DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Serial Reversal Learning in Goffin's Cockatoos (*Cacatua goffiniana*)"

verfasst von / submitted by

Cornelia Habl, BSc MSc

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

Magistra der Naturwissenschaften (Mag.rer.nat.)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 190 445 344
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Lehramtsstudium Unterrichtsfach Biologie und Umweltkunde, Unterrichtsfach Englisch
Betreut von / Supervisor:	Priv.-Doz. Dr.rer.nat. Alice Isabel Marie Auersperg

Index

4. Material and Methods

    4.6. Analysis

5. Results

1. Abstract

For several decades, researchers have been testing animals’ flexibility, as they need this ability to deal with changing environments. Serial reversal learning tasks were frequently used, as they provide an ideal set-up for testing behavioural flexibility comparatively. In these tasks, test subjects have to learn a rewarded stimulus which, upon reaching a learning criterion, changes contingencies (S+ becomes S-), and subjects have to adapt and relearn the rewarded stimulus. Only in recent years, after discovering that the mammalian forebrain and the avian pallium are homologous, avian species (especially corvids and parrots) have gained more traction as test species. I conducted a serial reversal task with Goffin’s cockatoos (*Cacatua goffiniana*), an Indonesian parrot species which has proven to be a prolific tool-user and manufacturer in the lab and an opportunistic and extractive forager in the wild. They were tested in 15 reversals with orange and blue containers changing contingencies, and were able to improve their performance significantly with each reversal. They performed similar to ravens and other parrot species when acquisition phase and first reversal were compared (1:1.8 ratio), but outperformed kea, another island-dwelling parrot species. Interestingly, neither sex nor the colour of training or testing containers had an effect on their performance. The individual data also reveals that the intra-specific variation is quite pronounced, as is true for other avian species. To gain more data for species-wide comparisons, this experiment is part of a large-scale study collecting data from different avian species (corvids, parrots, pigeons, Australian mynahs).

2. Zusammenfassung

Über die letzten Jahrzehnte haben Wissenschaftler versucht die Flexibilität von Tieren zu testen, da diese oft mit veränderlichen Umweltbedingungen konfrontiert sind. Dafür wurden oft Umkehrversuche (‘Serial reversal learning tasks‘) verwendet, da der Versuchsaufbau ideal ist um flexibles Verhalten vergleichend zu testen. Die Testsubjekte müssen dabei einen belohnten Stimulus lernen, der nach Erreichen eines Lernkriterions zu dem unbelohnten Stimulus wird, und sich anpassen und den neuen belohnten Stimulus lernen. Erst in den letzten Jahren, nach der Entdeckung, dass das der Säugetier Cortex und das Vogel Pallium homolog sind, wurden Vogelarten öfter in der Kognitionsforschung verwendet, besonders Corviden und Papageien. In dieser Studie habe ich Goffin Kakadus (*Cacatua goffiniana*) mit einem seriellen Umkehrversuch getestet. Diese indonesischen Kakadus haben in Gefangenschaft eine/n ausgeprägte Werkzeugherstellung und -gebrauch gezeigt, in der Wildnis sind sie hingegen nur opportunistische und extrahierende Futtersucher. Die Kakadus haben ihre Performance über 15 Umkehrversuche mit orangen und blauen Containern signifikant verbessert. Wenn man ihre Performance in der Aneignungsphase und der ersten Umkehrung vergleicht, schneiden sie im Vergleich mit anderen Vogelarten ähnlich gut ab wie Raben und andere Papageien (Verhältnis 1:1.8). Sie schnitten jedoch besser ab als eine andere inselbewohnende Papageienart, der neuseeländische Kea. Interessanterweise hatten weder das Geschlecht noch die Farbe der Trainings oder Test Container einen Einfluss auf die Performance der Kakadus. Die individuellen Daten zeigen außerdem, dass die intra-spezifische Variation, wie bei anderen Vogelarten, sehr hoch ist. Um daher mehr Daten für einen artenübergreifenden Vergleich zu erlangen, ist dieses Experiment teil einer großangelegten Studie, die Daten von verschiedenen Vogelarten sammelt (Rabenvögel, Papageien, Tauben, Mynahs).

3. Introduction

Many non-human animals do not solely rely on instinct, automated processes or genetically determined behaviours, but are able to adapt to demanding situations by applying combinations from a range of cognitive modules, which help them deal with variability in both social and environmental contexts (Sol 2009; Bond et al. 2007). These cognitive processes are called executive functions and typically involve inhibition, working memory, attentional control and cognitive flexibility (Diamond 2013). These functions in turn then serve as scaffolding for higher executive functions such as reasoning, planning and problem-solving, all of which are areas of interest in cognition research.

For decades, animal cognition research was mainly focussed on laboratory rats, pigeons and primates, as it was heavily influenced by behaviourism (Shettleworth 2009). Until the Cognitive revolution, little focus was set on avian species, as it was deemed unlikely that such phylogenetically distant taxa (last common ancestor with mammals 324 million years ago, dos Reis 2015) and with brains mainly composed of Striatum (now disproved, Güntürkün & Bugnyar 2016), could possibly be able to perform complex cognitive processes. However, over the years, several studies have shown that not only animals with a cortex possess the necessary neural foundation for these higher cognitive processes, but that neural structures like the avian and mammalian forebrain (Pallium and Cortex) are homologous and functionally similar (Güntürkün & Bugnyar 2016). While relative brain size used to be the best predictor for complex cognitive abilities, a higher number of neurons predict faster learning speed and greater flexibility in dealing with novel situations more accurately (Güntürkün et al. 2017). This is especially relevant for the avian taxa, as passerines (e.g. corvids) and parrots possess double packaging density of neurons in the pallium compared to primates. Moreover, corvids and parrots also have large brains in relation to their bodies, which furthermore increases their number of neurons and might explain their ability to perform cognitively complex behaviours (Olkowicz et al. 2016). Therefore, corvids and parrots are promising candidates for researching the aforementioned executive functions.

3.1. Serial Reversal Learning

Researchers have attempted to study and quantify the scope of cognitive flexibility for many decades. Therefore, Reversal Learning tasks are a well-established part of scientific behaviour research. Incidentally, they provide an ideal set-up for studying and comparing different species' ability to deal with changing contingencies. In a classic set-up, subjects first learn to choose between a rewarded (S+) and an unrewarded stimulus (S-), which they are confronted with until they reach a set learning criterion. Upon this, the contingencies change (S+ becomes the S- and vice versa) and subjects need to adjust their choices accordingly, until the criterion is reached anew and contingencies change again (Lambert et al. 2018). Usually, individuals reach criterion faster with each successive reversal, suggesting that not just associative learning but, additionally, rule learning is used (Bebus et al. 2016).

