diameter and had a raised side of 3cm.

2.3. Procedure

2.3.4 Experimental Procedure

The pre-experience group was first given the discrimination task with objects that differed in colour as well as weight and after reaching criterion were given five sessions of the weight-only task with visually identical objects. The weight-only group started from the beginning with only the weight cues (Figure 1). The trials were otherwise the same for both groups.

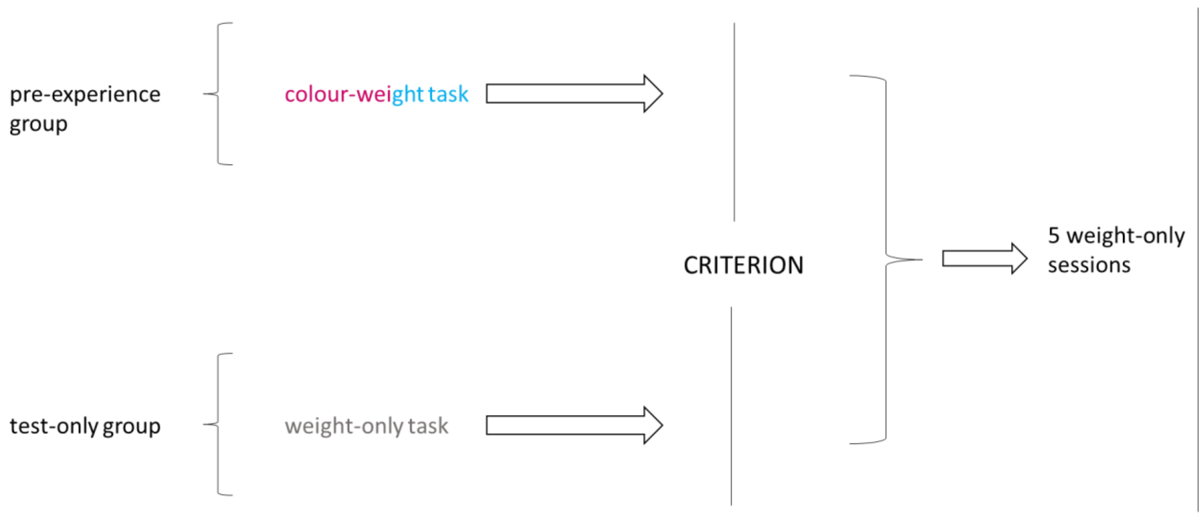

Figure 1. A representation of the testing plan of the two groups until they reach criterion and the post-criterion phase.

To reach criterion and enter the next phase the subject needed to put the rewarded object into the dish either on eight consecutive trials, or on nine out of ten trials. For the test-only group, there was a limit of ten sessions to reach criterion in the weight-only task. If the criterion was not met the individual was moved to the pre-experience group and given training with colour cues. Once the individual reached criterion in either group five extra sessions were given of the weight-only task to check whether their performance could be maintained.

To check if a subject needed to be habituated or trained in the command to be used ("give me"), all individuals were given one session prior to testing where they needed to move both spheres into the transparent dishes ten times. They were rewarded with a low-value reward in form of a sunflower seed. The subjects were otherwise naïve to the testing procedure at the start of the testing sessions. During all sessions the subject was visually isolated from the group. All trials were recorded by a video camera mounted on a tripod.

Before starting the trial, the bird was asked to sit on a wooden chair opposite of the experimenter. During the whole process the experimenter avoided making unnecessary movements and wore mirrored sunglasses to avoid cuing the birds. For the bird to stay on the chair, one hand was held in front of it and the voice command "stay" was given. The two transparent dishes were already placed in the middle of the table in a straight line approximately 20 cm apart. Next, the test objects were placed in front of the bird from the experimenters left to the right and approximately 10 cm apart.

After setting up the test, Step 1 of the trial consisted of the experimenter pointing to the object on the left and then to the plate on the left and using the voice command "give me". As soon as the bird placed the object onto the plate it received a sunflower seed and the vocal praise "fine". Then the experimenter pointed to the second object and the second plate and said "give me" to instruct the bird to transfer the second object. This was again rewarded with a sunflower seed and the word "fine".

In the second step, the black dish was placed centrally between the experimenter and the transparent dishes. For this second stage the experimenter pointed into the dish and said, "give me" to indicate that the bird should place one object into it (see figure 2 for a depiction of a trial).

If the desired object was placed into the black dish the bird was rewarded with 1/8th of a cashew nut and the verbal praise "fine". If the incorrect object was placed into the dish the bird received a verbal cue that the choice was wrong and was asked to return to the chair for additional 30 seconds. Thereafter the objects were removed from the dishes before a new trial started. No time limit was given for the conclusion of a trial. Each subject could take as long as it wished to complete the task. After finishing the trial, the subject was asked to return to the chair and the objects and the black dish were removed before the next trial started after a short delay of about 20 seconds.

