Masterarbeit / Master's thesis

Titel der Masterarbeit / Title of the Master's Thesis

"Direct reputation formation in pet dogs *(Canis lupus familiaris)* and its correlation with experience in years"

verfasst von / submitted by

Eva Martinelli, BEd

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

Master of Education (MEd)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 199 502 529 02
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Lehramt Sek (AB) Unterrichtsfach Biologie und Umweltbildung Unterrichtsfach Spanisch
Betreut von / Supervisor:	Mag. Dr. Friederike Range

Abstract

The ability to form reputations is an important skill that can lead to successful cooperation, which in turn can increase an animal's chance of survival. Reputations can be formed through direct experience, which is the most accurate way to form meaningful opinions because it is first-hand experience. Previous research has focused on whether dogs *(Canis lupus familiaris)* can form reputations of unfamiliar humans after directly interacting with them because of their ability to cooperate with them. However, the results of these studies are mixed, which is why we wanted to investigate this further. Our study investigated whether dogs would prefer a generous person, who fed the dog, over a selfish person, who did not feed the dog, in a food-giving situation by approaching them first or spending more time with them. In a pilot study, which consisted of a direct experience-phase and choice trials and whose results were implemented in the main study, 20 dogs were tested to see how many choice trials dogs needed to form a reputation above chance. In the main study, 40 dogs were tested in two sessions, one consisting of a direct experience-phase and choice trials, and one consisting of choice trials only. In the main study, we also tested whether older dogs would outperform younger dogs due to greater life experience and socialization. Unlike our pilot study, our main study revealed that dogs cannot form opinions about people through direct interaction, and the results did not differ between age groups. Based on the negative results of our study, it seems that direct reputation formation is not a simple skill and that further research is needed to come to a general conclusion as to whether dogs are capable of reputation formation and whether age has an effect on this ability.

Keywords: Direct reputation formation, pet dogs, age effect, food-giving situation, generous/selfish partner, life experience, socialization, first-hand experience

Zusammenfassung

Die Fähigkeit Reputationen zu formen, ist eine wichtige Fähigkeit, da sie zu einer erfolgreichen Zusammenarbeit führen kann, was wiederum die Überlebenschancen eines Tieres erhöhen kann. Man spricht hier im Englischen von „reputation formation“. Eine Reputation kann durch direkte Erfahrung gebildet werden, diese ist die einfachste Art, um Reputationen zu formen, da die Eindrücke aus erster Hand stammen und somit am verlässlichsten sind. Etliche Studien haben sich mit der Frage beschäftigt, ob Hunde *(Canis lupus familiaris)* aufgrund ihrer Kooperationsfähigkeit eine Reputation von unbekannten Menschen bilden können, nachdem sie direkt mit ihnen interagiert haben. Die Ergebnisse dieser Studien sind jedoch durchwachsen, weshalb wir diese Frage weiter untersuchen wollten. In unserer Studie wurde untersucht, ob Hunde in einer Situation, in der sie Futter bekommen, eine großzügige Person, die den Hund füttert, einer egoistischen Person, die den Hund nicht füttert, vorziehen würden, indem sie sich ihr zuerst näherten oder mehr Zeit mit ihr verbringen würden. In einer Pilotstudie, die aus einer direkten Erfahrungsphase (“experience-phase“) und Auswahlversuchen (“choice-trials“) bestand und deren Ergebnisse in die Hauptstudie einflossen, wurden 20 Hunde getestet, um festzustellen, wie viele Auswahlversuche die Hunde benötigten, um Reputationen zu formen, welche nicht nur durch Zufall entstanden. In der Hauptstudie wurden 40 Hunde in zwei Sitzungen getestet, von denen eine aus einer direkten Erfahrungsphase und Auswahlversuchen bestand und die andere nur aus Auswahlversuchen. In der Hauptstudie testeten wir auch, ob ältere Hunde aufgrund ihrer größeren Lebenserfahrung und Sozialisierung besser abschneiden würden als jüngere Hunde. Im Gegensatz zu unserer Pilotstudie zeigte unsere Hauptstudie, dass Hunde sich durch direkte Interaktion keine Meinung über Menschen bilden können, und die Ergebnisse unterschieden sich nicht zwischen den Altersgruppen. Aufgrund der negativen Ergebnisse unserer Studie scheint es, dass die direkte Reputationsbildung keine einfache Fähigkeit ist und dass weitere Forschung notwendig ist, um zu einer allgemeinen Schlussfolgerung zu kommen, ob Hunde zur Reputationsbildung fähig sind und ob das Alter einen Einfluss auf diese Fähigkeit hat.

