universität wien

MASTERARBEIT / MASTER´S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

“Wet noses to the touchscreen: Influence of age and experience on discrimination learning, novelty avoidance and inferential reasoning in pet dogs”

verfasst von / submitted by

Carmen Sophie Schwarzl, BSc

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

Master of Science (MSc)

Wien, 2023 / Vienna, 2023

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:	Univ.-Prof. Dott. Leonida Fusani, MPhil PhD

Acknowledgements

There are so many people to thank, without whom this project would have never come to life. A big to thank you to the Clever Dog Lab- team for providing equipment, testing facility and of course helping with the experiments and the dogs. Special thanks to Ludwig Huber and Leonida Fusani for taking time and effort to review my thesis. Additionally, I want to thank Mark O´Hara for providing an essential software for this study. I also want to thank all the dogs and owners that participated in this study for offering their time and interest.

My biggest thanks go to Lisa Wallis for her supervision, open ear whenever I needed one and in general giving me the opportunity to learn and improve my skills in a way that would have not been possible for me without her.

Table of contents

Zusammenfassung

Die Fähigkeit zu lernen, ermöglicht es Tieren ihre Informationsaufnahme zu verbessern. Haushunde (*Canis lupus familiaris*) wurden als Versuchstiere ausgewählt, um die Lernfähigkeit und logisches Schlussfolgern zu untersuchen. Um dies zu beurteilen, wurden eine einfache Unterscheidungsaufgabe, Neuheitsversuche und Aufgaben zur Schlussfolgerung durch Ausschluss unter Verwendung eines Touchscreens mit zwei Auswahlmöglichkeiten durchgeführt. Die Einflüsse von Geschlecht, Rasse, Kastrationsstatus, Alter und Vorerfahrung (in einer ähnlichen Studie) auf das Testergebnis wurden analysiert. Jeder der 43 teilnehmenden Hunde wurde auf ein eigenes Stimuluspaar (S+ und S-) trainiert und mit zwei Kriterien (Unterscheidungsvermögen und Neophilie) konfrontiert, bevor die Fähigkeit Inferenz durch Ausschluss getestet wurde. Die Ergebnisse zeigten, dass nur Alter und Erfahrung die Leistung der Hunde in dem Sinne beeinflussten, dass jüngere Hunde die einfache Unterscheidung schneller lernten und weniger Fehler machten. Hunde mit mehr Erfahrung lernten schneller, einen ablenkenden, negativen neuen Reiz zu meiden und machten allgemein auch weniger Fehler (in beiden Kriterien). Überraschenderweise zeigten nur 30 % der Hunde Inferenz durch Ausschlussverfahren. Darüber hinaus wurde kein Einfluss irgendeines Prädiktors (Alter, Rasse, Erfahrung, Kastratenstatus, Geschlecht, Leistung in Unterscheidungsversuchen und Leistung in Versuchen mit neuen Reizen) auf die Leistung der Tiere bei der Inferenz durch Ausschluss festgestellt.

Abstract

The ability to learn allows animals to enhance their ratio of information acquirement. Pet dogs (*Canis lupus familiaris*) were chosen as study animals to investigate learning ability and inferential reasoning. To assess this, we employed a simple discrimination task, novelty trials and inference by exclusion tasks using a two-choice touchscreen paradigm. The influences of sex, breed, neuter status, age, and previous experience (in a similar study) on the testing outcome were analysed. Each one of the 43 participating dogs was trained on their own stimulus pair (S+ and S-) and confronted with two criteria (discrimination ability and novelty attraction) before continuing to test for inference by exclusion ability. Results showed that only age and experience influenced the dogs´ performance in the sense that younger dogs were quicker in learning the simple discrimination and made fewer errors. Dogs with more experience learned to avoid a distracting, negative novel stimulus more quickly than non-experienced ones and made fewer errors in both criteria. Surprisingly, only 30% of the dogs exhibited inference by exclusion above chance. Furthermore, there was no influence of any predictor (age, breed, experience, neuter status, sex, performance in discrimination trials and performance in novelty trials) found on the animals´ inference by exclusion performance.

Keywords: dogs, learning, touchscreen, inference, age, experience, discrimination

Introduction

Animals are confronted with various challenging situations throughout their lives. Learning and memory are important factors for acquiring and remembering information (Wallis et al., 2016). Having certain knowledge about your surroundings is very helpful in making decisions, however information can sometimes be incomplete. When animals are confronted with various alternatives, different strategies can be applied to solve problems. Besides choosing randomly, previous information acquired through trial and error learning can be the base a decision is made up upon, yet a more complex strategy can also be used, for example by inferring via exclusion (O'Hara et al., 2015). One way to test an animal's learning ability is discrimination learning, which often includes a two-choice task, where two (often visual) stimuli get presented at the same time, but only one is rewarded (mostly with food) and the other one is not. When confronted with such a two-choice discrimination task, each stimulus must be thought about individually and the subject should, through a learning process, infer which "label" ("rewarded" or "non-rewarded") to attach to which stimulus. The subject has to decide which stimulus to choose and which one to inhibit choosing, thus parallel processing of the stimuli is needed (Wallis et al., 2016), because they are presented together at the same time. This parallel processing requires working memory, a form of short-term memory that enables an animal to create, access, simultaneously manipulate and maintain mental representations of stimuli or objects (Troche et al., 2014). Working memory is assisted by the prefrontal cortex. During aging, specific working memory deficits have been linked to a lower NMDAR (*N*-methyl-D-aspartate receptor) function in the prefrontal cortex (due to oxidative stress) (Foster, 2012; Guidi et al., 2015; McQuail et al., 2016). Keeping that in mind, it is not surprising that older individuals exhibit lower working memory capacity and therefore perform worse in tasks, which rely on this. Furthermore, one study showed that training not only improves the ability to recognize a stimulus, but also increases the capacity to ignore distracting information and attend towards a target stimulus (Zhao et al., 2013). Without paying attention to stimuli or items it is not possible to learn about or memorize them.

Inference by exclusion, which is defined as "selecting the correct alternative by logically excluding all other potential alternatives" (Call, 2006) can be used as a strategy by animals when they have to decide something on the basis of already given but incomplete information. It is a process that can be exhibited on a linguistically represented proposition or implication, however language is not a necessary prerequisite for problem solving, reasoning and inference (Vigo & Allen, 2009). This means that non-human animals can solve problems on a more advanced level that requires higher cognitive flexibility and learning abilities. Previous studies have shown that great apes (chimpanzees, gorillas, orang-utans and bonobos) (Call, 2006), ravens (Schloegl et al., 2009), humans (children and adults) and some dogs (Aust et al., 2008), grey parrots (Mikolasch et al., 2011; Pepperberg et al., 2013) and Goffin cockatoos (O'Hara et al., 2015) show inference by exclusion abilities.

In a study by Call et al. in 2006, orang-utans, chimpanzees, gorillas and bonobos were tested on their inference by exclusion abilities via a two-choice paradigm. The animals were

able to witness, through a plexiglass partition, the placement of two treats on a platform (a piece of banana and a grape). After that, two opaque bins were put over the rewards preventing the animals from seeing what was underneath. The experimenter then reached inside one of the bins and extracted the reward in their fist, placing it in the middle of the platform behind another opaque barrier without the animal being able to see what the treat was. Then, the experimenter acted as if they did the same for the second hidden reward, but instead left the reward in its place and only moved their empty fist to the opaque barrier. The subject was not able to see which food item really was removed until the experimenter lifted the opaque barrier in the middle, therefore making the food visible to the animal. This visible food item then got discarded in front of the animal for them to realize this piece of food is not accessible for them anymore. The platform was then moved towards the test subject so it could choose one of the bins (and get the treat if there was one underneath it). Only one food piece was still underneath a bin and the animal had to choose the one which still contained a reward by logically excluding the other one. Subjects as a group did choose the correct container above chance. However, the inference performance worsened over the test trials and there were no species or age differences found. This experiment showed that two-choice paradigms are suitable for testing inference by exclusion ability in non-human primates, that it is present in apes and can be tested for on a non-linguistic level.

Inference on what a novel object should be labelled as ("rewarded" or "non-rewarded") can be drawn by applying a label to the unknown object on the base of knowledge already internalised. This is called fast mapping and although it is commonly used in relation to how children learn words (Coutanche & Thompson-Schill, 2014; Spiegel & Halberda, 2011; Horst & Samuelson, 2008; Heibeck & Markman, 1987), it can also be applied to other species (Kaminski et al., 2004; Bloom & Markson, 1997). Rejecting a known, distracting object or stimulus is the underlying incentive for mapping novel labels to new objects or stimuli, rather than mapping novel labels to novel things (map-novelty-to-novelty) (Halberda, 2006). In the study of Kaminski et al. (2004) the authors showed that a dog demonstrated the ability to fast map and used this in order to infer by exclusion. A nine-year-old Border Collie, Rico, was able to successfully match a novel word to a novel object. The animal was familiar with the labels of over 200 items (mostly children´s toys). In an experiment, the dog was asked to fetch a specific toy from a selection of his other toys, and Rico fetched 37/40 items correctly (he brought the item that was asked for). After that, the Border Collie was presented with novel items which he had never seen before. The test sessions included known items and one novel item that was presented at the same time and the dog was asked to fetch either a known item or the novel item. When confronted with a new label, Rico correctly retrieved the novel item in 7/10 sessions. This study showed that dogs in general might have the ability to make decisions based on exclusion learning. However, one critique of this study was that Rico could have chosen the novel stimuli because it was more salient or appealing to him (Markman & Abelev, 2004). Pet dogs have been shown to exhibit neophilia in a preference paradigm (Kaulfuß & Mills, 2008) and to tackle that, Pilley and Reid (2011) carried out a study with a similar procedure, but with more controls and test conditions. Their test subject was again a Border Collie, named

Chaser, who was 3 years old at the time of the study. When tested for inference by exclusion ability, Chaser was able to successfully retrieve a novel object (which was asked for using a new word) when it was presented together with seven familiar ones. Chaser was asked to first retrieve known objects, and on the 3rd run in each trial she was asked to retrieve a newly named novel object she had never seen before. Chaser succeeded in retrieving the novel item correctly in 8/8 test trials. Two years after the initial testing the authors conducted a re-test with the exact same setting, but a new group of unfamiliar items. Again, the dog retrieved the novel object correctly in 8/8 trials. This showed that Chaser was able to use inferential reasoning by exclusion to map words to objects or toys.