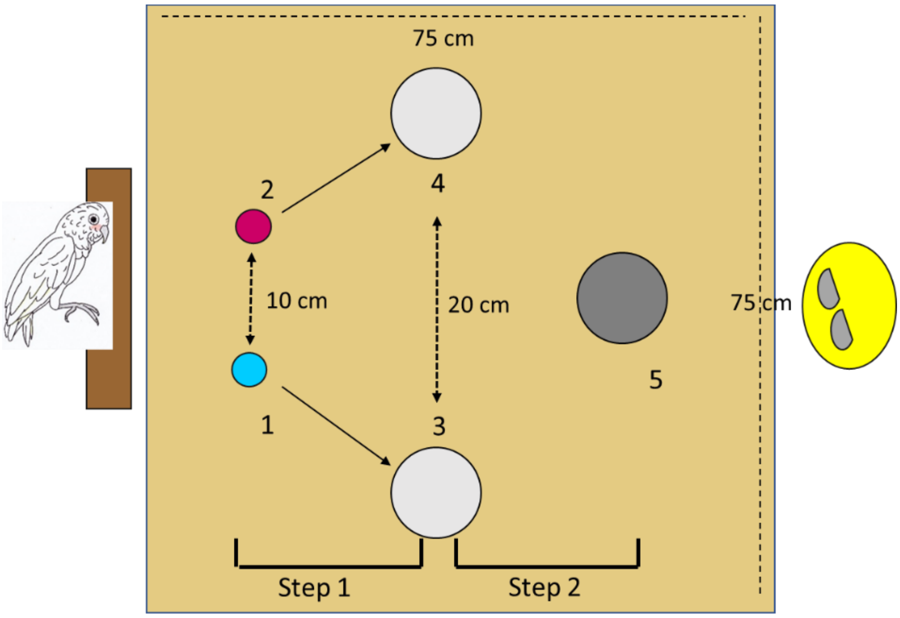

Figure 2. Placement of the test subject and the objects during the sessions. The numbers symbol the order of placement. 1 and 2 are the objects, which are placed by the experimenter, 3 – 5 show the order of placements by the subject. The arrows show the movement of the objects during the first step of each trial. Distances between the objects and plates are shown by the dotted lines.

A session was terminated only if a subject displayed signs of stress or indicated otherwise that it no longer wished to participate. The subject could indicate the unwillingness to participate by refusing to step off the chair or by distancing themselves from the testing table and not coming back when called. To confirm live coding choices of the subject trial videos were analysed. Those videos were also used to retrospectively record picking up and switching behaviour.

2.3.5 Object placement

Four different placement orders of the rewarded object were generated through the use of the website random.org. During each block the desired object was placed five times on each side with an accepted maximum of four times in a row. These blocks were rotated for each session of each subject in the order 1-4. (Table 2)

Table 2. Four semi-randomised test sessions containing ten trials each. The named side showed the rewarded objects placement.

rewarded object placement			
1	2	3	4
left	left	right	left
right	left	right	left
right	left	left	right
left	right	left	left
right	left	left	right
left	right	right	right
right	left	right	left
right	right	left	left
left	right	right	right
left	right	left	right

2.3.6 Deviations from procedure

One subject from the pre-experience group, Olympia, was wrongly given an extra session after reaching criterion during the colour-weight task. This session's data was not included in the statistical analysis of model 2 (see Analysis). Two subjects of the pre-experience group, Fini and Muki, were wrongly given an additional colour-weight session after having already reached criterion which interrupted the weight-only task after two sessions for Fini and one session for Muki. As it is possible that this additional session could have influenced subject learning during the weight-only task the data of these two subjects was left out of the analysis in models 1-3. However, all data of all subjects was included in the appendix and was used during the analysis of models 4 and 5.

2.3.7 Habituation and pre-training

Two adult birds, Olympia and Konrad, needed to be habituated to the balls and the dishes. To do that they were fed baby mash and sunflower seeds out of all dishes prior to starting the tests. The sub-adults needed to be trained in the "give me" command and habituated to the testing chamber. For the training of the exchange the protocol of Auersperg et al. 2013 was used.

2.3.8 Dropping and switching behaviour

Dropping behaviour was defined as picking up one object, dropping it down again and picking up the same object during step 2 without switching between the two objects. Switching behaviour was defined as picking up, putting down and switching to the other object. These behaviours could occur as many times as the bird wanted before placing one object into the black dish. Switching was possible in different directions, from the correct object to the incorrect one, from the incorrect to the correct one or starting and ending with no overall change. Dropping behaviour may have produced different sound cues based on the different weight of the dropped object.

2.4. Analysis

Six Generalized Linear Mixed Models (GLMMs) with a binomial error structure and logit link function (McCullagh & Nelder, 1989) were run to answer three main research questions and two post-hoc questions. Before fitting the models, all predictor covariates were z-transformed (Forstmeier & Schielteth, 2011). To avoid pseudo-replication, a random intercept of subject and the combined factor of session and subject, ('dayID') was included. This was done to account for day specific variation, for example motivation and mood of the test subject as well as all theoretically identifiable random slopes in all of our models. All models were fitted with the maximal random effects structure (Barr et al., 2013). If the majority of the correlations appeared unidentifiable or any resulted in non-numerical output, correlations of random slopes and its random intercept were removed to balance type I and II errors (Matuschek *et al.*, 2017).

Model 1 was used to analyse the main research question, whether the birds were able to learn to differentiate objects only by their weight. This model used the data from the five weight-only sessions, which were given after the birds reached criterion. To control for possible sound cues produced by different weights Model 1 was run again containing only the trials of the weight-only sessions after reaching criterion where no dropping of the object occurred. Model 2 tested if the subjects met criterion quicker when they had additional colour cues. Therefore, it used the data of the sessions before subjects met criterion. Model 3 investigated whether experience with the objects with an additional colour cue influenced success with visually identical objects. It used the data from all sessions of the weight-only task of both groups. In these three models our fixed effects for predicting success were session number, trial number, group and whether the rewarded object was heavy or light (weight kind). Group and weight kind were centred and dummy coded before inclusion in model 1 to attain a meaningful probability measure across both groups and weight types. To account for the possibility of a sigmoidal-shaped learning curve session number was included as 'session+session²'. A three-way interaction between 'session+session²', trial number and group was included to account for the possibility of different learning speeds between the two groups.