These choice stimuli can either be presented in the form of 3D objects (like levers, cups, strings etc.), 2D visuals like pecking keys or areas on a touchscreen, or by using other modes of perception like auditory or olfactory stimuli (O'Hara et al. 2015a).

Importantly, while all of the following studies have value standing on their own, comparing them conclusively is difficult or even impossible due to their sometimes widely different methodologies and measures of success.

Some of the earliest studies were performed by Gossette and Feldman, who tested different taxa (1966) and later compared their results with different species of new world monkeys (1968; 1969). They noted that yellow-headed amazons (*Amazona oratrix*) and red-billed blue magpies (*Urocissa erythroryncha*) performed best among ten avian species, and showed similarly low error levels to Capuchin monkeys (*Cebus sp*.). However, other (less encephalised) avian species like the Bobwhite quail (*Colinus virginianus*), Guinea fowl (family: Numididae) and trumpeters (*Psophia sp*.) performed considerably worse than the monkeys, only pigeons showed error levels closer to parrots.

Some studies have specifically focussed on corvids, e.g. Bond et al. 2007 compared pinyon jays (*Gymnorhinus cyanocephalus*), Clark's nutcrackers (*Nucifraga columbiana*) and California scrub jays (*Aphelocoma californica*) in their ability to discriminate colour and location of rewarded stimuli using the reversal learning paradigm. Despite their different ecologies, all species improved their performance similarly across serial reversals, as did New Caledonian crows (*Corvus moneduloides*) and carrion crows (*Corvus corone*) in another study by Teschke et al. 2013. According to the authors, one reason why all of these species showed similar performances despite their different (socio-) ecologies could be their use of generalised learning strategies or methodological issues.

Another method of measuring success in a reversal learning study, is to compare the ratios between the number of trials at certain points of the experiment. Studies with red-shouldered macaws (*Diopsittaca nobilis*), black-headed caiques (*Pionites melanocephalus*) and ravens (*Corvus corax*) all revealed an approximately 1:2 ratio between the number of trials to initial acquisition and the number of trials after the first reversal (Range et al. 2006; 2008; van Horik & Emery 2018). Interestingly, Kea (*Nestor notabilis*) performed at a 1:3 ratio when confronted with serial reversals featuring either 3D objects or touchscreen images (O'Hara et al. 2015a). Also, they acquired the initial distinction and the first reversal significantly faster in the solid object condition than in the touchscreen condition, highlighting again the difficulty and lack of validity of comparing studies using different methodologies.

Jelbert et al. (2016) tested New Caledonian crows (*Corvus moneduloides*) in an A-not-B task using the hand-tracking paradigm. Like the Kea, the crows also showed an approximately 1:3 ratio in a colour reversal learning task.

The mid-session reversal is another variety of the reversal learning task, offering individuals the possibility to learn to predict the change of contingencies, as they usually change mid-session, after the same number of trials. The strategy humans use in this sort of set-up is called 'win-stay/ lose-shift', while pigeons have been shown to use 'reversal estimation' as a strategy. Cook and Rosen (2010) tested humans and pigeons comparatively and noticed that humans tended to first stay with the previously rewarded stimulus until they chose incorrectly once or twice, and then shifted to the newly rewarded stimulus. Pigeons on the other hand made either anticipatory or perseverative errors, so they shifted from the reinforced stimulus either too soon or too late. In a recent study with kea, they seemed to use both strategies and some novel response patterns as well (Laschober 2019).

3.2. Goffins cockatoo

Goffins cockatoos (*Cacatua goffiniana*) are an opportunistic species of parrot endemic to the Tanimbar Islands (Indonesia), with feral populations in Singapore, Indonesia and Puerto Rico (Osuna-Mascaro & Auersperg 2018). In captivity, they have previously shown astonishing abilities in tool manufacture and use (e.g. Auersperg et al. 2012, 2014, 2016), even though they are not species-wide tool users but rather extractive foragers in the wild (O'Hara et al. 2019). These parrots are an interesting candidate for a reversal learning study, as they are known for their neophilia, behavioural flexibility and inhibition control (e.g. Auersperg et al. 2011, 2013a, 2015; Laumer et al. 2016), but have also shown great persistence when solving challenging and time-consuming tasks (Auersperg et al. 2013b). This persistence, typical for extractive foragers, might, in turn, pose a difficulty in succeeding in a reversal learning set-up.

3.3. Hypotheses

In this study we want to use a serial reversal learning paradigm to investigate cognitive flexibility in Goffins cockatoos. First of all, we want to measure their learning speed and determine, whether factors like container colour (in either the training or the testing) or sex have an effect on the number of trials before reaching criterion. Goffins have shown a preference for the colour yellow in a play experiment (Auersperg et al. 2015), therefore we predict, that they may prefer the orange over the blue container, and therefore switch easier in reversals from blue to orange than the other way around. In the study with ravens by Range et al. (2006), males outperformed females in a reversal learning task, we therefore predict that sex will have an effect on the number of trials before reaching criterion.

Lastly, we have adapted the methodology to compare their performance with other avian species (parrots, corvids, pigeons, Australian Mynahs) in the near future.

4. Material and Methods

4.1. Subjects

Eleven adult and two subadult, captive- born and hand-reared Goffin's cockatoos (*Cacatua goffiniana*), participated in this study. The subjects (8 males, 5 females) were kept in a single-species group at the Goffin Lab of the University of Veterinary Medicine (Vienna, Austria) and were housed in a large, enriched aviary with an indoors and outdoors area (in total ca. 200 m^2 ground, space up to 6m high). All parrots were kept on an ad libitum diet (fresh and dried fruits, boiled vegetables, boiled seeds, pellets and fresh water). For further information about the subjects see Table 1.

Table 1. Subjects participating in the study. Subjects listed with sex (M= Male; F= Female) and year of hatching.

SUBJECT	SEX	HATCHED
DOLITTLE	M	2011
MAYDAY	F	2011
FIGARO	M	2007
MONEYPENNY	F	2010
KONRAD	M	2010
PIPIN	M	2008
KIWI	M	2010
JANE	F	2016
FINI	F	2007
ZOZO	M	2010
TITUS	M	2016
HEIDI	F	2010
MUPPET	M	2010

4.2. Experimental History

Ten of the adult birds participating in this experiment were previously tested in abstract two-choice tasks on a touchscreen investigating inference by exclusion (O’Hara et al. 2015b) and comparing matching and non-matching abilities (Gruber 2016). Moreover, they partook in several studies that required them to switch back between using or making one of two tool types (Laumer et al., 2016; Auersperg et al., 2018), however none of the birds were ever tested in a classical reversal learning paradigm.