In the study of Aust et al. (2008) the ability to infer by exclusion was compared in three different species, namely pigeons (n=6), pet dogs (n=6), adult humans (n=6) and children (n=8). Utilizing the touchscreen paradigm, the subjects were all presented with a simple two-choice discrimination task. The individuals had to learn to discriminate between four positive and four negative stimuli. After reaching certain criteria during the training, the testing phase was initiated and in test 1 a positive stimulus was replaced by a novel positive stimulus the subjects had not seen before. This novel stimulus was presented with one of the familiar negative stimuli. In test 2, new test trials were presented to the subjects consisting of the new positive stimulus from test 1 and a new negative stimulus the subjects have never seen before, and their ability to infer by exclusion (by choosing the new positive stimulus from test 1 and rejecting the novel negative stimulus) was monitored. Results showed that in the more demanding test 2 none of the pigeons, half of the dogs, three out of five adults and six out of eight children showed inferential reasoning by exclusion. The authors concluded that pigeons do not rely on logic-based strategies like inferential reasoning, whereas almost all humans and half of the dogs did. However, one must keep in mind that, unlike training trials, test trials were not rewarded (for all species), because the authors did not want the choices of the subjects to be influenced by reinforcement (food reward) and one trial learning, which could have affected the animal´s choice due to missing feedback. The authors found that male dogs were slower learners than females and that dogs that needed more training sessions performed better in the inference trials.

In another touchscreen study based on Aust et al. (2008), Wallis et al. (2016) trained 95 Border Collies in the same discrimination task and after reaching certain criteria, with inference by exclusion trials. Younger dogs were faster in reaching the discrimination criteria and needed less correction trials, which were implemented after a wrong choice, where the same stimuli pairing appeared again and the dog had another chance to choose the correct stimulus. In general, dogs that needed more correction trials showed inference by exclusion more often than those individuals that needed less correction trials. Furthermore, when it comes to inference by exclusion performance per se, older dogs exhibited this strategy more often than younger dogs. Wallis et al. presented the dogs with a reduced reward ratio (in the training and in the test), meaning that not all trials were rewarded, to prepare the dogs for unrewarded trials and to prevent them from possibly switching strategies in the test due to lack of feedback. In young dogs this might have had a big impact, because they show a greater flexibility and therefore could change strategies more often than older individuals

due to missing feedback. In general, less than 10% of the tested dogs showed choice behaviour based on inference by exclusion.

In 2015, O'Hara et al. conducted a study in which Goffin cockatoos were tested on their inference by exclusion ability, also using a two-choice touchscreen paradigm. Coloured clip art pictures on a white background were used as stimuli. The animals were presented with diverse stimuli pairings during training and test sessions. Results showed that 8/12 Goffins exhibited an inference by exclusion response pattern significantly above chance. However, the authors could not find a significant influence of sex, prior touchscreen experience, discrimination-learning performance during the training (the number of sessions needed to learn which stimulus was positive and which one was negative) or the avoidance of novel negative stimuli in the training (novelty trials). In the current study, we decided to test dogs' inference by exclusion ability using the touchscreen apparatus and the same protocol used in O´Hara et al. to make the studies comparable, but with some minor changes. For example, dogs were given more test sessions than the birds (for more details please refer to the "Material and Methods" section and to O´Hara et al., 2015 for more details about the Goffin cockatoos).

In the current study, a touchscreen apparatus was used to inhibit social cuing from the owners or the experimenter and to allow the assessment of individual learning abilities (Range et al., 2008). This automated setup has been shown to be well suited for cognitive testing in various animal species (Aust et al., 2008; Bussey et al., 2008; Bussey et al., 2012; O'Hara et al., 2015; Spinelli et al., 2004). Pet dogs were chosen as subjects, because they offer a higher variability and complexity than laboratory animals which are often from the same breed and experience the same living conditions and lifestyles. Pet dogs were an ideal choice of test subjects, because inference by exclusion abilities in dogs have already been shown in previous studies (Aust et al., 2008; Kaminski, 2004; Pilley & Reid, 2011; Wallis et al., 2016) and it is relatively easy to obtain a high number of testing subjects. Previous studies on this topics in pet dogs had either a very small sample size (Aust et al., 2008) or were conducted with only one type of dog breed (Wallis et al., 2016).

Here, 43 dogs of different breeds and ages were tested in a two-choice discrimination task. The aim of this study was to determine the influence of age, breed, previous experience, neuter status, and sex on discrimination learning, novelty avoidance and inference by exclusion ability. Similar to Wallis et al. (2016) we expected to find age differences, more precisely, younger dogs should be quicker in learning the basic discrimination task and make fewer errors, due to higher working memory capacity in younger individuals (Milgram et al., 2002). We anticipated that some dogs would exhibit inference by exclusion. On the contrary to the results from Wallis et al., we did not expect older dogs to perform better in the inference by exclusion trials, but younger dogs to do so, since it was found that younger dogs show better learning abilities, are cognitively more flexible and that aging adversely affects cognitive processes, such as memory and attentiveness (Chapagain et al., 2017; Guidi et al., 2015; Wallis et al., 2014). In the former study, inference by exclusion trials were not rewarded, which could have led, especially in younger dogs, to them switching strategies often and therefore perform worse in these

trials. Here, all correct choices were rewarded, hence leading us to predict that younger dogs would maintain their problem-solving strategy, learn the stimulus associations more quickly due to feedback and consequently also perform better in inference by exclusion trials. Additionally, we included various dog breeds looked for possible breed differences, since such differences have been found in previous studies (Horschler et al., 2019; McGetrick & Range, 2018; Serpell & Hsu, 2005) and also examined if the animal´s neuter status would influence their performance, considering hormones are known to affect cognitive processes (Little, 2013). Aust et al. (2008) found a sex effect (male dogs learned slower) and an effect of the amount of training (dogs that exhibited inferential reasoning needed more training), that we also expected to find in the current study. Some dogs had already participated in a previous, similar study (Wallis et al., 2016) and we therefore expected them to be faster in understanding the tasks and reaching criteria, because they were already familiar with the procedures.

Material and Methods

Ethical approval and owner consent

This study was discussed and approved by the institutional ethics and animal welfare committee at the University of Veterinary Medicine Vienna in accordance with Good Scientific Practice guidelines and national legislation (http://www.vetmeduni.ac.at/fileadmin/v/z/forschung/GoodScientificPractice_English.pdf). All subjects that participated in the study were family pets, and reward-based training was utilized in all tests conducted, with no potentially harmful experimental manipulations. The dog owners signed a document prior to testing, stating that their participation in this study was completely voluntarily, gave their general consent to participate and agreed that the acquired data and information could be used for scientific purposes.

Subjects and variables

For each subject, the following variables were collected: age (dogs aged 6 years and under were put in age group 1, dogs over the age of 6 years were put in age group 2), breed (Border Collies and non-Border Collies), previous experience (yes or no), neuter status (intact or not) and sex (F,female; M,male). Overall, 43 dogs (22F, 21M) participated, of which 21 were intact (8 F, 13 M) and 22 were neutered (13 F, 8 M). The age of the animals ranged from 7 months to 12 years. There were 22 individuals in age group 1 and 21 individuals in age group 2, 18 dogs were Border Collies and 23 other breeds (including mixed breeds and mongrels) (Table 1). Of all the participating dogs, 18 had previous touchscreen experience with an inference by exclusion test and 25 did not (of which nine individuals did not have any experience with a touchscreen at all but received extra pre-training before participating in the study). The dogs were recruited using an internal database of the Clever Dog Lab (Vienna, Austria) and contacting the dog owners directly.

Table 1. Name, sex (F= female, M= male), age group, experience status, breed and neuter status of all dogs that participated in the study (N=43).