Models 4 and 5 were used to analyse post-hoc questions regarding the switching behaviour. Model 4 was used to assess if switching behaviour predicted the probability of success. Model 5 was used to analyse if the choice of the first object picked up (correct or incorrect) during Step 2 predicted if the subject would switch the objects. Group, task, session, trial and weight kind were included as fixed effects predicting success in model four and probability of switching in model five. In model four an interaction between task and switching behaviour was included, as the likelihood of switching theoretically could depend on the task. In model five an interaction between task and choice of first object was included, as the effect of the first choice on probability to switch objects could depend on the task.

To confirm that none of the model assumptions were violated after fitting each model the dispersion parameter was evaluated to rule out overdispersion, the 'Variance Inflation Factor' (Field, 2005) was calculated and to 'Best Linear Unbiased Predictors' were visually inspected to check if they were approximately normally distributed (Baayen, 2008; Harrison et al., 2018). Furthermore, model stability was evaluated by comparing estimates from the model based on all data with estimates acquired from models in which the levels of random effects were excluded one at a time (Nieuwenhuis *et al*., 2012). One comparison was done between the full model containing all terms and a null model lacking the term of interest. All models except model one were compared in this way, because for model one the intercept was the main interest as an intercept of 0 represented chance level with a probability of 0.5. The full-null model comparisons were done to avoid 'cryptic multiple testing' (Forstmeier & Schielzeth, 2011). In case of a significant result during the full-null comparison the complexity of the model structure was reduced by stepwise removal of non-significant interactions to obtain informative estimates of the fixed effects terms. This was done from higher to lower order terms.

All analyses were done using R-Studio (RStudio Team; version: 1.3.1093) of the open-source software R (R Core Team; version: 3.6.3). The GLMMs were fitted using the function ‘glmer’ (Bates et al., 2015; package ‘lme4’; version: 1.1.25). The optimizer ‘bobyqa’ was used and all models had iterations set to 100,000. The function ‘vif’ of the package ‘car’ (Fox & Weisberg, 2019; version: 3.0.10) was used to evaluate ‘Variance Inflation Factors’. Roger Mundry kindly provided the functions that were used to evaluate model stability and the calculation of confidence intervals. The function ‘bootMer’ of the package ‘lme4’ was used to derive confidence intervals. For both fixed effects and random effects 1,000 parametric bootstraps were used. For evaluating the likelihood ratio tests the functions ‘anova’ and ‘drop1’, with the argument ‘test’ set to ‘Chisq’, were used.

3. Results

Looking at the responses of subjects reveals that the test-only group required a mean of 60.6 trials, with a subject range of 30 to 88, to reach criterion. However, one subject of the test-only group did not manage to reach criterion within the limit of ten sessions and therefore the data was neither included here nor in the analysis for model 1 and the associated analysis disregarding the trials where dropping behaviour occurred. The seven birds which reached criterion in the weight-only task were able to maintain their performance in the five sessions given to them afterwards, with 90.9% correct choices. For these individuals the range of correct choices was between 76% - 98% and six out of seven birds achieved 90% correct or higher. The eight birds in the colour task required a mean of 40.8 trials, with a range of 20 to 70 trials, to reach criterion. In the five sessions after reaching criterion the data of two birds of the colour-weight task needed to be excluded because of an interruption of the session order. The six remaining birds achieved 77.6% correct choices, with a range of 60% - 92% between the individual subjects.

Model 1 confirmed that birds in both groups clearly learned to discriminate the objects based on weight: they chose correctly significantly above chance on the sessions given after criterion (intercept estimate of 1.713, SE ± 0.232, $z = 7.392$, $p < 0.001$) The estimated probability of choosing the correct object differed for each individual, however out of 16 subjects only three had a probability lower than 80% (see Appendix 1).

The analysis of the same weight-only sessions excluding 305 trials where dropping behaviour occurred showed similar results to the first model (intercept estimate of 1.472, SE ± 0.266, $z = 5.523$, $p < 0.001$). The probability to choose the correct object was lower for many birds, however seven subjects still managed to have over 80% correct choices and no subject had a probability lower than 68% (see Appendix 4).

Model 2 revealed a significant difference between the speeds at which both groups met criterion. The success rate for the pre-experience group in the colour-weight task was significantly higher than the success rate for the subjects in the test-only group during the weight-only task until criterion (Figure 3). The model analysing all sessions until the subject reached criterion showed a full-null model comparison of $\chi^2 = 19.364$, $df = 6$, $p = 0.004$. The group estimate was -0.624, SE ± 0.197, $z = -3.167$, $p = 0.002$ (see Appendix 7).

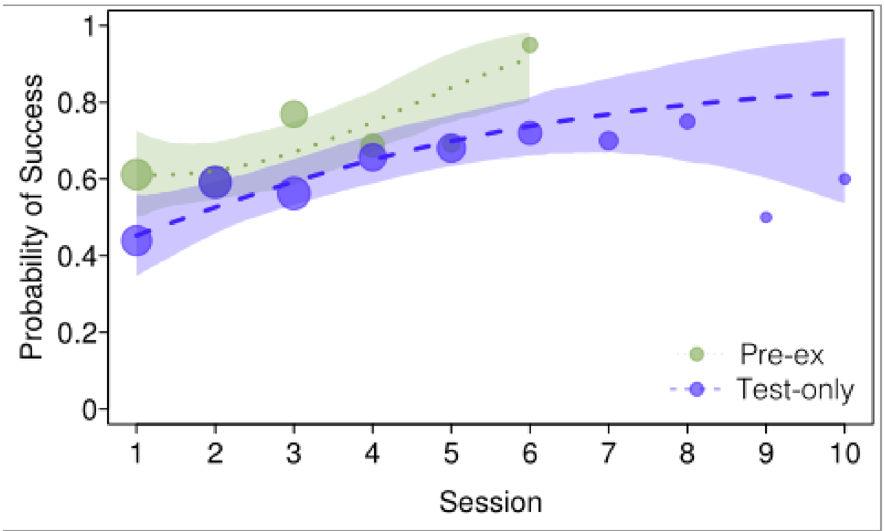

Figure 3. Influence of the group on probability of success, over all sessions for the pre-experience group in the colour-weight task and the test-only group in the weight-only task until each group reached criterion.