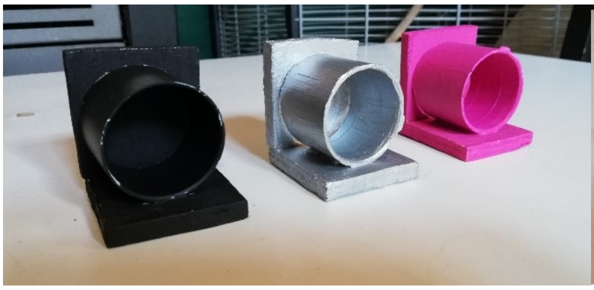
4.3. Ethical Statement

All animals are permanently kept (before and after the experiment) in a well-established group at the “Goffin Lab”. All have CITES certificates and were registered at the district’s administrative animal welfare bureau (Bezirkshauptmannschaft St. Pölten Schmiedgasse 4–6, A-3100, St. Pölten, Austria). The housing conditions comply with the Austrian Federal Act on the Protection of Animals (Animal Protection Act—§24 Abs. 1 Z 1 and 2; §25 Abs. 3—TSchG, BGBl. I Nr. 118/2004 Art. 2). As my experiments are strictly non-invasive and based purely on behavioural observations, they are not classified as animal experiments in accordance with Austrian Law and do not require permission (Austria: §2. Federal Law Gazette No. 501/1989). The birds are not wing-clipped and only participate in the study on a voluntary basis. The animals are never food deprived and are closely monitored by trained personnel on a daily basis.

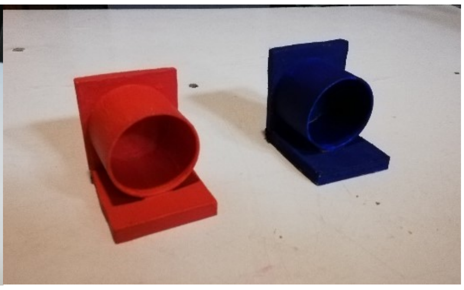
Throughout my testing period, no elevated levels of stress or aggression could be detected.

4.4. Material

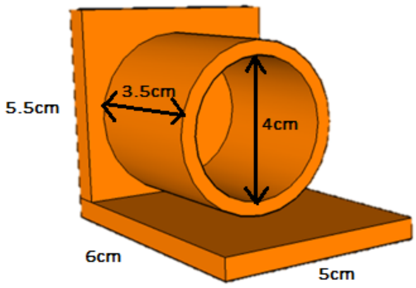
For my experiment, I used containers made of plywood and short PVC pipes, which were painted with different acrylic colours. The containers for the training phase were painted black, pink and silver (Picture 1), and the containers for testing were dark blue and orange (Picture 2). They measured 5.5cm in height, 5cm in width and 6cm in length, with the pipes being only 3.5cm long and 4cm in diameter (Picture 3).



Picture 1: Training containers



Picture 2: Test containers



Picture 3: Measurements

4.5. Procedure

After a group habituation, during which sunflower seeds were sprinkled around and inside the training containers until birds were comfortable touching the containers, the birds proceeded to individual training in a familiar testing room.

At first, birds had to learn to walk around a single container (the one not used as S+ and S- in the Training respectively) and obtain a food reward. Once they succeeded in this 5 times, they started with the actual training. They were divided into 3 groups, each group was assigned a different coloured rewarded container (S+), see Table 2.

Table 2. Training groups; according to their S+ in the training.

S+ Black	S+ Silver	S+ Pink
Kiwi	Dolittle	Fini
Jane	Mayday	Zozo
Muppet	Figaro	Titus
Olympia**	Moneypenny	Heidi
Irene*	Konrad	Muki**
	Pipin	

* Dropped out after the training
**Dropped out during the training

4.5.1. Training

The training took place between 24th October and 7th of November 2019. Every bird was presented with 2 containers, their assigned S+ and S-. A piece of cashew nut was put into the S+ container, which was only visible if the bird walked around the container. Once the session started, the bird was perched on a chair and the containers were put on the table approximately 40cm away from the bird and 30cm apart and at an angle, so that the bird could only look into one container at a time. The positions of the containers were pseudo randomised previous to the session, but were never in the same place more than two times consecutively in order to avoid side bias. Furthermore, the experimenter stood a step away from the table, hands either behind the back or crossed in front, and was wearing sunglasses during all training and testing procedures to avoid cueing the bird through gaze following.

After the containers were set on the table, the bird had to wait 1-2 seconds and was then given a verbal and gestural command to start. If the bird chose correctly, it was allowed to eat the nut and the inter trial interval (ITI) started, during which the bird returned to the chair and the experimenter made a note of the choice and rebaited the containers. This took approximately 5 seconds and then the next trial started. If the bird chose incorrectly, it was ordered back on the chair, the containers were removed, and a correctional trial started after an ITI of ca. 5 seconds. This was repeated until the bird made the correct choice.

Birds were tested in 10 trials per session, and passed the training when the made 17/20 correct first choices. They were never tested in more than 4 sessions per day to avoid saturation and boredom.

4.5.2. Testing

The first part of testing consisted of the Acquisition phase, during which bird were first presented with their respective S+ testing container (see Table 3). The first birds started testing on the 3rd of November 2019, and the last bird was tested on the 5th of January 2020.

Birds were divided into 2 Test groups, each group stared with a different colour as S+ in the Acquisition phase. The test procedure was the same as in the training (including correctional trials), and once birds chose the correct container 17/20 times on the first try, they proceeded to the first reversal, whereupon the S+ became the S- and vice versa. Once they chose the rewarded container in 17/20 trials, the contingencies changed again. This procedure was continued for 15 reversals. Importantly, new contingencies (reversals) were only started on a new day, so no colour would be S+ and S- on the same day.

Table 3: Test groups; according to their S+ in the Acquisition Phase

S+ Orange	**S+ Blue**
Dolittle	Fini
Mayday	Zozo
Figaro	Titus
Moneypenny	Heidi
Konrad	Muppet
Pipin	Olympia*
Kiwi	Irene*
Jane	Muki*

*Birds did not participate in the testing

4.6. Analysis

All training and testing trials were recorded using a digital camcorder (JVC Camcorder Model No.: GZ-HM30BE) and coded *in situ*. I recorded whether birds chose the S+, whether the S+ was situated left (L) or right (R), the number of correctional trials when they chose the S-, and the number of trials until they reached criterion. Throughout the experiment, data was curated using Microsoft Excel.