Name	Sex	Age group	Experienced	Breed	Neutered
Amy	F	1	Yes	Border Collie	No
Bambi	F	1	No	Mix	Yes
Cheynna	F	1	No	Australian Shepherd	No
Cliff	M	1	No	Border Collie	No
Clio	F	1	No	Mix	Yes
Dexter	M	1	Yes	Border Collie	No
Igor	M	1	No	Border Collie	No
Jamie	M	1	Yes	Border Collie	No
Kieki	F	1	No	Maltese Terrier	Yes
Kilio	M	1	Yes	Mix	Yes
Lenny	M	1	No	Podenco	Yes
Linus	M	1	No	Australian Shepherd	No
Lola	F	1	No	Mix	No
Luna	F	1	No	Husky	No
Merlin	M	1	Yes	Border Collie	Yes
Muffin	F	1	No	Mix	Yes
Noah	M	1	No	Border Collie	No
Ronja	F	1	No	Mix	Yes
Suri	F	1	Yes	Border Collie	Yes
Talie	M	1	Yes	Husky	No
Tuuka	F	1	No	Mix	Yes
Ultimo	M	1	Yes	Border Collie	No
Chilli	F	2	No	Australian Shepherd	Yes
Cleo	F	2	Yes	Border Collie	No
Clooney	M	2	Yes	Border Collie	No
Fenja	F	2	Yes	Border Collie	No
Flamme	M	2	Yes	Pyranean Shepherd	No
Ginger	F	2	No	Mix	Yes
Gismo	M	2	Yes	Border Collie	Yes
Guinness	F	2	Yes	Border Collie	No
Habibi	F	2	No	Mix	Yes
Hybie	F	2	No	Mix	Yes
Lenny	M	2	Yes	Border Collie	No
Luke	M	2	Yes	Border Collie	Yes
Maggie	F	2	Yes	Border Collie	Yes
Mago	M	2	No	Golden Retriever	No
Miley	F	2	No	Border Collie	No
Patrasch	M	2	No	Mix	Yes
Scully	F	2	No	Rottweiler- Mix	Yes
Sokrates	M	2	No	Podenco- Mix	Yes
Taiko	M	2	No	Akita Inu	Yes
Tika	F	2	No	Husky- Mix	Yes
Todi	M	2	Yes	Mix	No

Touchscreen

A touchscreen apparatus was used for this study. The dogs were trained to touch the screen with their nose to trigger a response. The apparatus consisted of a large rectangular box, which contained a height adjustable computer screen and a food dispenser (48 x 100 x 60 cm [w x h x d]) which was connected to a computer. Pieces of commercial dry dog food were used as a reward for correct choices. The food was, after a correct choice, dispensed

automatically via a metal tube underneath the screen ("reward hole"). The computer screen (15" TFT 600 × 800 pixel resolution) was covered with an infrared touchframe (Carroll Touch, Round Rock, TX, USA; 32 vertical× 42 horizontal resolution) and whenever this infrared field was disrupted (by the dog´s nose) this elicited a reaction of the computer (including audio feedback and potential food output). To work on the computer screen, the dog had to enter a testing niche. From this niche, view protectors were located on the sides and on the top to prevent any social cues (from the owner or experimenter) influencing the dog´s choice (Fig. 1). The computer screen (and touch frame) could be moved up and down and therefore adjusted to the height of the dog.
The study was conducted at the Clever Dog Lab in Vienna, Austria in a separate testing room.

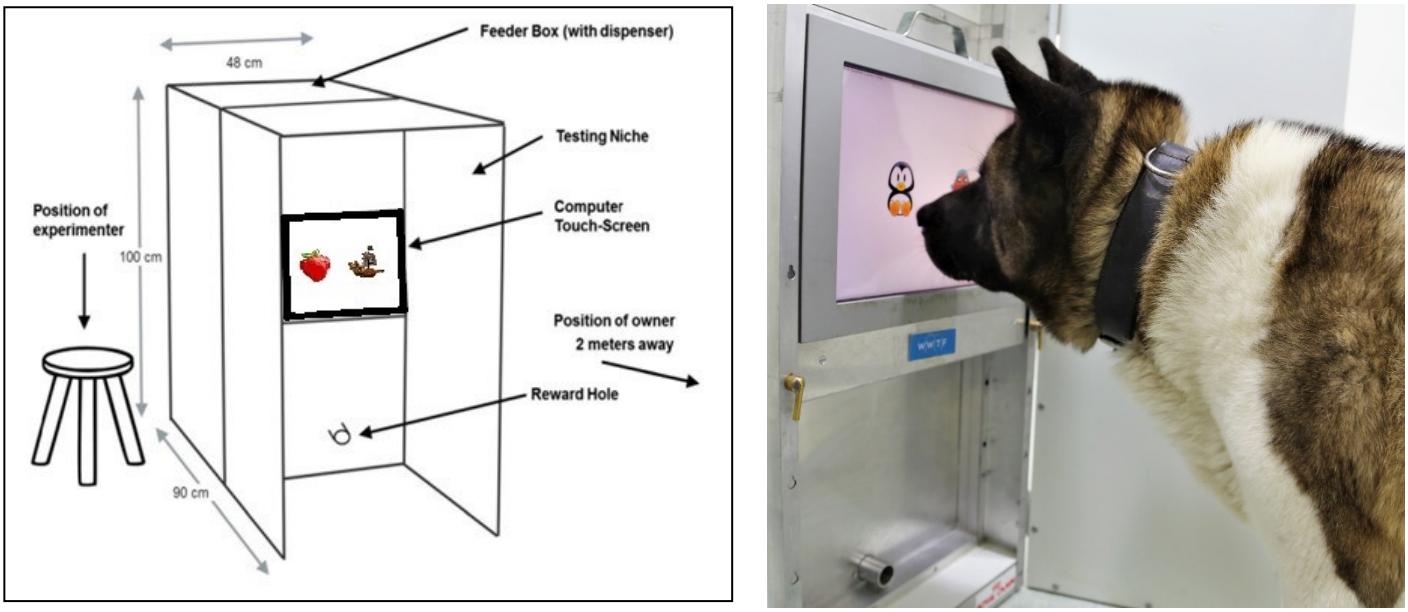

Fig. 1. Left: Schematic illustration of the touchscreen apparatus used in this study. Illustration taken from Wallis et al. (2016).
Right: A subject dog working on the touchscreen. The animal is about to touch one stimulus to elicit a response from the computer.

Stimuli and procedure

Dogs and owners visited the lab once to three times a week depending on the availability of the owner. All dogs started with the "training" and after its completion they continued with the "test" condition. The animals participated in about four sessions each visit (depending on the dog's motivation and concentration that day), each session consisting of 20 trials. In general, each lab visit lasted about 30mins per dog. Throughout the whole procedure the owner could be present in the testing room, sitting silently on a chair about 1.5m from the apparatus and the dog. The experimenter sat right next to the apparatus to make sure everything was working properly (e.g., that food was dispensed correctly via the reward hole). Furthermore, sometimes it was necessary for the experimenter to hold some dogs

back from the screen, because they were standing too close to it and would trigger a response accidentally (because their nose or other body part was already touching the screen and the computer would count it as a touch). The experimenter held the dog back without looking at the screen or giving any comments. Apart from those occasions, the dogs worked independently on the touchscreen (Fig. 1).

For this study, we used the same stimuli that were used in O´Hara et al. (2015). Consisting of 190 different freeware clip art stimuli from the "Open Clip Art Library" (http://www.openclipart.org/). All stimuli were differently coloured and had different shapes (examples Fig. 2). Each dog was randomly assigned one positive (rewarded) and one negative (non-rewarded) stimulus (S+ and S-). These stimuli represented the dog´s baseline training set. Novel stimuli (novS) that were neither the dog´s positive nor negative stimulus were interspersed in the "training" and "test" trials. Each session contained at least two new novel stimuli. In the training the novel stimuli were always negative (novS-), however in the test sessions there were also novel positive stimuli (novS+) additionally to novel negative ones (for more exact information on the procedure see "training" and "test" below). A forced choice paradigm was used, where two stimuli were presented on the touchscreen at the same time (S+ and S-). The appearance of the positive and negative stimulus on either the right or the left side of the screen was randomized by the computer program (Cognition Lab, version 1.9). Whenever the dog touched the correct stimulus a short high pitched sound was elicited, the stimuli disappeared (screen turned completely white), a food reward was dispensed and after a few seconds the next trial was shown on the screen. Whenever the animal touched the negative stimulus, a lower pitched sound was elicited and a time out followed, characterized by the whole screen turning red (with no stimuli shown) for three seconds. Afterwards, a correction trial was initiated (the same pairing of stimuli appeared again) until the correct choice was made. These correction trials were present throughout all training trials and during the test only for the baseline trials, but were absent in the four test trials, in which the stimuli disappeared without any (auditory or visual) feedback when the dog made the wrong choice, and the next pairing was presented. We chose a 100% reward ratio, also in the test trials, meaning that each correct choice was rewarded.

Training

The main goal of the training was for each dog to learn to distinguish their S+ from their S- and to learn to refrain from touching novel stimuli when paired with their known positive stimulus. Each session consisted of 20 trials, 18 of which comprised of the dog´s S+ and S- (baseline) and two trials were novelty trials, consisting of the dog´s known S+ paired with a novel negative stimulus (novS-) in each of the two novelty trials. Novelty trials occurred in each session on either the 4^{th} or 5^{th} and 14^{th} or 15^{th} trial, resulting in two novel stimuli per session (Fig. 2). Training allowed correction trials, meaning that after making a wrong choice (by either touching S- or novS-) the same stimuli pairing appeared again after the time out and the dog had another chance to choose, until a correct choice was made. During the training the subject had to reach two different Criteria before moving to the testing phase.