Comparing all weight-only trials of both groups showed that the subjects of the pre-experience group were more successful than the subjects of the test-only group (Figure 4). The full-null model comparison for Model 3 was significant: $\chi^2 = 18.768$, $df = 6$, $p = 0.005$. The fixed effects result for group was an estimate of -1.110, SE ± 0.272, $z = -4.086$, $p > 0.001$ (see Appendix 9).

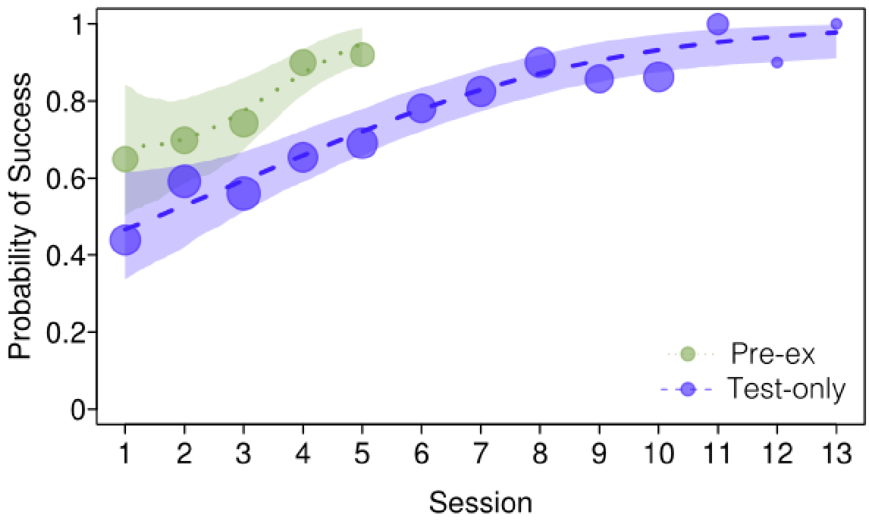

Figure 4: Influence of group on probability of success during all sessions of the weight-only task.

Switching behaviour occurred in 237 trials out of 1640 trials. In most cases the switching behaviour involved a single switch from one object to the other. A few times subjects switched back and forth between objects multiple times with a maximum of four switches during a single trial. During the colour-weight task switching behaviour occurred 25 times overall. During the weight-only task switching behaviour occurred 212 times (Table 3).

Table 3. Number of trials per combination of task, direction, and number of switches.

		Number of switches				
Task	Overall direction of switch	1	2	3	4	Total n
weight-only (n=212)	incorrect -> correct	181	0	3	0	184
	correct -> correct	0	17	0	2	19
	correct -> incorrect	7	0	1	0	8
	incorrect -> incorrect	0	1	0	0	1
colour-weight (n=25)	incorrect -> correct	19	0	0	0	19
	correct -> correct	0	0	0	0	0
	correct -> incorrect	5	0	0	0	5
	incorrect -> incorrect	0	1	0	0	1
	Total n	212	19	4	2	237

The statistical analysis of the influence of switching behaviour on success showed that, although this behaviour occurred in only 237 out of 1640 trials, it does indeed predict success. The full-null model comparison for Model 4 resulted in $\chi^2 = 18.669$, df = 2, p < 0.001. The fixed effects results showed a switch estimate of 1.806, SE ± 0.438, z = 4.119, p < 0.001 (see Appendix 11).

The first object picked up predicted switching behaviour of the subjects. The full-null comparison for Model 5 resulted in $\chi^2 = 40.177$, df = 2, p < 0.001. The fixed effects results showed a first object estimate of 9.270, SE ± 1.511, z = 6.136, p < 0.001 (see Appendix 13) – i.e. a subject was more likely to switch if they first picked up the incorrect object.

4. Discussion

The results confirmed our main prediction that Goffin's cockatoos can discriminate between visually identical objects based upon only weight. Comparing the number of trials it took for both the pre-experience group and the test-only group to reach criterion, the birds which received the colour cues in addition to the weight cue did in fact learn faster to discriminate between the objects.

The analysis of all weight-only trials for both groups showed that the group which was given pre-experience through the colour-weight task performed better than the test-only group when discriminating the visually identical objects. These results suggest that the birds paid attention to the weight as well as the colour cues during the colour-weight trials and were therefore able to use the weight as a differentiating factor during the weight-only trials more quickly than the birds without previous experience with the weighted objects. This is to some extent corroborated by previous findings that New Caledonian crows and kea could learn about the weight of objects when they differed systematically in colour (Lambert *et al*., 2017). The data in the experiment of Lambert *et al*. (2017) suggests that the birds learned about the properties of the tools during exploration. Therefore, it might be possible to suggest that New Caledonian crows, kea and Goffin's learn about an objects weight while handling it. This could explain the learning about object weight by the Goffin's during this learning discrimination task when they were presented with the differently coloured objects. These similarities between species seem to indicate that exploration is an important factor in learning about object properties.