4.6.1. GLMM Model

For further statistical analysis, we fitted the models in R (version 3.6.3; R Core Team. 2020) using the function glmer of the package lme4 (version 1.1-21; Bates et al. 2015). We established model stability by dropping levels of the random effects one at a time and comparing the estimates of the full model fitted to these subsets with those obtained for the full data set. This then created a model of good stability (see results). We determined confidence intervals of model estimates and the fitted model by means of a parametric bootstrap (N=1000; function bootMer of the package lme4). The sample analysed for this model comprised a total of 3900 trials, conducted in 390 sessions with 13 individuals and in which the individuals chose the correct response in 3183 trials.

To test whether reversal number affected the probability of correct choices we fitted a Generalized Linear Mixed Model (GLMM; Baayen 2008) with binomial error structure and logit link function (McCullagh & Nelder 1989). As key fixed effect we included reversal number. However, as we expected its effect to depend on session and trial number, we also included the three-way interaction between reversal, session, and trial number and also all two-way interactions and main effects which it encompasses. To further control for their effects, we also included test group (factor with levels 'blue' and 'orange'), training group (factor with levels '1', '2', and '3'), rewarded (factor with levels 'blue' and 'orange'), position of the correct choice (factor with levels 'left' and 'right') and sex of the individual as fixed effects. To avoid pseudreplication, we included random intercept effects for the identity of the individual and also of session ID (nested in individual). To keep type I error level at the nominal level of 0.05, we included random slopes (Schielzeth & Forstmeier 2009; Barr et al. 2013) of reversal, session, and trial number and all their interactions up to third order, position of the correct choice, and rewarded within individual ID and of trial number and position of the correct choice within session ID (to include factors as random slopes we manually dummy coded and then centred them). We also included parameters for the correlations among random intercept and random slow effects.

To avoid 'cryptic multiple testing' (Forstmeier & Schielzeth 2011), we compared the full model as described above with a null model lacking reversal number and its interactions with trial and session number in the fixed effects part but being otherwise identical. This comparison was based on a likelihood ratio test (Dobson 2002). We also utilised likelihood ratio tests to test the individual effects which we based on comparing a given model with sub-models lacking the terms to be tested one at a time. In case the full-null model revealed significance, we successively removed non-significant interactions (beginning with the highest order interaction) in order to be able to infer the significance of lower level effects.

5. Results

5.1. Training

The training was conducted with 16 individuals, however only 14 completed it successfully (see Table 2), as both birds (Muki, Olympia) did not come to the test compartment reliably. Of the remaining 14 individuals, 13 completed the training (criterion 17/20 trials) within 2 sessions, only 1 individual (Figaro) needed 3 sessions. On average, they made 1.6+/-0.3 SE incorrect choices in session 1, but only 0.2+/-0.1 SE in session 2. Likewise, the number of correctional trials decreased from 3.3 +/-0.7 SE on average in session 1, to 0.4 +/-0.3 SE in session 2.

5.2. Acquisition Phase & First reversal

13 individuals participated in the testing phase of the study, 1 individual (Irene) stopped coming to the test compartment reliably after she successfully completed the training, and was therefore excluded from the testing phase. Of the remaining 13 individuals, 12 completed the acquisition phase in 2 sessions (criterion 17/20 trials), and 1 subject (Zozo) completed it in 3 sessions.

The comparative data for the acquisition phase and the first reversal (Table 4) shows that on average, birds needed 2.1+/-0.08 SE sessions to complete the acquisition phase, and an average of 2.5 +/-0.2 SE sessions for the first reversal, scoring their performance at a 1:1.025 ratio. They made twice as many wrong choices in the first reversal as in the acquisition phase (1.9 +/-0.3 SE and 3.6+/-0.5 SE), and needed 5 times as many correctional trials (4.3 +/-1.5 SE and 20+/-2.8 SE; Figure 1).

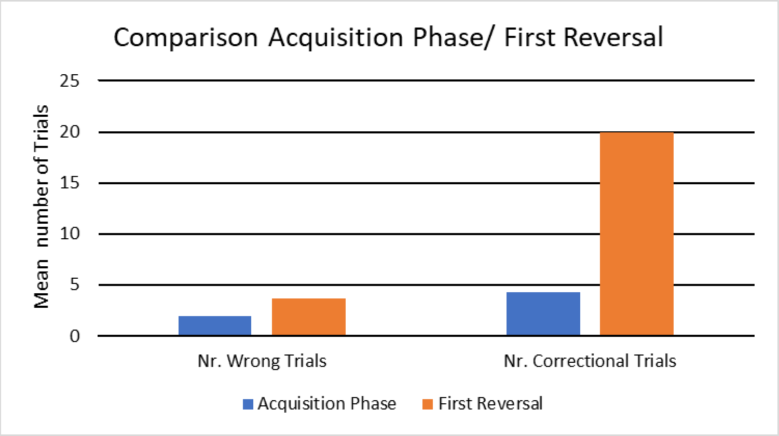


Figure 1: Comparison between Acquisition Phase and First Reversal: regarding the mean number of trials necessary to pass criterion (17/20 trials)

Table 4: Individual data. Listing sessions, wrong choices and correctional trials for both Acquisition Phase and First reversal. Mean values include Standard Error (SE).

	Acquisition Phase			*First Reversal*		
	Sessions	Wrong choices	Correctional Trials	Sessions	Wrong choices	Correctional Trials
Dolittle	2	1	1	2	3	12
Mayday	2	0	0	3	3	6
Figaro	2	2	2	2	3	19
Moneypenny	2	1	1	2	3	11
Konrad	2	0	0	2	2	21
Pipin	2	3	9	3	4	27
Kiwi	2	3	4	3	4	23
Jane	2	3	5	2	3	12
Fini	2	2	3	3	4	31
Zozo	3	4	20	2	2	17
Titus	2	3	7	4	8	42
Heidi	2	2	2	2	1	11
Muppet	2	1	2	3	7	27
Mean	2.076 +/-0.08	1.923 +/-0.3	4.307 +/-1.5	2.538 +/-0.2	3.615 +/-0.5	19.923 +/-2.8

5.3. GLMM

First, we calculated a three-way interaction between reversal, trial, and session number (see Appendix Table A1), none of which were significant. Because neither the two-way interactions between reversal number on the one hand, and trial or session number on the other (in the model lacking the three-way interaction; see Appendix Table A2), we removed these terms from the final model. The final model revealed that the probability of a correct choice increased with reversal number (χ^2=20.461, df=1, p<0.001; see Table 5; Figure 2). Overall, reversal number and or its interactions with trial and session number had a clear effect on the probability of a correct choice (full-null model comparison, likelihood ratio test: χ^2=22.492, df=6, p<0.001).