Criterion 1 (Crit1) was reached when the dog made 16/18 correct choices when confronted with its baseline S+ and S- in two consecutive sessions, indicating that the subject can distinguish between its positive and negative stimulus. For criterion 2 (Crit2), only the novelty trials were of interest. The dogs had to inhibit touching the novS- (and instead touch their S+) in both the session´s novelty trials for two consecutive sessions. Crit1 and Crit2 could be reached independently from each other, meaning it was not necessary, but possible, for the animals to reach both criteria in the same session. Training trials were not limited in number, resulting in the number of sessions varying between subjects, testing was initiated as soon as the dog had reached both criteria. Furthermore, the number of errors was assessed, with one error being one correction trial, (one preservative response towards the negative stimulus). Out of simplicity, henceforth the number of correction trials that occurred in the dogs´ baseline training (Crit1) were labelled "Errors1" and the number of correction trials that occurred during the dogs´ novelty trials (Crit2) were called "Errors2".

Testing

Test sessions were limited to 30, and after a dog finished these sessions (each consisting of 20 trials), the testing for the dog (and their participation in the study) was completed. Test trials differed from the training in the sense that in this phase the baseline appeared in only 16/20 trials, the other four trials were test trials. The test trials were semi-randomly interspersed on the 4^{th} or 5^{th}, 8^{th} or 9^{th}, 12^{th} or 13^{th} and 16^{th} or 17^{th} trial. During the first test trial per session the dog´s S- was presented with a novel positive stimulus novS+, during the second test trial the novS+ (from the previous test trial) was presented with another novel negative stimulus, novS-2. The same setting was conducted for the dog´s S+ in the following test trials, such that in the third test trial the dog´s initial S+ was presented together with another, also completely novel stimulus novS- and in the fourth test trial this novS- was then presented together with another completely novel (positive) stimulus, novS+2 (Fig. 2). Keep in mind that S+ and S- were always the dog´s known baseline stimuli, whereas novS+, novS-, novS+2 and novS-2 changed each session, meaning they were always represented by new stimuli the dog was unfamiliar with and were not presented in any other session again. So, in each test session the animals were presented with four completely novel stimuli and had to figure out whether those stimuli were positive or negative by logically inferring by exclusion (Fig. 2). For a test session to be considered successful, the dog had to choose the correct stimulus in all these four test trials. The baseline trials in these sessions were not of interest for this analysis, however, were present in these sessions for the animal to keep their working routine, and to have easy trials in between the more difficult test trials to gain food and keep motivation to work high.

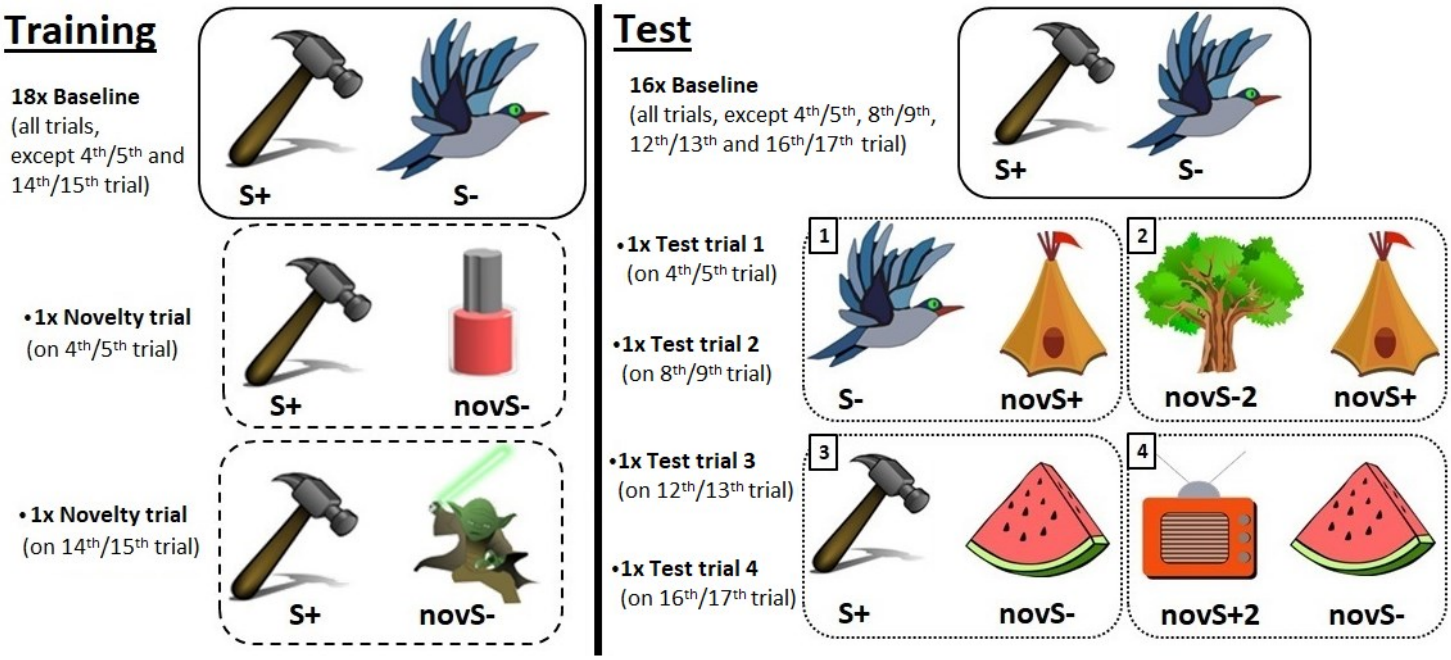

Fig. 2. Example of a training and test setup. Only S+ and S- were the same throughout the whole study (each subject had their own combination of stimuli), all other novel stimuli were replaced by new ones in the next session.

Analysis

Statistical analysis was conducted using the software R (R version 4.1.0 (2021-05-18), R: A Language and Environment for Statistical Computing). R packages used included: “stats” for basic statistical analysis and preparation, “mgcv” for creating all the models, “ggplot2” for the graphics and “stargazer” for generating a LaTeX output for the tables. Tables and graphs were all created using R (see above) and Overleaf (Online LaTeX editor, https://www.overleaf.com).

Generalized additive models (GAM) with negative binomial distribution (due to the highly dispersed data) and a log link function were used for all the analyses. Prior to the final GAM analysis, the most influential points per model were assessed using the Cook´s distance. Using the traditional cut-off value of 4/n (where n indicates the number of tested individuals) (Belsley et al., 1980:28), the influencing data points exceeding this value, were removed from the analysis as they heavily influenced and altered the statistical outcome. The number of observations (N) for each analysis is stated in detail below in the “Results” section and in the tables 1-3. For the reduction of full models, an initial $\alpha = .1$ was chosen (for reduction purposes), whereas in the final models, the more restrictive $\alpha = .05$ was chosen as the final boundary for further discussion. The fit of each model was computed using Restricted Maximum Likelihood (REML) as a parameter estimate in all models.

Results

Out of the 43 test subjects, 19 individuals reached criterion 1 before criterion 2 (n= 14 from age group 1 and n= 5 from age group 2) and 18 individuals learned criterion 2 before criterion 1 (n= 5 from age group 1 and n= 13 from age group 2), 6 individuals spontaneously reached both criteria in the same session (n= 3 from age group 1 and n= 3 from age group 2).

Training phase

Crit1

The overall number of sessions needed to reach Crit1 ranged from 4 (the minimum number of sessions required for a dog to reach criteria) to 74 sessions. After excluding five overinfluential points (individuals), GAMs were calculated using n= 38 observations. For the full model, the number of sessions needed to reach Crit1 (which varied for each dog) was used as the response variable and age, breed, experience, neuter status, and sex were set as the explanatory variables. Only age, experience and neuter status were kept as they were significant at least on the α= .1 level, whereas all other predictors were removed from the final model (Crit1 and Crit1 (reduced), Table 2). In the final model, however, only age remained significant at the α= .05 level. This final model showed that older dogs (age group 2) needed significantly more sessions to reach Crit1 (the simple discrimination learning) than younger dogs (Fig. 3).

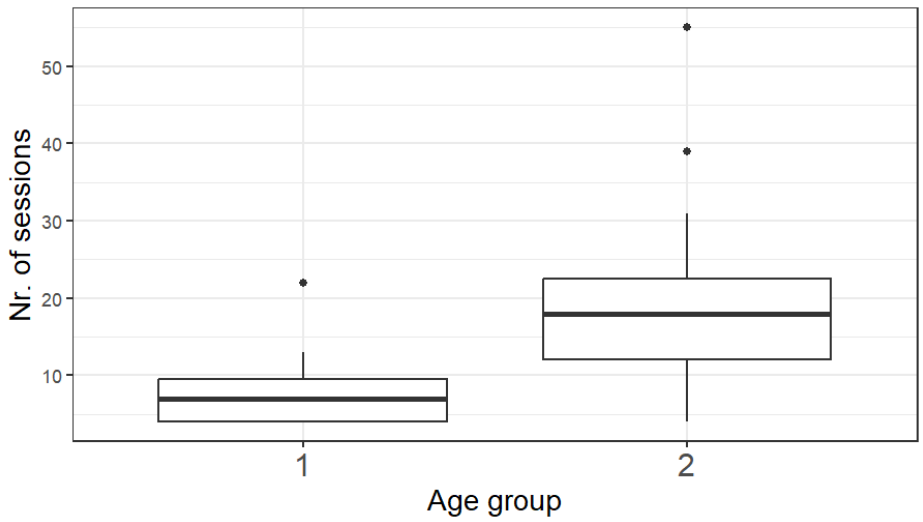

Fig. 3. Number of sessions needed to reach Crit1, grouped by age group. Error bars with a 95% confidence interval are shown, influential points are not depicted (n= 38).