Comparing the results of the Goffin's cockatoos with apes which have been tested on similar tasks, the birds were able to learn to differentiate between visually identical weighted objects much more quickly. The seven out of eight subjects which reached criterion in the weight-only task in our study needed an average of 60.6 trials to do so. In contrast, chimpanzees needed an average of 895.2 trials in a study by Povinelli (2011) and great apes needed a median of 331 'exchanges' with the experimenter to reach the criterion used in a study by Schrauf and Call (2009). Nevertheless, the set criterion differed between studies: for the chimpanzees in Povinelli's (2011) study the probability of reaching criterion by chance was 0.001 and for the great apes in Schrauf and Call (2009) the probability was <0.0000001. For the Goffin's cockatoos the chance level probability of making eight correct choices in a row was 0.0039, and to get nine out of ten trials correct within a session it was 0.0098. While the criterions differed for each study, which makes direct comparisons difficult, it could be said that this studies criterion is similar to the one used in Povinelli's study. Therefore, we can say that the Goffin's were faster in reaching criterion than the chimpanzees in these two different studies.

A further difference between these studies that needs to be considered is their methodology. These studies' methods focused on the subjects lifting the differently weighted objects and deciding which one to place into the dish. Schrauf and Call (2009) required subjects to exchange six out of twelve correct objects with the experimenter which differs from this study. Povinelli (2011) focused on sorting objects based on their weight into differently coloured plates which does differ as our test subjects only needed to choose one object to exchange and not sort both objects.

One possible explanation for the seemingly faster differentiation by Goffin's cockatoos compared to primate species might be a higher sensitivity to weight. Avian species may be required to pay more attention to 'weight-related signals' when transporting or using materials because of their evolutionary adaptations for flight (Povinelli, 2011) and the energetic costs of carrying weight during aerial movement.

Moreover, when comparing the species, the weight of the object relative to subject weight could be a contributing factor to the time it takes to master weight-related tasks. The near 20g difference between the objects in this study amounts to 6-8% of the birds body weight. In comparison, the 300g difference between objects of a study done with chimpanzees amounted to 0.00007% of the average species body weight (Schrauf &

Call, 2009). Therefore, it is possible that the birds were better at differentiating between the weight of the objects because the relative weight of the objects to the body weight of the subjects was larger and therefore more salient. Another important difference lies in the body strength of an animal. Even closely related species differ greatly in body strength based on their size, for example humans and chimpanzees. It is possible that weight perception could be affected by this difference in body strength and therefore the same weight could be experienced differently between the species. The lighter weight in comparison to body weight for the apes could be perceived differently than how the weight difference is perceived by the birds.

We found that switching behaviour was functional during the test setting and predicts success. Our exploration of the data indicated that a possible factor for success was the correction of incorrect choices through switching behaviour. The majority of switches occurred when the first object picked up during Step 2 of the trial was incorrect. If an object was dropped during the trial, the sound that was produced differed noticeably between the heavy and the light objects. Therefore, the possibility exists that the birds learned which object was rewarded based on the sound the objects made when they were dropped. However, when trials where dropping occurred were left out of the statistical analysis the performance of the birds remained significantly above chance level. When looking at the level of the individual, the two subjects with the highest dropping rate increased the frequency of the behaviour during their later sessions where they were already performing with a high rate of success. A possible explanation for this behaviour is that these birds made an association between dropping the objects and having success. Altogether, it is unlikely that an association of the sound created by the dropped objects influenced the outcome of the trials, even though sound cues may be an important input for birds if they need to discriminate between small food items (Jablonski *et al*., 2015). Nonetheless, further studies with investigating weight discrimination should minimize auditory cues.

It is challenging to make comparisons between species solely based on a criterion that was set by the experimenter and which can be reached by chance. It is important to question if subjects reaching criterion is sufficient information to draw conclusions from about whether a species does or does not possess a particular cognitive ability. For the statistical analysis of the data GLMMs were used to avoid multiple testing issues and to avoid committing a type 1 error (i.e. concluding that the Goffin's cockatoo has the ability to discriminate based on weight alone when in fact they don't).

Few studies have been done involving weight discrimination tasks, and nearly all of these tested primate species (e.g. Hanus & Call, 2008; Schrauf & Call, 2011; Povinelli, 2011; Neilands *et al*., 2016;). Even though it may be difficult to make direct comparisons between different species, because relative strength and morphology could affect how weight differences are perceived, studies with different species could provide important insight in the cognitive abilities of different animals. To enable comparison between species test set-ups should be as similar to each other as possible. Furthermore, statistical analysis should be appropriate with regards to multiple testing issues. Since this study suggests quicker learning with regard to weight information in birds compared to primates, it is important to consider the ecological factors present during the evolution of the species and how these may have influenced cognitive abilities.

5. References

Auersperg A.M.I., Laumer I., Bugnyar T., 2013: Goffin cockatoos wait for qualitative and quantitative gains but prefer 'better' to 'more'. Biology Letters, 9(3), 20121092. DOI: 10.1098/rsbl.2012.1092.

Auersperg A.M.I., Borasinski S., Laumer I., & Kacelnik A., 2016: Goffin’s cockatoos make the same tool type from different materials. Biology Letters, 12(11), 20160689.

Baayen R., 2008: A practical introduction to statistics using R. Analyzing Linguistic Data. Cambridge University Press.

Barr D.J., Levy R., Scheepers C., & Tily H.J., 2013: Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68(3), 255–278.

Bates D., Mächler M., Bolker B., & Walker S., 2015: Fitting linear mixed-effects models using lme4. 67(1), 1–48.

Beinhauer I., Bugnyar T., & Auersperg A.M.I., 2019: Prospective but not retrospective tool selection in the Goffin’s cockatoo (Cacatua goffiniana). Behaviour, 156 (5-8), 633-659. Doi: https://doi.org/10.1163/1568539X-00003515

Field A., 2005: Discovering statistics using SPSS. Thousand Oaks, CA, US.