Table 5. Results of the final reduced model comprising the only significant interaction (yellow). (estimates, together with standard errors, confidence limits, and significance tests).

term	Estimate	SE	lower CI	upper CI	χ^2	df	P
intercept	2.202	0.349	1.513	3.019			(1)
Reversal(2)	0.290	0.058	0.169	0.421	18.229	1	<0.001
Session(3)	0.812	0.065	0.709	0.986			(1)
Trial(4)	0.741	0.080	0.608	0.920			(1)
test.grouporange(5)	0.184	0.289	-0.448	0.812	0.384	1	0.536
Traininggroup2(6)	0.050	0.203	-0.334	0.473	0.424	2	0.809
Traininggroup3(6)	-0.145	0.334	-0.861	0.560			
rewardedorange(7)	0.024	0.108	-0.190	0.241	0.049	1	0.825
pos.corrright(8)	-0.263	0.153	-0.561	0.015	2.913	1	0.088
sexM(9)	-0.215	0.152	-0.537	0.053	1.905	1	0.168
Session:Trial	-0.502	0.077	-0.665	-0.357	20.461	1	<0.001

(1) not indicated because of having a very limited interpretation

(2) z-transformed to a mean of zero and a standard deviation (sd) of one; mean and sd of the original variable were 8.000 and 4.321, respectively

(3) z-transformed to a mean of zero and a standard deviation (sd) of one; mean and sd of the original variable were 1.500 and 0.500, respectively

(4) z-transformed to a mean of zero and a standard deviation (sd) of one; mean and sd of the original variable were 5.500 and 2.873, respectively

(5) dummy coded with blue being the reference category

(6) dummy coded with '1' being the reference category; the indicated test refers to the overall effec of training group

(7) dummy coded with blue being the reference category

(8) dummy coded with left being the reference category

(9) dummy coded with female being the referencecategory

Moreover, in none of the models calculated did sex (χ^2=1.905, df=1, p=0.168), training group identity (χ^2=0.424, df=2, p=0.809), or testing group identity (χ^2=0.384, df=1, p=0.536), or the location (χ^2=2.913, df=1, p=0.088) and the colour of the rewarded container (χ^2=0.049, df=1, p=0.825) have an effect on the probability of correct choices (Table 4). Hence, we did not conduct any post-hoc tests.

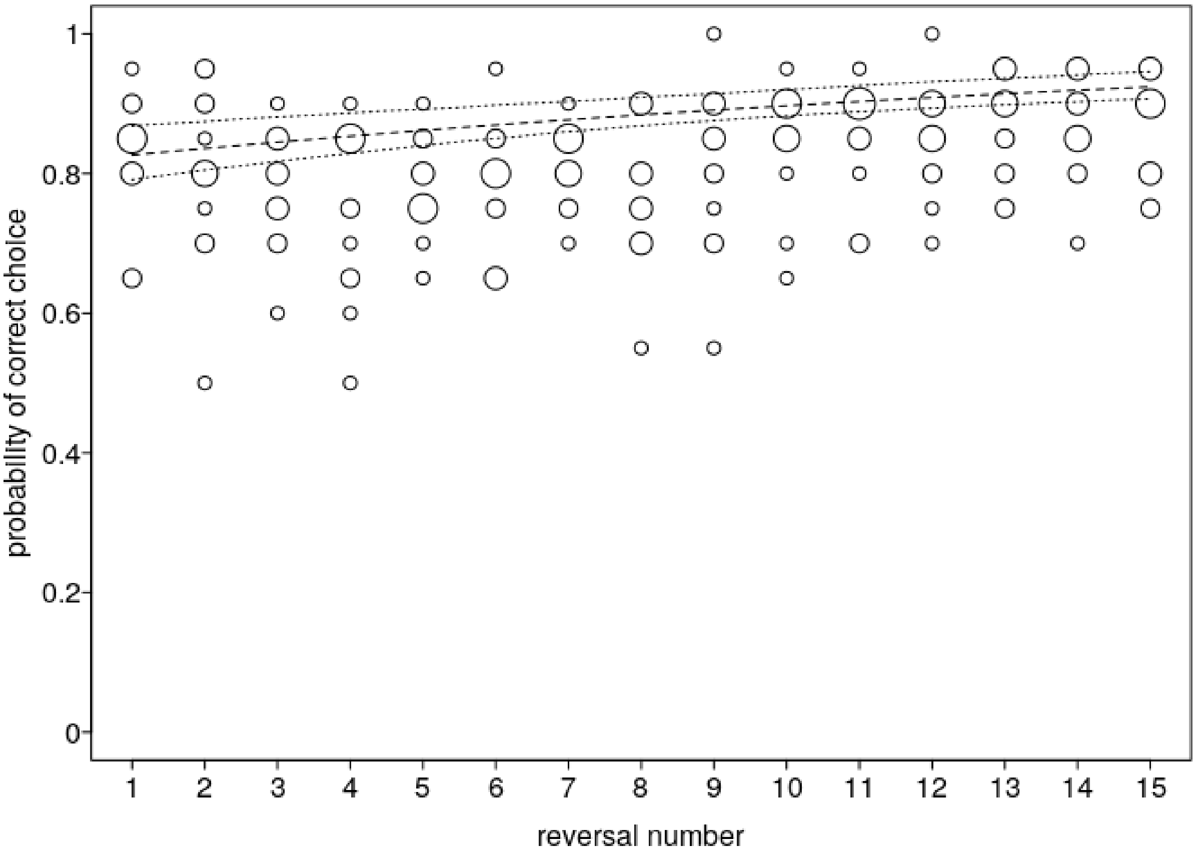


Figure 2: Probability of a correct choice as a function of reversal number. The dots show the probability of a correct choice per individual and reversal number, the size of the dots depicts the number individuals with a given probability (n=1 to 6). The dashed line depicts the fitted model, the spotted lines the confidence interval.