Schlagwörter: direkte Reputationsbildung, Alterseffekt, futtergebende Situation, großzügiger/ egoistischer Partner, Lebenserfahrung, Sozialisierung, direkte Erfahrung

Acknowledgments

First of all, I would like to thank my co-supervisor Dr. Hoi-Lam Jim for always helping out with the data collection and always being there if questions arose. Also many thanks for her improvements on the masters' thesis and helpful tips regarding the whole study.

Furthermore, I would also like to thank Mag. Dr. Friederike Range, who made this possible in the first place and who also actively supported me with open questions.

Further thanks go to Kadisha Belfiore, BSc, who studied indirect reputation formation in pet dogs and with whom I was able to conduct this study. We recruited dogs together, shared the organizational work and supported each other with statistics and open questions.

Finally, I would like to thank all the volunteers who supported us as partners without whom this study would not have been possible. Accordingly, thanks go to Natasa Ciric, Stefanie Karner and Eva Kopf who volunteered during the pilot study. Furthermore, I would like to thank Angela Stojan, Klara Schulmeister, Michaela Tost, Sari Komulainen, Zsanett Györi, Katharine Creagh and Mayte Martínez for participating in the main study. They invested a lot of time and energy to make this possible.

Table of contents

1 Introduction

1.1 Reputation formation

Reputation or social evaluation refers to knowledge about an individual's typical behavior based on knowledge of that individual's past behavior (Russell, 2007) and is a useful skill for all animals that interact with other individuals to predict their behavior. There are two ways to form reputations: through direct or indirect experience (eavesdropping) (Subiaul et al., 2008). Direct judgments are more accurate than indirect ones, as they are based on first-hand experience.

Now, when opinions are formed on an indirect or direct basis, two different types of biases can arise on which they are based. These biases contribute to an animal's fitness. First, there is the negativity bias, which is an aversion towards anything negative, and second, there is the positivity bias, which is the preference towards anything positive. Both are important for survival because on one hand, interacting with negative stimuli like an antisocial individual may be dangerous, thus avoiding it would increase an individual's chance of survival. On the other hand, interacting with prosocial individuals may lead to significant gains, such as resources, help, cooperation, etc. (Rault, 2019), thus a preference towards it would be advantageous (Abdai & Miklósi, 2016). For example, capuchin monkeys *(Cebus apella)*, which are known to be prosocial and cooperate in many situations, show an aversion towards an antisocial partner after observing that the partner refused to help a conspecific, but they do not show a preference for a prosocial partner. Therefore, they showed a negativity bias, which means an avoidance of the antisocial partner (Anderson et al., 2013). Group-living animals, including humans, inevitably interact with unfamiliar individuals from time to time, so reputation formation is an invaluable skill for when they do so to reduce the risk of bad decision in future interactions with the same individual.

Hamlin and colleagues conducted a series of experiments between 2007 and 2014 to test whether human infants could differentiate between a prosocial and an antisocial partner. Hamlin et al. (2007) found that 6- and 10-month-old infants chose a prosocial partner more often than an antisocial or neutral partner when given the choice. They even looked longer at the prosocial partner, suggesting that infants could socially evaluate others. In Hamlin et al. (2010) it is shown that even 3-month-old infants can socially evaluate others based on interactions. Furthermore, Hamlin and Wynn (2011) found that infants of 5 and 9 months chose the prosocial partner more often, infants of 3 months looked at this partner more frequently and in Hamlin et al. (2011) it is stated that 5- and 8-month-old infants, as well as 19- to 24-month-old infants are capable of selective social evaluation. They tested if the infants would prefer positive acting towards a prosocial partner and negative acting towards an antisocial one and they found that from 8-months onwards infants can consider the whole context and not only the outcome. However, Scarf et al. (2012) was concerned about the methods and believed that some aspects could have influenced the results. For example, there were two events which could have been associated with the prosocial individual, but only one which could have been associated with the antisocial individual and this could have led to the fact that the helper was viewed as more positive. Additionally, Hamlin (2013) found that 8-month-old infants were capable of distinguishing between the behavior of people and not only the positive or negative outcome. But according to Hamlin (2014) even 5-month-old infants consider the context if given enough repetitions. These studies (Hamlin et al., 2010, Hamlin et al., 2011, Hamlin, 2013) show that only negativity bias can be detected in infants, and at around 5 to 6 months of age, positivity bias emerges (Hamlin et al., 2010, Hamlin et al., 2011, Hamlin, 2013).

Another study by Dahl et al. (2013) found that 17- and 22-month old infants help a prosocial and an antisocial partner with equal frequency, which means that they were not capable of forming reputations, while 26-month-old infants prefer to help the prosocial one. The results here show that the ability of reputation formation changes as the age of the infant increases. Herrmann et al. (2013) also found that infants of 30 months of age preferred the prosocial partner. Overall, one should be careful to draw conclusions about the functionality of the reputation formation system in children. Also, there are methodological differences, such as different age groups and the usage of inanimate objects as well as humans as partners, across these studies, which could be why they found divergent results (Abdai & Miklósi, 2016).