Crit2

The overall number of sessions needed to reach Crit2 ranged from 4 (the minimum number of sessions required for a dog to reach criteria) to 41 sessions. After excluding four overinfluential points (individuals), GAMs were generated using n= 39 observations. For the full model the number of sessions needed to reach Crit2 (which varied for each dog) was used as the response variable and age, breed, experience, neuter status, and sex were set as the explanatory variables. The only significant variable at the α= .01 level was experience and therefore this remained in the model, whereas all other predictors were removed from the final model (Crit2 and Crit2 (reduced), Table 2). This result remained significant ($p < 0.05$) in the final model and therefore showed, that dogs with previous experience needed significantly fewer sessions to reach Crit2 (Fig. 4), than dogs without prior experience.

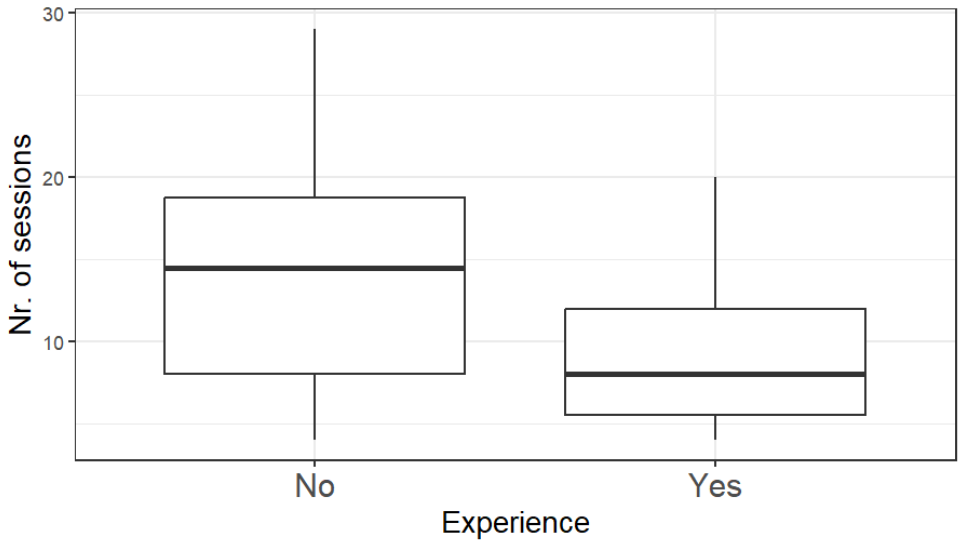

Fig. 4. Number of sessions needed to reach Crit2, grouped by experience. Error bars with a 95% confidence interval are shown, influential points are not depicted (n= 39).

Table 2. GAMs (full and reduced) of nr. of sessions needed to reach Crit1 and Crit2 (all read from top to bottom). Estimates, 95% CI (in brackets), t- scores and p- values are given for each variable. The explanatory variables stated on the left are age group, breed, experience, neuter status (intact) and sex. Age group 2= 6 years and older, F= female, Y= yes, BC= Border Collie. High significance ($p < 0.05$) is marked bold.

	Dependent variable: Nr. of sess. to Crit1 and Crit2			
	Crit1	Crit1 (reduced)	Crit2	Crit2 (reduced)
age gr. 2	**0.85 (0.48, 1.22)** **t = 4.48, p < 0.05****	**0.88 (0.51, 1.25)** **t = 4.66, p < 0.05****	0.21 (−0.13, 0.55) t = 1.19, p = 0.24	
breed BC	0.32 (−0.21, 0.86) t = 1.19, p = 0.24		0.09 (−0.34, 0.53) t = 0.43, p = 0.67	
exp Y	−0.44 (−0.94, 0.05) t = −1.75, p = 0.09*	−0.32 (−0.73, 0.08) t = −1.57, p = 0.12	**−0.60 (−1.04, −0.17)** **t = −2.72, p < 0.05****	**−0.43 (−0.76, −0.09)** **t = −2.48, p < 0.05****
inta Y	−0.42 (−0.86, 0.03) t = −1.85, p = 0.07*	−0.32 (−0.72, 0.08) t = −1.59, p = 0.12	0.23 (−0.14, 0.61) t = 1.22, p = 0.23	
sex F	0.25 (−0.12, 0.62) t = 1.30, p = 0.20		0.02 (−0.34, 0.38) t = 0.11, p = 0.92	
AIC	251.6	250.3	256.5	251.7
Observations (N)	38	38	39	39
Adjusted R^2	0.37	0.40	0.06	0.11
Log Likelihood	−124.79	−124.16	−127.23	−124.84
REML	123.55	123.83	126.59	125.26

*p<0.1; **p<0.05

Errors1

The overall number of Errors made during baseline trials to Crit1 ranged from 1 to 389. After excluding two overinfluential points, GAMs were generated using n= 41 observations. For the full model, the overall number of errors made in the simple discrimination task (Crit1) was used as response variable and age, breed, experience, neuter status, and sex were set as the explanatory variables. The only significant variables were age and experience, both on the α= .01 level and thus remained in the model, whereas all other predictors were removed from the final model (Errors1 and Errors1 (reduced), Table 3). These remained significant ($p < 0.05$) in the final model and therefore showed, that older dogs made more errors than younger dogs and additionally individuals with previous experience were less inclined to make errors during discrimination trials (Fig. 5).

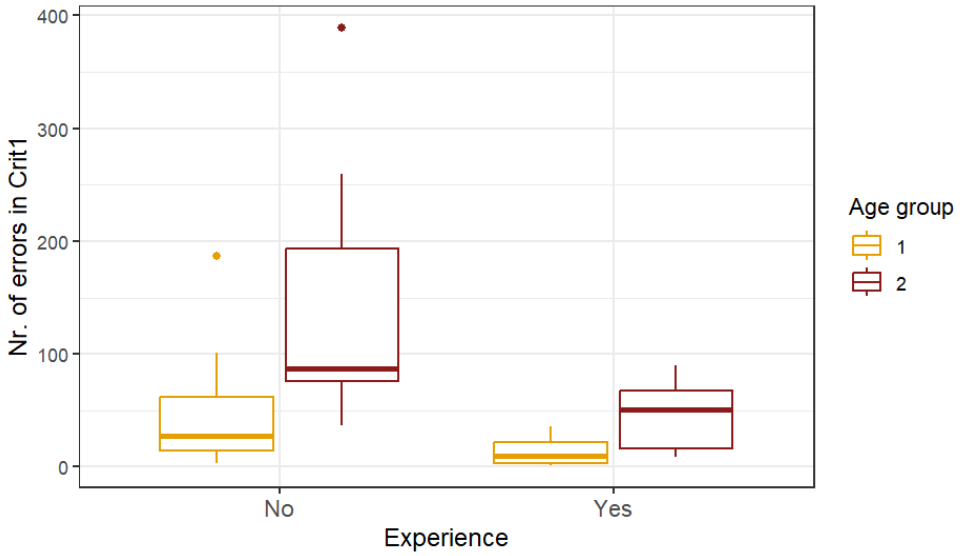

Fig. 5. Number of errors made during baseline trials in Crit1, grouped by experience and age group. Error bars with a 95% CI are shown, influential points are not depicted (n= 41).

Errors2

The overall number of Errors made during novelty trials in Crit2 ranged from 0 to 63. After excluding two overinfluential points, GAMs were generated using n= 41 observations. For the full model, the overall number of errors made in the novelty avoidance task (Crit2) was used as the response variable and age, breed, experience, neuter status, and sex were set as the explanatory variables. The only significant variables on the α= .1 level were experience and age and thus remained in the model, whereas all other predictors were removed from the final model (Errors2 and Errors2 (reduced), Table 3). In the final model, however, only experience remained significant on a α= .05 level. This showed that experienced dogs made significantly less errors in the novelty avoidance task (Crit2), than inexperienced dogs (Fig. 6).

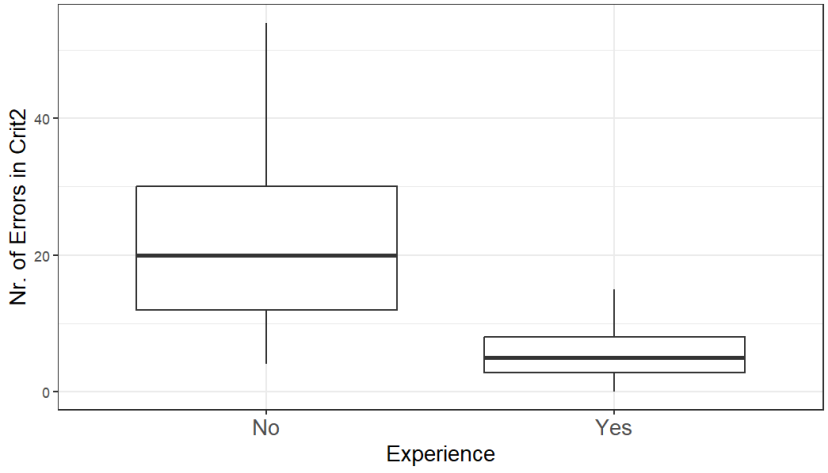

Fig. 6. Number of errors made during novelty trials in Crit2, grouped by experience. Error bars with a 95% CI are shown, influential points are not depicted (n= 41).