6. Discussion

I tested Goffin's cockatoos, a parrot species that has previously shown great behavioural flexibility in many experiments (e.g. Auersperg et al. 2012, 2016), in a serial reversal learning task. When looking at the difference between the acquisition phase and the first reversal, Goffins performed at a 1:1.025 ratio, while similar studies with ravens (Range et al. 2006, 2008), red-shouldered macaws and black-headed caiques (van Horik & Emery 2018) showed a 1:2 ratio between initial acquisition and first reversal. Importantly though, both of these studies did not use correctional trials in their testing, making a direct comparison difficult. However, if Goffins' results were corrected for this approximation (converting the correctional trials into actual trials and adding them to the sessions to criterion count), the ratio would amount to 1:1.8 (2.5+/-0.2 SE and 4.5 +/-0.4 SE sessions to criterion in acquisition phase and first reversal respectively). Hence, we can estimate, that

Goffins' performance is similar to the other parrot species and ravens in the aforementioned studies.

While both parrots and corvids frequently perform cognitively complex behaviours at similar levels (Güntürkün & Bugnyar 2016; Lambert et al. 2018), one explanation why Goffin's cockatoos may have outperformed other parrots and corvids is their island origin. Many species found on islands were discovered to possess unparalleled behavioural flexibility, probably due to ecological selection pressures (Lefebvre & Bolhuis 2003). However, in a reversal study also using correctional trials, O'Hara et al. (2015a) tested kea, another island dwelling parrot species. Surprisingly, they were found to perform at an even worse ratio (1:3). Bebus et al. (2016) conducted a reversal learning study with Florida Scrub Jays and found out that more neophobic birds performed better at the task than neophilic ones. Incidentally, individuals that learned the initial reward contingencies quickly were slower to reach criterion on each reversal event, highlighting the trade-off between associative learning and reversal learning. The Kea in O'Hara's study might have faced the opposite problem, as kea are an extremely neophilic species, therefore their cognitive performance may have been confounded through their level of neophilia.

Interestingly, while Goffin's cockatoos are generally classified as neophilic as well, they have previously shown a great ability for inhibition (Auersperg et al. 2013a), another executive function closely associated with behavioural flexibility (Diamond 2013). In a reversal learning task this comes into effect because after subjects first learn to choose the rewarded container in the acquisition phase, they then experience failures in the first trials of the first reversal. Later, they have to inhibit their response to the previously rewarded container and switch. This illustrates that behavioural flexibility alone is not enough to be successful in this sort of task, and that inhibition control plays a crucial role (O'Hara et al. 2015a).

Another possible explanation for Goffins' superior results could simply be a methodological one, as no 2 studies mentioned above used the same set-up, making a direct comparison difficult. Differences include learning criteria, use of correctional trials, side-bias corrections, administration of more than one reversal on the same day, and number of total reversals.

Additionally, Goffins' experimental history has to be addressed, as they have previously participated in 2 choice discrimination tasks on the touchscreen (O'Hara et al. 2015b; Gruber 2016). The idea that these experiments could have influenced

Goffins' performance in this study however, seems very unlikely. First, the 2 studies mentioned were both conducted on a touchscreen, and several years before data collection for this study began. Also, even though a choice between 2 stimuli has to be made, the crucial elements of e reversal learning study are quite different. Moreover, results from a reversal learning study on Kea (O'Hara et al. 2015a) showed that previous experience in a discrimination task on a touchscreen (O'Hara et al. 2012) had no effect on the study in question.

A quality that did not contribute to Goffin's' performance as much as I expected, is persistence. In the wild, they frequently exhibit extractive foraging behaviours (O'Hara et al. 2019), which might have put selective pressure on behavioural persistence, as it enables their success in these foraging situations. In the lab, Goffins have also demonstrated goal-oriented persistency in artificial fruit tasks (Auersperg et al. 2011, 2013b) and in many tool-manufacture experiments (e.g. Auersperg et al. 2016). In a reversal learning task, persistence could be seen as an antagonist to inhibition, making it difficult for animals to switch from the rewarded to the previously unrewarded stimulus. Indeed, Goffins showed 5 times the number of correctional trials in the first reversal, compared to the acquisition phase. Looking at the individual data though, it becomes apparent that the level of persistence displayed might vary greatly between individuals (Correctional trials in the first reversal between 6 and 42), and might therefore depend on birds' personality. Range et al. (2006) remarked that in ravens, personality very much influenced birds' performance in their reversal learning task.

Other than in Range's study however (Male ravens performed better in reversals than females), no influence of subjects' sex could be detected in this study, nor in any other study conducted with these individuals.

Other than that, a 2015 study conducted by Auersperg et al. with these cockatoos did reveal a pronounced colour preference, as they preferred yellow and red objects over blue ones. Conceivably, I therefore predicted that container colour might have an effect on the probability of a correct choice or on their ability to switch from a preferred to a less preferred colour. Interestingly, the GLMM models revealed that neither the colour of the training nor the testing containers had any effect on individuals' choices.

7. Conclusion

My results suggest that Goffin's cockatoos are able to succeed in a reversal learning study within 15 reversals, and perform at similar ratios as other parrot or corvid species. They completed the acquisition phase similarly fast as Kea, but showed greater flexibility during the first reversal. Importantly however, the cockatoos' seemingly superior performance might also be a result of different methodologies used in the other studies.

What seems to be certain is that Goffins manage the trade-off between associative learning (necessary to learn the rewarded stimulus) and rule learning (noticing that contingencies change) quite well. Accordingly, factors like the colours of the training and testing containers had no effect on their choices, demonstrating that the associations they formed did not impair their performance in the reversal sessions.

The success of managing these opposing strategies is also illustrated in their ability to apparently forego their proclivity for persistency, and react flexibly to changing circumstances. As both persistence and behavioural flexibility are hallmark qualities of opportunistic foragers, they act as antagonists in situations like reversal learning experiments. Indeed, what seems to have made the difference in this tug-of-war is another quality Goffin's cockatoos have previously demonstrated to possess, namely inhibition control. Auersperg et al. (2013a) were able to show that Goffins waited for qualitatively and quantitatively better food options, while inhibiting consuming less preferred reward options. One individual actually waited up to 80 seconds for a qualitatively preferred food, remarkably while holding the non-preferred food in its beak.

Importantly, when comparing the performance of individuals in this study, the intraspecific variability was quite pronounced, which makes it harder to draw definitive conclusions on the species level. As Cussen (2017) remarked, more individuals of more species need to be tested to get a clearer picture of parrots' cognitive abilities and limitations. Incidentally, this study is part of a large-scale comparison between a number of avian species (e.g. parrots, corvids, mynahs, pigeons), which will hopefully shed more light on avian cognitive abilities.