1.2 Studies of non-human apes

It is unknown whether animals possess the ability to form reputations or whether it is a uniquely human skill (Subiaul et al., 2008). Many primate species perform behaviors that benefit others; for example, chimpanzees *(Pan troglodytes)* engage in food sharing, grooming, consolation and coalition forming (Boesch & Boesch-Achermann, 2000, de Waal, 1996, Goodall, 1986, Muller & Mitani, 2005, O'Connell, 1995, Preston & de Waal, 2002) and often have preferences with whom to interact suggesting that they form at least preferences if not reputations of others.

Herrmann et al. (2013) directly compared 2.5-year-old human children, chimpanzees, bonobos *(Pan paniscus)* and orangutans *(Pongo pygmaeus)* in their ability to distinguish between a nice and a selfish human partner. They received food from the nice experimenter whereas the selfish experimenter withheld food. Afterwards, the subjects had the chance to approach one of the two experimenters. The results showed that only orangutans and human children significantly preferred the nice experimenter over the selfish experimenter, which suggests they were able to form a reputation of the humans, while bonobos and chimpanzees did not.

Subiaul et al. (2008) also tested whether chimpanzees are capable of learning which human is more likely to give food. There was one generous partner who gave food and one selfish partner who withheld food. Seven chimpanzees were tested and four showed a significant preference for the generous partner in 7 out of 8 trials. One might argue that forming reputations might be especially important for chimpanzees, because they hunt in groups and predicting the behavior of conspecifics could be helpful for more success (Watts & Mitani, 2002), but studies on chimpanzees are normally conducted using humans because it is not possible to control animals in an experimental context and tell them what to do, i.e., be generous or selfish, therefore, it is furthermore logical to also test reputation formation on species that rely on humans for social information, such as domesticated animals (Botigué et al., 2017). Even though chimpanzees were successful here, these results of captive animals, since artificial situations were created, are not generalizable to the wild population. Ecological validity refers to representing the integrity of a real-life situation (Schmuckler, 2001), which is why studies testing chimpanzees with humans have less ecological validity, as they do not interact with humans in the wild, than studies with domesticated species that also interact with humans in real-life situations and therefore offer more ecological validity.

1.3 Studies of domesticated animals with a focus on dogs

Due to the fact that domesticated animals, as already mentioned, also interact with humans in everyday situations, it is particularly relevant to test whether they are capable of forming reputations. There are several ways to test this, mostly by creating a human-animal interaction without food as a reward (Trösch et al., 2020, Leete et al., 2020, Nitzschner et al., 2012), a food-giving situation (Chijiiwa et al., 2020, Carballo et al., 2015, Carballo et al., 2017, Heberlein et al., 2016, Heberlein et al., 2017, McGetrick et al., 2021, Jim et al., 2022) or a helping situation (Piotti et al., 2017).

Reputation formation was tested in different domesticated species, for example Trösch et al. (2020) tested in an unable vs. unwilling paradigm whether horses *(Equus caballus)* could discriminate between different actions performed by an individual. They were tested in different conditions and the results showed that horses were able to form a reputation about humans and also that they were capable of spontaneously discriminating between the conditions without any training beforehand.

Furthermore, Leete et al. (2020) tested if cats *(Felis catus)* could distinguish between different behaviors. This study showed that they were not able to form any reputation because they were not more likely to approach the friendly partner faster and did not spend more time with this partner. However, Chijiiwa et al. (2020) compared the behavior of cats and dogs *(Canis lupus familiaris)* in a different setting. This study showed that, because they tended to choose a container which was preferred by a human, both species are attracted by human behavior, even though they have different domestication histories. This again proves the before mentioned ecological validity.

Several studies (Nitzschner et al., 2012, Carballo et al., 2015, Carballo et al., 2017) tested dogs using a generous and a selfish partner in a food-giving situation to find out if they would prefer the generous partner over the selfish one. Nitzschner et al. (2012) found that the dogs spent more time next to the nice experimenter than the ignoring experimenter, whereas in Carballo et al. (2015) in the beginning the dogs could not discriminate between the two partners even though they got food from the generous partner. Only after repeated interactions were they able to form a reputation. Then Carballo et al. conducted another study in 2017 to find out if dogs' different experiences with humans leads to better or worse discrimination between generous and selfish behavior, testing family dogs, shelter dogs and puppies in a food-giving situation. They found that the adult dog groups preferred the generous person, but puppies did not. This might be due to lack of experience with humans. Thus, this study suggests that years of experience matter when it comes to discriminating between humans. However, more research needs to be done on the impact that ontogeny plays in this regard. In Carballo et al. (2015) and (2017) food was used as a reward, whereas in Nitzschner et al. the human-animal interaction alone was decisive.