Table 3. GAMs (full and reduced) of nr. of errors made within Crit1 and Crit2 (all read from top to bottom). Estimates, 95% CI (in brackets), t- scores and p- values are given for each variable. The explanatory variables stated on the left are age group, breed, experience, neuter status (intact) and sex. Age group 2= 6 years and older, F= female, Y= yes, BC= Border Collie. High significance ($p < 0.05$) is marked bold.

	Dependent variable: Nr. of errors made within Crit1 and Crit2			
	Errors1	Errors1 (reduced)	Errors2	Errors2 (reduced)
age gr. 2	**1.25 (0.69, 1.81)** **t = 4.41, p < 0.05****	**1.16 (0.63, 1.70)** **t = 4.24, p<0.05****	0.42 (−0.02, 0.87) t = 1.85, p = 0.07*	0.36 (−0.06, 0.79) t = 1.67,p = 0.10*
breed BC	0.11 (−0.66, 0.88) t = 0.27, p = 0.79		0.20 (−0.43, 0.83) t = 0.63, p = 0.53	
exp Y	**−1.23 (−1.98, −0.48)** **t = −3.21, p < 0.05****	**−1.17 (−1.73, −0.62)** **t = −4.14, p < 0.05****	**−1.66 (−2.27, −1.05)** **t = −5.34, p < 0.05****	**−1.42 (−1.88, −0.97)** **t = −6.14, p < 0.05****
inta Y	0.24 (−0.37, 0.86) t = 0.78, p = 0.44		0.11 (−0.40, 0.61) t = 0.41, p = 0.68	
sex F	0.33 (−0.25, 0.90) t = 1.12, p = 0.27		−0.18 (−0.64, 0.28) t = −0.77, p = 0.45	
AIC	414.6	410.3	295.1	290.6
Observations (N)	41	41	41	41
Adjusted R^2	0.23	0.34	0.28	0.36
Log Likelihood	−206.29	−204.13	−146.54	−144.29
REML	202.47	202.93	144.04	143.75

*p<0.1; **p<0.05

Testing phase

Inference by exclusion ability

Dogs did not choose by inference by exclusion on a group level, but on the individual level 13/43 (30%) exhibited inference by exclusion above chance, of which nine dogs were from age group 1 and four dogs from age group 2. From these dogs, five had previous experience and eight did not.

For the full GAM the total number of sessions where the dog chose by inference by exclusion (4/4 correct test trials), was used as the response variable and Crit1, Crit2, Errors1, Errors2, age, breed, experience, neuter status, and sex were set as the explanatory variables. Results showed, that only Crit2 and Errors2 were significant on the α= .1 level, whereas all other predictors were not and therefore removed from the final model (Table 4). In the final model, however, Errors2 did not remain significant and Crit2 did not reach the more restrictive level of α= .05. Therefore, in this analysis, there was no strong influence of the tested variables found on the dogs´ inference by exclusion ability.

Table 4. GAMs (full and reduced) of nr. of inference by exclusion trials (all read from top to bottom). Estimates, 95% CI (in brackets), t- scores and p- values are given for each variable. The explanatory variables stated on the left are age group, breed, experience, neuter status (intact) and sex. Age group 2: 6 years and older, F= female, Y= yes, BC= Border Collie.

	Dependent variable: Inference by exclusion trials	
	Inf. by excl.	Inf. by excl.(reduced)
age gr. 2	−0.19 (−0.58, 0.21) t = −0.92, p = 0.36	
breed BC	−0.12 (−0.56, 0.31) t = −0.56, p = 0.58	
exp Y	−0.32 (−0.79, 0.15) t = −1.33, p = 0.19	
inta Y	0.05 (−0.33, 0.44) t = 0.27, p = 0.79	
sex F	−0.09 (−0.46, 0.28) t = −0.48, p = 0.64	
Sess. to Crit1	−0.02 (−0.06, 0.03) t = −0.82, p = 0.42	
Errors within Crit1	0.003 (−0.005, 0.01) t = 0.71, p = 0.48	
Sess. to Crit2	0.02 (−0.003, 0.05) t = 1.74, p = 0.09*	0.02 (−0.003, 0.04) t = 1.73, p = 0.09*
Errors within Crit2	−0.02 (−0.04, 0.001) t = −1.90, p = 0.06*	−0.01 (−0.02, 0.003) t = −1.60, p = 0.12
AIC	157.5	150.2
Observations (N)	38	38
Adjusted R^2	0.24	0.09
Log Likelihood	−77.73	−74.09
REML	89.21	80.42

*p<0.1

Discussion

The aim of this study was to assess possible influences of age, breed, previous experience, neuter status, and sex on discrimination learning and assigning labels to novel objects via inference by exclusion in pet dogs, using a two-choice paradigm on a touchscreen setup. We did not find any effects of sex, breed or neuter status, so in this study, none of these factors seem to impact working memory or inference by exclusion ability. However, when we examined the number of sessions needed in the discrimination task (Crit1), significant effects of age were found. Previously, in a similar discrimination task, performance was shown to be effected by age in the sense that older dogs needed more sessions to reach a certain criterion (Wallis et al., 2016). Here, we also found that older dogs needed longer to reach a discrimination task criterion. Such age effects on learning, memory and discrimination ability have also been found in other animal studies (Milgram et al., 1994; Tapp, 2003; Lacreuse et al., 2005; Snigdha et al., 2012) and are well known in the human literature (Park et al., 2003; Schönknecht et al., 2005). In general, age effects seem to be absent in easy discrimination tasks and become more apparent the more difficult the task gets, with older individuals needing more time to reach a certain criterion and also making more errors than younger individuals (Matzel & Kolata, 2010; Mell et al., 2005; Tapp, 2003; Voytko, 1999). In this study, clip art pictures of different colour, size and composition were used, requiring the dogs to learn and memorize each stimulus individually, because there were no distinct features that would have made it possible for the animal to use generalisation rules (Wallis et al., 2016). This makes the discrimination learning in this study a complex task, resulting in age effects. Dogs from age group 2 needed longer to reach the discrimination task criterion (needed more sessions) and had a higher error rate. It seems that older dogs had more problems with switching their initial strategy when they preferred the negative over the positive stimulus. Aged dogs show reduced cognitive flexibility (Adams et al., 2000; Chan et al., 2002; Wallis et al., 2016) and have lower working memory (Chan et al., 2002; Tapp et al., 2003), which can be linked to an overall decline in discrimination learning (Mutter et al., 2006). Therefore, it is no surprise that, in our study, older individuals performed worse than younger ones in the discrimination trials. Yet it seems that previous experience in a similar test led to a smaller error rate, meaning individuals that were already familiar with such a study setup needed less correction trials (thus they were able to overcome an initial wrong response and more flexibly switched their response behaviour). In contrast to our predictions, we failed to find an effect of experience on the speed of discrimination learning. There was no difference in the number of sessions needed to reach this criterion for experienced and inexperienced dogs, only a difference in the number of errors they made was found.

Crit2 however did not consist of a simple discrimination, but instead of training the dogs to inhibit touching a novel (but negative) stimulus. Neophilia has been demonstrated in dogs previously (Kaulfuß & Mills, 2008), where dogs picked a novel toy more often than any of two other, already familiar ones when they were free to choose a toy to play with. Here, dogs were confronted with new stimuli in every session, and we considered these trials as

truly novel, as they were presented with a different, new clip art picture in every trial. In the study of Kaulfuß and Mills (2008) the authors did not examine age differences. Other studies, though, with various subject species found that younger individuals tend to be more neophilic towards unknown objects than older ones (Bergman & Kitchen, 2009; Biondi et al., 2010; Benson-Amram & Holekamp, 2012). However, these results are not consistent as in other studies, no such neophilic tendencies were found in younger individuals (Benson-Amram et al., 2013; Kendal et al., 2005). In the current study, we did not find any effect of age on the novelty trials and therefore we can say that, at least in this setup, younger and older dogs exhibit neophilia equally often. Regardless, experience had a significant effect on the number of sessions needed to reach this criterion. The more experienced dogs needed less sessions and made fewer errors in these novelty trials. Experienced individuals have previously been shown to exhibit improved task performance or problem solving (Benson-Amram et al., 2013; Marshall-Pescini et al., 2008). Training has a positive effect on selective and sustained attention (Chapagain et al., 2017), more precisely, it was demonstrated, that attention increased when the dogs were trained throughout their lifetime (e.g., search and rescue training, obedience, trick training, etc.). Attention and training were shown to promote faster problem solving abilities in a detour task (Marshall-Pescini et al., 2016), dogs with training experience (e.g. agility, search and rescue, etc.) had a higher success rate compared to non-trained dogs. In our study, dogs with previous experience were faster to avoid touching the novel stimulus and thus also needed less correction trials (made fewer errors). These trained animals were quicker in remembering and retaining their positive stimulus and switching to the positive stimulus after a mistake.