8. Acknowledgements

I would like to thank Alice Auersperg for the opportunity to conduct this experiment, as well as her supervision and support throughout the project. A big thank you as well to the whole Goffin Lab team for their help during data collection. I owe special thanks to Roger Mundry for his invaluable help with the statistical analysis. And last but not least, I owe thanks to my friends and family for always supporting me in all my parrot endeavours!

9. References

Auersperg, A.M.I., von Bayern, A.M.P., Gajdon, G.K., Huber, L. & Kacelnik, A., 2011. Flexibility in problem solving and tool use of kea and New Caledonian crows in a multi access box paradigm. *PLoS ONE*, 6: e20231.

Auersperg, A.M.I., Szabo, B., von Bayern, A.M.P. & Kacelnik, A., 2012. Spontaneous innovation in tool manufacture and use in a Goffin's cockatoo. *Current Biology*, 22, R903-R904.

Auersperg, A.M.I., Laumer, I.B. & Bugnyar, T., 2013a. Goffin cockatoos wait for qualitative and quantitative gains but prefer 'better' to 'more'. *Biology Letters*, 9: 20121092.

Auersperg, A.M.I., Kacelnik, A. & von Bayern, A.M.P., 2013b. Explorative learning and functional inferences on a five-step means-means-end problem in Goffin's cockatoos. *PLoS ONE*, 8: e68979.

Auersperg, A. M. I., von Bayern, A. M. P., Weber, S., Szapari, A., Bugnyar, T., & Kacelnik, A., 2014. Social transmission of tool use in the Goffin cockatoo. *Proceedings of the Royal Society, Series B*, 281, 20140972. http://dx.doi.org/10.1098/rspb.2014.0972.

Auersperg, A. M. I., van Horik, J., Bugnyar, T., Kacelnik, K., Emery, N., & von Bayern, A. M. P., 2015. Combinatory actions during object play in psittaciformes (*Diopsittaca nobilis, Pionites melanocephala, Cacatua goffini*) and corvids (*Corvus corax, C. monedula, C. moneduloides*). *Journal of Comparative Psychology*, 129, pp.62–71. http://dx.doi.org/10.1037/a0038314

Auersperg, A. M. I., Borasinski, S., Laumer, I., & Kacelnik, K., 2016. Goffin's cockatoos make the same tool type from different materials. *Biology Letters*, 12, 20160689. http://dx.doi.org/10.1098/rsbl.2016.0689

Auersperg, A.M.I., Köck, C., O'Hara, M., Huber, L., 2018. Tool making cockatoos adjust the lengths but not the widths of their tools to function. *PLoS ONE*, 13(11): e0205429. https://doi.org/10.1371/journal.pone.0205429

Baayen, R.H., 2008. Analyzing Linguistic Data. Cambridge University Press. Cambridge.

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, pp. 255–278.

Bates, B., Maechler, M., Bolker, B., Walker, S., 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67, pp.1-48.

Bebus, S.E., Small, T.W., Jones, B.C., Elderbrock, E.K. & Schoech, S.J., 2016. Associative learning is inversely related to reversal learning and varies with nestling corticosterone exposure. *Animal Behaviour*, 111, pp. 251-260.

Bond, A.B., Kamil, A. & Balda, R.P., 2007. Serial reversal learning and the evolution of behavioral flexibility in three species of north american corvids (*Gymnorhinus cyanocephalus, Nucifraga columbiana, Aphelocoma californica*). *Journal of Comparative Psychology*, 121(4), pp.372–379.

Cook, R.G. & Rosen, H.A., 2010. Temporal control of internal states in pigeons. *Psychonomic Bulletin & Review*, 17(6), pp.915–922.

Cussen, V.A., 2017. Psittacine cognition: individual differences and sources of variation. *Behavioural processes*, 134, pp.103-109.

Diamond, A., 2013. Executive functions. *Annual Review of Psychology*, 64, pp. 135-168.

Dobson, A.J., 2002. An Introduction to Generalized Linear Models. Chapman & Hall/CRC. Boca Raton.

dos Reis, M., Thawornwattana, Y., Angelis, K., Telford, M.J., Donoghue, P.C. & Yang, Z., 2015. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Current Biology*, 25, pp. 2939-2950.

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, pp.47–55.

Gossette, R.L., Gossette, M.F., Riddell, W., 1966. Comparisons of successive discrimination reversal performances among closely and remotely related avian species. *Animal Behaviour*, 14, pp. 560-564.

Gossette, R.L., 1968. Examination of retention decrement explanation of comparative successive discrimination reversal learning by birds and mammals. *Perceptual and Motor Skills*, 27, pp. 1147-1152.

Gossette, R.L., Slonim, P., 1969. Spatial SDR performances across three species of new world monkeys. *Psychonomic Science*, 14(5), pp. 221-223.

Gruber, R., 2016. Matching to Sample in the Goffin's cockatoo (*Cacatua goffiniana*) [Master Thesis]. Vienna (Austria): University of Vienna.

Güntürkün, O., Bugnyar, T., 2016. Cognition without cortex. *Trends in Cognitive Science*, 20, pp.291-303.

Güntürkün, O., Ströckens, F., Scarf, D., Colombo, M., 2017. Apes, feathered apes, and pigeons: differences and similarities. *Current Opinion in Behavioral Sciences*, 16, pp.35–40.

Jelbert, S.A., Taylor, A.H. & Gray, R.D., 2016. Does absolute brain size really predict self-control? Hand-tracking training improves performance on the A-not-B task. *Biology Letters*, 12: 20150871.

Lambert, M., Jacobs, I., Osvath, M., von Bayern, A.M.P., 2018. Birds of a feather? Parrot and corvid cognition compared. *Behaviour*, 156 (5-8), pp. 505-595: DOI:10.1163/1568539X-00003527

Laschober, M., 2019. Midsession Reversal Learning in Kea (*Nestor notabilis*) [Master Thesis]. Vienna (Austria): University of Veterinary Medicine Vienna.

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: 28380.

Lefebvre, L.M. & Bolhuis, J.J., 2003. Positive and negative correlates of feeding innovations in birds: evidence for limited modularity. In: *Animal Innovation*, Oxford University Press, p. 39-62.

McCullagh, P., Nelder, J.A., 1989. Generalized linear models. Chapman and Hall. London.

O'Hara, M., Gajdon, G.K. & Huber, L., 2012. Kea Logics: How these birds solve difficult problems and outsmart researchers. In S. Watanabe, ed. *Logic and Sensibility*. Tokyo: Keio University, pp. 23–38.

O'Hara, M., Huber, L. & Gajdon, G.K., 2015a. The advantage of objects over images in discrimination and reversal learning by kea, *Nestor notabilis*. *Animal Behaviour*, 101, pp.51–60.