Forstmeier W., & Schielzeth H., 2011: Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner’s curse. Behavioral Ecology and Sociobiology, 65(1), 47–55.

Fox J., & Weisberg S., 2019: An R Companion to Applied Regression (Third Edition). R package version 3.0.

Fragaszy D.M., & Mangalam M., 2020: Folk Physics in the Twenty-first Century: Understanding Tooling as Embodied. Animal Behavior and Cognition, 7(3), 457–473.

Habl C., & Auersperg A.M.I., 2017: The keybox: Shape-frame fitting during tool use in Goffin's cockatoos (Cacatua goffiniana). PLOS ONE, 12(11), e0186859 doi: https://doi.org/10.1371/journal.pone.0186859

Hanus D., & Call J., 2008: Chimpanzees infer the location of a reward on the basis of the effect of its weight. Current Biology, 18(9), R370–R372. doi:https://doi.org/10.1016/j.cub.2008.02.039

Harrison X.A., Donaldson L., Correa-Cano M.E., Evans J., Fisher D.N., Goodwin C.E., Robinson B.S., Hodgson D.J., & Inger R., 2018: A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ, 6, e4794.

Heinrich B., Joerg C.C., Madden S.S., Sanders Jr. E.W., 1997: Black-Capped Chickadees and Red-Breasted Nuthatches 'Weigh' Sunflower Seeds. The Auk, 114(2), 298–299.

Jablonski P.G., Lee S., Fuszara E., Fuszara M., Jeong C., & Lee W.Y., 2015: Proximate mechanisms of detecting nut properties in a wild population of Mexican Jays (Aphelocoma ultramarina). Journal of Ornithology, 156(S1), 163–172.

Jelbert S.A., Miller R., Schiestl M., Boeckle M., Cheke L.G., Gray R.D., Taylor A.H., & Clayton N.S., 2019: New Caledonian crows infer the weight of objects from observing their movements in a breeze. Proceedings of the Royal Society B: Biological Sciences, 286(1894), 20182332. doi:10.1098/rspb.2018.2332

Klüver H., 1933: Behaviour mechanisms in monkeys. Chicago: Chicago University Press.

Lambert M.L., Schiestl M., Schwing R., Taylor A.H., Gajdon G.K., Slocombe K.E., & Seed A.M., 2017: Function and flexibility of object exploration in kea and New Caledonian crows. Royal Society Open Science, 4(9), 170652. doi:10.1098/rsos.170652

Laumer I.B., Bugnyar T., & Auersperg A.M.I., 2016: Flexible decision-making relative to reward quality and tool functionality in Goffin cockatoos (Cacatua goffiniana). Scientific Reports, 6(1), 28380. doi: https://doi.org/10.1038/srep28380

Laumer I.B., Bugnyar T., Reber S.A., & Auersperg A.M.I., 2017: Can hook-bending be let off the hook? Bending/unbending of pliant tools by cockatoos. Proceedings of the Royal Society B: Biological Sciences, 284(1862), 20171026.

Laumer I.B., Jelbert S.A., Taylor A.H. *et al.*, 2021: Object manufacture based on memorized template: Goffin's cockatoos attend to different model features. Animal Cognition 24, 457-470 doi: https://doi.org/10.1007/s10071-020-01435-7

Matuschek H., Kliegl R., Vasishth S., Baayen H., & Bates D., 2017: Balancing Type I error and power in linear mixed models. Journal of Memory and Language, 94, 305–315.

McCullagh P., & Nelder J.A., 1989: Generalized linear models. In Monographs on statistics and applied probability (2nd ed). Chapman and Hall.

McCulloch T.L., 1941: Discrimination of lifted weights by chimpanzees. Journal of Comparative Psychology, 32(3), 507–519.

Neilands P.D., Jelbert S.A., Breen A.J., Schiestl M., & Taylor A.H., 2016: How Insightful Is 'Insight'? New Caledonian Crows Do Not Attend to Object Weight during Spontaneous Stone Dropping. PLOS ONE, 11(12), e0167419.
doi:10.1371/journal.pone.0167419

Nieuwenhuis R., te Grotenhuis M., & Pelzer B., 2012: influence.me: Tools for detecting influential data in mixed effect models. 4, 38–47.

O'Hara M., Mioduszewska B., Haryoko T., Prawiradilaga D.M., Huber L., & Auersperg A., 2019: Extraction without tooling around—The first comprehensive description of the foraging- and socio-ecology of wild Goffin's cockatoos (Cacatua goffiniana). Behaviour, 156(5–8), 661–690.

O'Hara M., Mioduszewska B., Mundry R., Yohanna, Haryoko T., Rachmatika R., Prawiradilaga D.M., Huber L., Auersperg A.M.I., 2021: Wild Goffin's cockatoos flexibly manufacture and use tool sets. Current Biology.

Povinelli D., 2011: World without weight: Perspectives on an alien mind. Oxford university press, Oxford New York. doi:10.1093/acprof:oso/9780198570967.001.0001

R Core Team, 2020: R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Robinson H.B., 1964: An experimental examination of the size weight illusion in young children. Child Dev. 35, 91–107. DOI: 10.1111/j.1467-8624.1964.tb05921.x

RStudio Team, 2019: RStudio: Integrated Development for R (1.2.5033) [Computer software]. RStudio, Inc. http://www.rstudio.com/

Schrauf C., Huber L., & Visalberghi E., 2008: Do capuchin monkeys use weight to select hammer tools? Animal Cognition 11(3), 413-422. doi:10.1007/s10071-007-0131-2

Schrauf C., & Call J., 2009: Great apes' performance in discriminating weight and achromatic color. Animal Cognition, 12(4), 567–574. doi:10.1007/s10071-009-0216-1

Schrauf C., & Call J., 2011: Great apes use weight as a cue to find hidden food. American Journal of Primatology, 73(4), 323–334.