None of the tested variables were found to significantly influence inference by exclusion trials. Wallis et al. (2016) found an age effect when it came to reasoning by exclusion, older dogs performed better. We could not replicate these findings, because here no age effect was detected. Older and younger individuals performed equally, perhaps because of a smaller sample size and having two age groups instead of five age groups, which would have made a performance gradient more observable. In the touchscreen study of Aust et al. (2008) the authors found that 3 out of 6 dogs used the logic-based strategy of inferring by exclusion. These subjects needed more training trials to master the discrimination task and the authors suggest that this might have been necessary for the dogs to search for logic-based solutions rather than relying on rote learning. This does not align with our findings and could be due to the different testing setups. In Aust et al. (2008) the animals conducted 16 test sessions in general (two times the four test sessions for test 1 and two times the four test sessions for test 2) whereas in our study each dog conducted 30 test sessions, with each session presenting novel stimuli. Generally speaking, our study relied more on concept learning and logical reasoning because the subjects had more sessions with novel test stimuli, so understanding the concept was required to be successful throughout many trials. Whereas in the former study, since there were the same eight test stimuli appearing again, associative learning could have taken place, due to the repetition of the test trials. In our study we had a higher number of participating dogs than Aust et al. (2008) and a more diverse sample than Wallis et al (2016), allowing for closer examination of

inference ability on a more diverse sample and getting a more realistic representation of this strategy in the overall population.

Besides that, some dogs have clearly shown inference by exclusion abilities above chance. It seems that, different inter-individual variations and strategies can influence problem solving approaches and therefore play an important role in exhibiting inference by exclusion. This could be one explanation why this strategy is not represented in all dogs equally. Looking at our data it might be that the representation of inference ability on a population level is not as predominant as the 50% found in Aust et al. (2008) and not as little as the less than 10% found in Wallis et al. (2016), but rather somewhere in between around the 30% found here. It could be that this cognitively demanding strategy might not be exhibited by most dogs and they might utilize simpler problem-solving strategies (e.g., trial and error learning) instead, when seen on a population level. To get a clearer overall picture and more specifically on the rate of applying such complex cognitive strategies (inference learning) more studies on even bigger and diverse sample sizes are needed.

When we compare the Goffin cockatoos (from O'Hara et al., 2015) with our tested dogs, the birds performed better in the exclusion trials than the canids. On a group level about 66% (8/12) of the Goffins, compared to about 30% (13/43) of the dogs exhibited inference by exclusion response above chance. Despite this, the authors could not find a significant influence of sex, experience, discrimination-learning performance during training or the performance in the novelty trials, which was also the case in our study. One has to keep in mind, that some bird species show advanced abilities in physical cognition (Auersperg et al., 2011; Weir et al., 2002), whereas dogs do not (Collier-Baker et al., 2004; Müller et al., 2014). This could be the underlying proposition leading to a better inference by exclusion ability in birds. Still, more comparative studies on various species with a similar, comparable approach are needed.

However, we must also address the physiological component of the study setup, particularly how dogs perceive the stimuli and how this might have influenced their choice. Dogs have been shown to exhibit human like red-green colour blindness (Siniscalchi, et al., 2017) and are sensitive to the luminance levels of the stimuli shown on a computer screen, meaning when intensity levels increase, certain colour discriminations can become more difficult (Byosiere et al., 2019; Siniscalchi et al., 2017; Pretterer et al., 2004). This could have influenced the animals´ choice behaviour, as coloured stimuli were presented on a white background, resulting in a high overall brightness and luminance, making discrimination for some dogs challenging (Byosiere et al., 2019; Pretterer et al., 2004) and possibly impacting their performance and cognitive strategies. Future studies using such a setup should carefully select their testing stimuli, use a lower luminance, brightness and go for achromatic stimuli dogs can easily perceive, when testing using difficult, cognitively more demanding tasks.

References

Adams, B., Chan, A., Callahan, H. & Milgram, N. W. (2000). The canine as a model of human cognitive aging: Recent developments. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *24*(5), 675–692. https://doi.org/10.1016/S0278-5846(00)00101-9

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*(6), e20231. https://doi.org/10.1371/journal.pone.0020231

Aust, U., Range, F., Steurer, M. & Huber, L. (2008). Inferential reasoning by exclusion in pigeons, dogs, and humans. *Animal Cognition*, *11*(4), 587–597. https://doi.org/10.1007/s10071-008-0149-0

Benson-Amram, S. & Holekamp, K. E. (2012). Innovative problem solving by wild spotted hyenas. *Proceedings of the Royal Society B*, *279*(1744), 4087–4095. https://doi.org/10.1098/rspb.2012.1450

Benson-Amram, S., Weldele, M. L. & Holekamp, K. E. (2013). A comparison of innovative problem-solving abilities between wild and captive spotted hyaenas, Crocuta crocuta. *Animal Behaviour*, *85*(2), 349–356. https://doi.org/10.1016/j.anbehav.2012.11.003

Bergman, T. J. & Kitchen, D. M. (2009). Comparing responses to novel objects in wild baboons (Papio ursinus) and geladas (Theropithecus gelada). *Animal Cognition*, *12*(1), 63–73. https://doi.org/10.1007/s10071-008-0171-2

Biondi, L. M., Bó, M. S. & Vassallo, A. I. (2010). Inter-individual and age differences in exploration, neophobia and problem-solving ability in a Neotropical raptor (Milvago

chimango). *Animal Cognition*, *13*(5), 701–710. https://doi.org/10.1007/s10071-010-0319-8

Markson, L. & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature* 385, 813–815 . https://doi.org/10.1038/385813a0

Bussey, T. J., Padain, T. L., Skillings, E. A., Winters, B. D., Morton, A. J. & Saksida, L. M. (2008). The touchscreen cognitive testing method for rodents: how to get the best out of your rat. Learning & memory, 15(7), 516-523. doi: 10.1101/lm.987808.

Bussey, T. J., Holmes, A., Lyon, L., Mar, A. C., McAllister, K. A. L., Nithianantharajah, J., Oomen, C.A. & Saksida, L. M. (2012). New translational assays for preclinical modelling of cognition in schizophrenia: the touchscreen testing method for mice and rats. *Neuropharmacology*, 62(3), 1191-1203. https://doi.org/10.1016/j.neuropharm.2011.04.011

Byosiere, S.-E., Chouinard, P. A., Howell, T. J. & Bennett, P. C. (2019). The effects of physical luminance on colour discrimination in dogs: A cautionary tale. *Applied Animal Behaviour Science*, *212*, 58–65. https://doi.org/10.1016/j.applanim.2019.01.004

Call, J. (2006). Inferences by exclusion in the great apes: The effect of age and species. *Animal Cognition*, *9*(4), 393–403. https://doi.org/10.1007/s10071-006-0037-4

Chan, A. D. F., Nippak, P. M. D., Murphey, H., Ikeda-Douglas, C. J., Muggenburg, B., Head, E., Cotman, C. W. & Milgram, N. W. (2002). Visuospatial Impairments in Aged Canines (Canis familiaris): The Role of Cognitive-Behavioral Flexibility. *Behavioral Neuroscience*, *116*(3), 443–454. https://doi.org/10.1037/0735-7044.116.3.443

Chapagain, D., Virányi, Z., Wallis, L. J., Huber, L., Serra, J. & Range, F. (2017). Aging of Attentiveness in Border Collies and Other Pet Dog Breeds: The Protective Benefits of

Lifelong Training. *Frontiers in Aging Neuroscience*, *9*, 100. https://doi.org/10.3389/fnagi.2017.00100

Collier-Baker, E., Davis, J. M. & Suddendorf, T. (2004). Do Dogs (Canis familiaris) Understand Invisible Displacement? *Journal of Comparative Psychology*, *118*(4), 421–433. https://doi.org/10.1037/0735-7036.118.4.421

Coutanche, M. N. & Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology: General*, *143*(6), 2296–2303. https://doi.org/10.1037/xge0000020

Foster, T. C. (2012). Dissecting the age-related decline on spatial learning and memory tasks in rodent models: N-methyl-D-aspartate receptors and voltage-dependent Ca2+ channels in senescent synaptic plasticity. *Progress in Neurobiology*, *96*(3), 283–303. https://doi.org/10.1016/j.pneurobio.2012.01.007

Guidi, M., Kumar, A. & Foster, T. C. (2015). Impaired Attention and Synaptic Senescence of the Prefrontal Cortex Involves Redox Regulation of NMDA Receptors. *Journal of Neuroscience*, *35*(9), 3966–3977. https://doi.org/10.1523/JNEUROSCI.3523-14.2015

Halberda, J. (2006). Is this a dax which I see before me? Use of the logical argument disjunctive syllogism supports word-learning in children and adults. *Cognitive Psychology*, *53*(4), 310–344. https://doi.org/10.1016/j.cogpsych.2006.04.003

Heibeck, T. H. & Markman, E. M. (1987). Word Learning in Children: An Examination of Fast Mapping. *Child development*, 1021-1034. https://doi.org/10.2307/1130543

Horschler, D. J., Hare, B., Call, J., Kaminski, J., Miklósi, Á. & MacLean, E. L. (2019). Absolute brain size predicts dog breed differences in executive function. *Animal Cognition*, *22*(2), 187–198. https://doi.org/10.1007/s10071-018-01234-1