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

O'Hara, M., Mioduszewska, B., Haryoko, T., Prawiradilaga, D.M., Huber, L. & Auersperg, A.M.I., 2019. Extraction without tooling around — The first comprehensive description of the foraging- and socio-ecology of wild Goffin's cockatoos (*Cacatua goffiniana*). *Behaviour*, 156: BEH 3523.

Olkowicz, S., Kocourek, M., Lučan, R.K., Porteš, M., Fitch, W.T., Herculano-Houzel, S. & Němec, P., 2016. Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences USA*, 113, pp. 7255-7260.

Osuna-Mascaró, A.J., Auersperg, A.M.I., 2018. On the brink of tool use? Could object combinations during foraging in a feral Goffin's cockatoo (*Cacatua goffiniana*) result in tool innovations? *Animal Behavior and Cognition*, 5, pp. 229-234.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Range, F., Bugnyar, T., Schlögl, C. & Kotrschal, K., 2006. Individual and sex differences in learning abilities of ravens. *Behavioural Processes,* 73, pp. 100-106.

Range, F., Bugnyar, T. & Kotrschal, K., 2008. The performance of ravens on simple discrimination tasks: a preliminary study. *Acta Ethologica,* 11, pp. 34-41.

Schielzeth, H., Forstmeier, W., 2009. Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology*, 20, pp. 416-420.

Shettleworth, S.J., 2009. The evolution of comparative cognition: is the snark still a boojum? *Behavioural Processes*, 80, pp. 210-217.

Sol, D., 2009. Revisiting the cognitive buffer hypothesis for the evolution of large brains. *Biology letters*, 5, pp.130–133.

Teschke, I., Wascher, C.A.F., Scriba, M.F., von Bayern, A.M.P., Huml, V., Siemers, B. & Tebbich, S., 2013. Did tool-use evolve with enhanced physical cognitive abilities? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368: 20120418.

van Horik, J.O. & Emery, N.J., 2018. Serial reversal learning and cognitive flexibility in two species of Neotropical parrots (*Diopsittaca nobilis* and *Pionites melanocephala*). *Behavioural Processes,* 157, pp. 664-672.

10. Appendix

Table A1. Results of the full model: (estimates, together with standard errors, confidence limits, significance tests as well as minimum and maximum of estimates obtained when excluding individuals and sessions one at a time).

term	Estimate	SE	lower CI	upper CI	χ^2	df	P	min	max
intercept	2.192	0.355	1.488	3.008			(1)	1.963	2.388
Reversal(2)	0.214	0.077	0.077	0.368			(1)	0.181	0.238
Session(3)	0.805	0.065	0.720	0.983			(1)	0.787	0.829
Trial(4)	0.772	0.081	0.645	0.963			(1)	0.734	0.856
test.grouporange(5)	0.178	0.290	-0.446	0.837	0.382	1	0.536	0.064	0.388
Traininggroup2(6)	0.039	0.199	-0.380	0.474	0.386	2	0.824	-0.045	0.138
Traininggroup3(6)	-0.146	0.334	-0.854	0.553				-0.326	0.162
rewardedorange(7)	0.028	0.107	-0.185	0.256	0.069	1	0.793	-0.011	0.074
pos.corrright(8)	-0.189	0.166	-0.522	0.141	1.292	1	0.256	-0.292	-0.103
sexM(9)	-0.208	0.152	-0.498	0.088	1.824	1	0.177	-0.339	-0.045
Reversal:Session	-0.073	0.080	-0.227	0.089			(1)	-0.097	-0.038
Reversal:Trial	0.000	0.069	-0.138	0.151			(1)	-0.022	0.037
Session:Trial	-0.509	0.082	-0.683	-0.356			(1)	-0.546	-0.465
Reversal:Session:Trial	0.074	0.081	-0.084	0.249	0.795	1	0.373	0.038	0.118

(1) not indicated because of having a very limited interpretation

(2) z-transformed to a mean of zero and a standard deviation (sd) of one; mean and sd of the original variable were 8.000 and 4.321, respectively

(3) z-transformed to a mean of zero and a standard deviation (sd) of one; mean and sd of the original variable were 1.500 and 0.500, respectively

(4) z-transformed to a mean of zero and a standard deviation (sd) of one; mean and sd of the original variable were 5.500 and 2.873, respectively

(5) dummy coded with blue being the reference category

(6) dummy coded with '1' being the reference category; the indicated test refers to the overall effect of training group

(7) dummy coded with blue being the reference category

(8) dummy coded with left being the reference category

(9) dummy coded with female being the reference category

Table A2. Results of the reduced model lacking the non-significant three-way interaction.
(estimates, together with standard errors, confidence limits, and significance tests).

term	Estimate	SE	lower CI	upper CI	χ^2	df	P
intercept	2.257	0.352	1.576	3.089			(1)
Reversal(2)	0.227	0.078	0.074	0.375			(1)
Session(3)	0.804	0.067	0.706	0.974			(1)
Trial(4)	0.772	0.084	0.628	0.963			(1)
test.grouporange(5)	0.174	0.290	-0.437	0.756	0.334	1	0.563
Traininggroup2(6)	0.041	0.199	-0.312	0.456	0.432	2	0.806
Traininggroup3(6)	-0.145	0.333	-0.864	0.531			
rewardedorange(7)	0.013	0.109	-0.212	0.221	0.011	1	0.916
pos.corrright(8)	-0.252	0.154	-0.596	0.045	2.627	1	0.105
sexM(9)	-0.210	0.152	-0.511	0.085	1.834	1	0.176
Reversal:Session	-0.100	0.077	-0.269	0.058	1.726	1	0.189
Reversal:Trial	-0.019	0.068	-0.157	0.111	0.105	1	0.745
Session:Trial	-0.532	0.080	-0.704	-0.391	21.515	1	<0.001

(1) not indicated because of having a very limited interpretation

(2) z-transformed to a mean of zero and a standard deviation (sd) of one; mean and sd of the original variable were 8.000 and 4.321, respectively

(3) z-transformed to a mean of zero and a standard deviation (sd) of one; mean and sd of the original variable were 1.500 and 0.500, respectively

(4) z-transformed to a mean of zero and a standard deviation (sd) of one; mean and sd of the original variable were 5.500 and 2.873, respectively

(5) dummy coded with blue being the reference category

(6) dummy coded with '1' being the reference category; the indicated test refers to the overall effec of training group

(7) dummy coded with blue being the reference category

(8) dummy coded with left being the reference category

(9) dummy coded with female being the reference category