Schrauf C., Call J., Fuwa K., & Hirata S., 2012: Do Chimpanzees Use Weight to Select Hammer Tools? PLOS ONE, 7(7), e41044. doi:10.1371/journal.pone.0041044

Visalberghi E., Addessi E., Truppa V., Spagnoletti N., Ottoni E., Izar P., Fragaszy D., 2009: Selection of effective stone tools by wild bearded capuchin monkeys. Current Biology 19:213-217. doi:http://dx.doi.org/10.1016/j.cub.2008.11.064

Visalberghi E., & Neel C., 2003: Tufted Capuchins (Cebus apella) Use Weight and Sound to Choose Between Full and Empty Nuts. Ecological Psychology, 15(3), 215–228.

6. Appendix

Appendix 1. Estimated probability of correct choices in the five sessions after reaching criterion for each individual in percent.

Particpant	Correct
Dolittle	85.76%
Figaro	80.16%
Heidi	82.02%
Irene	76.10%
Kiwi	89.41%
Konrad	85.22%
Mayday	69.84%
Moneypenny	88.79%
Muppet	83.16%
Olympia	88.38%
Pipin	85.96%
Titus	87.66%
Zozo	77.81%

Appendix 2. Fixed effect results for model 1, with the data (a) z-transformed to a mean of 0 and a standard deviation of 1 and (b) dummy coded and centred.

term	estimate	SE	z-value	lower CI	upper CI	χ2	df	p	min	max
Intercept*	1.713	0.232	7.392	1.369	2.334				1.593	1.831
session[a]	0.661	0.256	2.578	0.244	1.458	5.990	1	0.014	0.441	0.847
session[a]^2	0.521	0.245	2.127	0.134	1.219	6.369	1	0.012	0.423	0.691
trial[a]	-0.097	0.134	-0.727	-0.426	0.197	0.528	1	0.467	-0.150	-0.014
group[b]	0.786	0.536	1.466	-0.084	1.957	2.014	1	0.156	0.556	1.213
weight.kind[b]	0.037	0.483	0.076	-0.954	0.981	0.007	1	0.934	-0.327	0.232

Appendix 3. Random effect results for model 1. (1) 'intercept' denotes random intercepts, other entries random slopes. Data (a) was z-transformed to a mean of 0 and a standard deviation of 1.

term	effect [(1)]	sd
dayID	trial[a]	0.000
dayID	intercept	0.000
subject	intercept	0.450
subject	session[a]	0.441
subject	session[a]^2	0.353
subject	trial[a]	0.055
subject	session[a]:trial[a]	0.263
subject	session[a]^2:trial[a]	0.154

Appendix 4. Estimated probability of correct choices in the five sessions after reaching criterion for each individual in percent excluding the trials where dropping behaviour occurred.

Particpant	Prob. Correct
Dolittle	86.04%
Figaro	73.76%
Heidi	75.97%
Irene	68.86%
Kiwi	88.32%
Konrad	81.26%
Mayday	69.21%
Moneypenny	82.92%
Muppet	78.73%
Olympia	86.54%
Pipin	83.64%
Titus	86.13%
Zozo	72.71%

Appendix 5. Fixed effect results for model 1, excluding the trials where dropping behaviour occurred. Data (a) was z-transformed to a mean of 0 and a standard deviation of 1 and (b) was dummy coded and centred.

term	estimate	SE	z-value	lower CI	upper CI	χ2	df	p	min	max
Intercept*	1.472	0.266	5.523	1.084	2.178				1.329	1.616
session[a]	0.544	0.284	1.916	0.017	1.510				0.426	0.687
trial[a]	-0.191	0.213	-0.896	-0.691	0.244				-0.310	-0.069
group[b]	0.865	0.515	1.680	-0.124	2.012				0.571	1.072
session[a]^2	0.596	0.288	2.070	0.201	1.586				0.474	0.817
weight.kind[b]	-0.161	0.551	-0.292	-1.274	0.910	0.087	1	0.769	-0.537	0.300
session[a]:z.trial[a]	-0.244	0.265	-0.923	-0.946	0.344	0.819	1	0.365	-0.383	-0.087
session[a]:group[b]	-0.983	0.489	-2.009	-2.256	-0.070	4.711	1	0.030	-1.423	-0.583
trial[a]:group[b]	-0.465	0.367	-1.267	-1.347	0.261	1.511	1	0.219	-0.681	-0.268
trial[a]:session^2[a]	-0.011	0.297	-0.038	-0.691	0.616	0.001	1	0.970	-0.197	0.189
group[b]:session^2[a]	-0.160	0.469	-0.341	-1.215	1.065	0.114	1	0.735	-0.608	0.114

Appendix 6. Random effect results for model 1, excluding the trials where dropping behaviour occurred. (1) ‘intercept’ denotes random intercepts, other entries random slopes. Data (a) was z-transformed to a mean of 0 and a standard deviation of 1.

term	effect [(1)]	sd
dayID	trial[a]	0.000
dayID	intercept	0.000
subject	intercept	0.492
subject	session[a]	0.426
subject	session[a]^2	0.342
subject	trial[a]	0.304
subject	session[a]:trial[a]	0.023
subject	session[a]^2:trial[a]	0.338

Appendix 7. Fixed effect results for model 2. Data (a) was z-transformed to a mean of 0 and a standard deviation of 1, (b) was dummy coded with “pre-ex” being the reference category and (c) was dummy coded with “heavy” being the reference category.