Kaminski, J. (2004). Word Learning in a Domestic Dog: Evidence for „Fast Mapping“. *Science*, *304*(5677), 1682–1683. https://doi.org/10.1126/science.1097859

Kaulfuß, P. & Mills, D. S. (2008). Neophilia in domestic dogs (Canis familiaris) and its implication for studies of dog cognition. *Animal Cognition*, *11*(3), 553–556. https://doi.org/10.1007/s10071-007-0128-x

Kendal, R. L., Coe, R. L. & Laland, K. N. (2005). Age differences in neophilia, exploration, and innovation in family groups of callitrichid monkeys. *American Journal of Primatology*, *66*(2), 167–188. https://doi.org/10.1002/ajp.20136

Lacreuse, A., Kim, C. B., Rosene, D. L., Killiany, R. J., Moss, M. B., Moore, T. L., Chennareddi, L. & Herndon, J. G. (2005). Sex, Age, and Training Modulate Spatial Memory in the Rhesus Monkey (Macaca mulatta). *Behavioral Neuroscience*, *119*(1), 118–126. https://doi.org/10.1037/0735-7044.119.1.118

Little, A. C. (2013). The influence of steroid sex hormones on the cognitive and emotional processing of visual stimuli in humans. *Frontiers in Neuroendocrinology*, *34*(4), 315–328. https://doi.org/10.1016/j.yfrne.2013.07.009

Markman, E. M. & Abelev, M. (2004). Word learning in dogs? *Trends in Cognitive Sciences*, *8*(11), 479–481. https://doi.org/10.1016/j.tics.2004.09.007

Marshall-Pescini, S., Frazzi, C. & Valsecchi, P. (2016). The effect of training and breed group on problem-solving behaviours in dogs. *Animal Cognition*, *19*(3), 571–579. https://doi.org/10.1007/s10071-016-0960-y

Marshall-Pescini, S., Valsecchi, P., Petak, I., Accorsi, P. A. & Previde, E. P. (2008). Does training make you smarter? The effects of training on dogs' performance (Canis familiaris) in a problem solving task. *Behavioural Processes*, *78*(3), 449–454. https://doi.org/10.1016/j.beproc.2008.02.022

Matzel, L. D. & Kolata, S. (2010). Selective attention, working memory, and animal intelligence. *Neuroscience & Biobehavioral Reviews*, *34*(1), 23–30. https://doi.org/10.1016/j.neubiorev.2009.07.002

McGetrick, J. & Range, F. (2018). Inequity aversion in dogs: A review. *Learning & Behavior*, *46*(4), 479–500. https://doi.org/10.3758/s13420-018-0338-x

McQuail, J. A., Beas, B. S., Kelly, K. B., Simpson, K. L., Frazier, C. J., Setlow, B. & Bizon, J. L. (2016). NR2A-Containing NMDARs in the Prefrontal Cortex Are Required for Working Memory and Associated with Age-Related Cognitive Decline. *The Journal of Neuroscience*, *36*(50), 12537–12548. https://doi.org/10.1523/JNEUROSCI.2332-16.2016

Mell, T., Heekeren, H. R., Marschner, A., Wartenburger, I., Villringer, A. & Reischies, F. M. (2005). Effect of aging on stimulus-reward association learning. *Neuropsychologia*, *43*(4), 554–563. https://doi.org/10.1016/j.neuropsychologia.2004.07.010

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

Milgram, N. W., Head, E., Muggenburg, B., Holowachuk, D., Murphey, H., Estrada, J., Ikeda-Douglas, C. J., Zicker, S. C. & Cotman, C. W. (2002). Landmark discrimination learning in the dog: Effects of age, an antioxidant fortified food, and cognitive strategy. *Neuroscience & Biobehavioral Reviews*, *26*(6), 679–695. https://doi.org/10.1016/S0149-7634(02)00039-8

Milgram, N. W., Head, E., Weiner, E. & Thomas, E. (1994). Cognitive functions and aging in the dog: acquisition of nonspatial visual tasks. *Behavioral neuroscience*, *108*(1), 57. https://doi.org/10.1037/0735-7044.108.1.57

Müller, C. A., Riemer, S., Range, F. & Huber, L. (2014). Dogs' use of the solidity principle: Revisited. *Animal Cognition*, *17*(3), 821–825. https://doi.org/10.1007/s10071-013-0709-9

Mutter, S. A., Haggbloom, S. J., Plumlee, L. F. & Schirmer, A. R. (2006). Aging, Working Memory, and Discrimination Learning. *Quarterly Journal of Experimental Psychology*, *59*(9), 1556–1566. https://doi.org/10.1080/17470210500343546

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

Park, H. L., O'Connell, J. E. & Thomson, R. G. (2003). A systematic review of cognitive decline in the general elderly population. *International Journal of Geriatric Psychiatry*, *18*(12), 1121–1134. https://doi.org/10.1002/gps.1023

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

Pilley, J. W. & Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes*, *86*(2), 184–195. https://doi.org/10.1016/j.beproc.2010.11.007

Pretterer, G., Bubna-Littitz, H., Windischbauer, G., Gabler, C. & Griebel, U. (2004). Brightness discrimination in the dog. *Journal of Vision*, *4*(3), 10. https://doi.org/10.1167/4.3.10

Range, F., Aust, U., Steurer, M. & Huber, L. (2008). Visual categorization of natural stimuli by domestic dogs. *Animal Cognition*, *11*(2), 339–347. https://doi.org/10.1007/s10071-007-0123-2

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

Schönknecht, P., Pantel, J., Kruse, A. & Schröder, J. (2005). Prevalence and Natural Course of Aging-Associated Cognitive Decline in a Population-Based Sample of Young-Old Subjects. *American Journal of Psychiatry*, *162*(11), 2071–2077. https://doi.org/10.1176/appi.ajp.162.11.2071

Serpell, J. A. & Hsu, Y. A. (2005). Effects of breed, sex, and neuter status on trainability in dogs. *Anthrozoös*, *18*(3), 196–207. https://doi.org/10.2752/089279305785594135

Siniscalchi, M., d'Ingeo, S., Fornelli, S., & Quaranta, A. (2017). Are dogs red–green colour blind? *Royal Society Open Science*, *4*(11), 170869. https://doi.org/10.1098/rsos.170869

Snigdha, S., Christie, L.-A., De Rivera, C., Araujo, J. A., Milgram, N. W. & Cotman, C. W. (2012). Age and distraction are determinants of performance on a novel visual search task in aged Beagle dogs. *AGE*, *34*(1), 67–73. https://doi.org/10.1007/s11357-011-9219-3

Spiegel, C. & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology*, *109*(1), 132–140. https://doi.org/10.1016/j.jecp.2010.10.013

Spinelli, S., Pennanen, L., Dettling, A. C., Feldon, J., Higgins, G. A. & Pryce, C. R. (2004). Performance of the marmoset monkey on computerized tasks of attention and working memory. Cognitive brain research, 19(2), 123-137. https://doi.org/10.1016/j.cogbrainres.2003.11.007

Tapp, P. D. (2003). Size and Reversal Learning in the Beagle Dog as a Measure of Executive Function and Inhibitory Control in Aging. *Learning & Memory*, *10*(1), 64–73. https://doi.org/10.1101/lm.54403

Tapp, P. D., Siwak, C. T., Estrada, J., Holowachuk, D. & Milgram, N. W. (2003). Effects of age on measures of complex working memory span in the beagle dog (Canis familiaris) using two versions of a spatial list learning paradigm. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *10*(2), 148–160. https://doi.org/10.1101/lm.56503

Troche, S. J., Wagner, F. L., Voelke, A. E., Roebers, C. M. & Rammsayer, T. H. (2014). Individual differences in working memory capacity explain the relationship between general discrimination ability and psychometric intelligence. *Intelligence*, *44*, 40–50. https://doi.org/10.1016/j.intell.2014.02.009

Vigo, R. & Allen, C. (2009). How to reason without words: Inference as categorization. *Cognitive Processing*, *10*(1), 77–88. https://doi.org/10.1007/s10339-008-0220-4

Voytko, M. L. (1999). Impairments in acquisition and reversals of two-choice discriminations by aged rhesus monkeys. *Neurobiology of Aging*, *20*(6), 617–627. https://doi.org/10.1016/S0197-4580(99)00097-4

Wallis, L. J., Range, F., Müller, C. A., Serisier, S., Huber, L. & Zsó, V. (2014). Lifespan development of attentiveness in domestic dogs: Drawing parallels with humans. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00071

Wallis, L. J., Virányi, Z., Müller, C. A., Serisier, S., Huber, L. & Range, F. (2016). Aging effects on discrimination learning, logical reasoning and memory in pet dogs. *AGE*, *38*(1), 6. https://doi.org/10.1007/s11357-015-9866-x

Weir, A. A. S., Chappell, J. & Kacelnik, A. (2002). Shaping of Hooks in New Caledonian Crows. *Science*, *297*(5583), 981–981. https://doi.org/10.1126/science.1073433

Zhao, X., Zhou, R. & Fu, L. (2013). Working Memory Updating Function Training Influenced Brain Activity. *PLoS ONE*, *8*(8), e71063. https://doi.org/10.1371/journal.pone.0071063