term	estimate	SE	z	lower CI	upper CI	x2	df	p	min	max
Intercept	0.876	0.190	4.611	0.534	1.309				0.666	1.028
session[a]	0.621	0.145	4.296	0.362	0.953	16.326	1	<0.001	0.464	0.703
trial[a]	-0.055	0.085	-0.645	-0.220	0.126	0.411	1	0.521	-0.101	-0.028
group[b]	-0.624	0.197	-3.167	-1.047	-0.249	9.873	1	0.002	-0.775	-0.487
weight.kind[c]	0.576	0.199	2.888	0.194	0.984	7.361	1	0.007	0.490	0.624

Appendix 8. Random effect results for model 2. (1) 'intercept' denotes random intercepts, other entries random slopes. Data (a) was z-transformed to a mean of 0 and a standard deviation of 1.

term	effect [(1)]	sd
dayID	trial[a]	0.000
dayID	intercept	0.000
subject	intercept	0.462
subject	session[a]	0.253
subject	session[a]^2	0.330
subject	trial[a]	0.101
subject	session[a]:trial[a]	0.119
subject	session[a]^2:trial[a]	0.159

Appendix 9. Fixed effect results for model 3. Data (a) was z-transformed to a mean of 0 and a standard deviation of 1, (b) was dummy coded with “pre-ex” being the reference category and (c) was dummy coded with “heavy” being the reference category.

term	estimate	SE	z	lower CI	upper CI	χ2	df	p	min	max
Intercept	1.815	0.264	6.865	1.360	2.382				1.516	2.128
session[a]	1.007	0.173	5.825	0.684	1.374	18.107	1	<0.001	0.932	1.115
trial[a]	-0.141	0.073	-1.950	-0.286	0.018	3.322	1	0.068	-0.182	-0.105
group[b]	-1.110	0.272	-4.086	-1.673	-0.595	13.141	1	<0.001	-1.386	-0.868
weight.kind[c]	0.623	0.225	2.766	0.199	1.099	5.717	1	0.017	0.514	0.821

Appendix 10. Random effect results for model 3. (1) 'intercept' denotes random intercepts, other entries random slopes. Data (a) was z-transformed to a mean of 0 and a standard deviation of 1.

term	effect [(1)]	sd
dayID	trial[a]	0.000
dayID	intercept	0.000
subject	session^2[a]:trial[a]	0.000
subject	session[a]:trial[a]	0.000
subject	trial[a]	0.000
subject	session^2[a]	0.134
subject	session[a]	0.450
subject	intercept	0.232

Appendix 11. Fixed effect results for model 4. Data (a) was dummy coded with “colour-weight” being the reference category, (b) was dummy coded with “pre-ex” being the reference category, (c) was z-transformed to a mean of 0 and a standard deviation of 1, (d) was dummy coded with “heavy” being the reference category and (1) switch is an integer with 1=yes and 0=no.

term	estimate	SE	z	lower CI	upper CI	x2	df	p	min	max
Intercept	1.538	0.236	6.523	1.098	2.053				1.450	1.666
switch[1]	1.806	0.438	4.119	1.219	3.328	17.272	1	<0.001	1.631	2.134
task[a]	-1.340	0.285	-4.708	-1.935	-0.836	18.015	1	<0.001	-1.451	-1.234
group[b]	0.638	0.236	2.708	0.203	1.121	6.022	1	0.014	0.495	0.713
session[c]	0.886	0.136	6.496	0.639	1.170	24.106	1	<0.001	0.845	0.956
trial[c]	-0.118	0.065	-1.802	-0.244	0.004	2.841	1	0.092	-0.145	-0.096
weight.kind[s]	0.422	0.154	2.732	0.104	0.792	6.505	1	0.011	0.369	0.523

Appendix 12. Random effect results for model 4. (1) ‘intercept’ denotes random intercepts, other entries random slopes. Data (a) was z-transformed to a mean of 0 and a standard deviation of 1 and (b) was dummy coded and centred.

term	effect [(1)]	sd
dayID	trial[a]	0.062
dayID	intercept	0.000
subject	intercept	0.383
subject	session[a]	0.378
subject	trial[a]	0.083
subject	task[b]	0.389
subject	switch[b]	0.705
subject	task[b]:switch[b]	1.493

Appendix 13. Fixed effect results for model 5. Data (a) was dummy coded with “correct” being the reference category, (b) was dummy coded with “colour-weight” being the reference category, (c) was z-transformed to a mean of 0 and a standard deviation of 1 and (d) was dummy coded with “heavy” being the reference category.

term	estimate	SE	z	lower CI	upper CI	χ2	df	p	min	max
Intercept	-9.594	1.570	-6.110	-12.897	-8.737				-11.040	-8.735
first.object[a]	9.270	1.511	6.136	8.453	12.251	39.836	1	<0.001	8.639	10.120
task[b]	-2.015	0.767	-2.626	-3.397	-0.493	6.763	1	0.009	-2.273	-1.489
session[c]	2.204	0.325	6.781	1.711	2.812	29.884	1	<0.001	1.940	2.411
trial[c]	0.186	0.163	1.136	-0.088	0.508	1.434	1	0.231	0.097	0.250
weight.kind[d]	0.921	0.622	1.480	-0.321	2.159	2.240	1	0.135	0.431	1.247

Appendix 14. Random effect results for model 5. (1) ‘intercept’ denotes random intercepts, other entries random slopes. Data (a) was z-transformed to a mean of 0 and a standard deviation of 1 and (b) was dummy coded and centred.

term	effect [(1)]	sd
dayID	intercept	4.414
dayID	trial[a]	0.619
dayID	first.object[b]	6.912
subject	intercept	1.441
subject	session[a]	0.397
subject	trial[a]	0.213
subject	task[b]	1.434
subject	first.object[b]	1.634
subject	task[b]:first.object[b]	2.149