To continue with more food-giving situations, McGetrick et al. (2021) tested whether dogs would reciprocate help they received from a human in a food-giving situation and found that dogs did not reciprocate help, which shows that they were not able to form a reputation in this setting. Another study by Heberlein et al. (2017) tried to find out whether dogs could form reputations about a competitive and a cooperative partner. The results showed that the dogs were more likely to lead the cooperative partner to the preferred food box and this effect even got stronger on the second test day, which shows that dogs were able to form reputations of the two partners. Interestingly, Heberlein and colleagues conducted a similar study in 2016 with dogs and wolves *(Canis lupus)* at the Wolf Science Center (WSC), which have had equal experiences and socialization with humans. They found out that dogs and wolves showed the location of the food more often to the cooperative partner than to the competitive one. This shows that they were also able to discriminate between the cooperative and the competitive partner. Although Heberlein et al. (2017) found that wolves and dogs adjusted their showing behavior to the cooperativeness of the human, Jim et al. (2022) conducted a study on whether dogs and wolves could form direct and indirect reputations of humans in a food-giving situation. Unfortunately, the results did show that dogs and wolves were not able to form reputations after indirect or direct experience. Testing the wolves and dogs at the WSC can elucidate whether differences in their abilities are due to genetics, since they share the same ontogeny. At the same time, testing dogs with different ontogenies in the same experimental setup could elucidate how life experiences affect their abilities, and age correlates with the amount of life experience or socialization with humans they have had. Therefore, this study forms a follow-up of Jim et al. (2022).

Another setting in which dogs get tested frequently is the helping situation without a food reward. Piotti et al. (2017) investigated whether dogs form opinions about humans based on their skillfulness. The dog either received help from a partner or did not but the results showed that dogs were not able to form reputations of humans based on quality of interaction or skillfulness.

1.4 Study aim and hypothesis

In general, the study consisted of two parts, the pilot study and the main study. The aim of the pilot study was to find out how many trials dogs needed to reliably choose the generous person above chance. Subiaul et al. (2008) did this with chimpanzees, finding that four of them preferred the nice partner after a certain number of trials (in this case after 15–75 trials, depending on the individual).

The aim of the main study was to investigate whether dogs were capable of distinguishing between humans after directly interacting with them. We also wanted to test whether dogs with more experience with humans are better at forming reputations of them than younger dogs. As dogs with more experience with humans are better at discriminating humans, we predicted that older dogs, which have spent more time with humans and are therefore more experienced, would significantly prefer the generous partner over the selfish partner after direct experience. We also predicted that younger dogs could discriminate between the generous or selfish partner after direct experience up to some point, but they will not perform as well as older dogs based on their lack of experience with humans. In addition, we wanted to find out whether dogs would perform better in their second session with a little time in-between sessions, than in their first session. We argued that the difference between session 1 and session 2 should be significant due to their gained experience.

We investigated this using a begging situation in which the dogs either received food from a human or the human withheld food because this kind of situation might be more relevant for dogs than getting help from a human and food increases motivation in dogs.

2 Material and Methods

2.1 Ethical statement

Ethical approval was obtained from the 'Ethik und Tierschutzkommission' of the University of Veterinary Medicine Vienna (Protocol number ETK-023/02/2021). The subjects' participation in the study was voluntary. If the animal was not motivated to participate, the session was canceled, and the animal excluded. The individual persons who participated in this study have given approval to publish photographs and videos that show their identity.

2.2 Pilot study

This study was conducted in May 2021.

2.2.1 Study subjects

We recruited all of the dogs from a database of owners at the Clever Dog Lab (CDL) at the University of Veterinary Medicine Vienna, who have volunteered to participate in behavioral studies.

We recruited 20 dogs of various breeds older than one year (see Table 1). The recruited dogs had a large age range from 1-13 years, and we counterbalanced the sex of the dogs (10 males, 10 females). One dog, marked red, dropped out of the study.

Table 1. Name, sex, age and breed of dogs participating in the pilot study.

Name	**Sex**	**Age**	**Breed**
Mr. Darcy	M	1	Golden Retriever
Fanny	F	1	Labrador
Easy	M	2	King Charles Spaniel
Coco	F	3	Jack Russell Terrier
Nala	F	3	Goldendoodle
Manni	M	4	Pyrenean Sheepdog
Lillybet	F	5	Samoyed
Filou	M	6	Poodle

Uno	M	6	King Charles Spaniel
Cleo	F	6	Labrador
Samira	F	7	Yorkshire Terrier
Ylvi	F	7	Australian Shepherd
Cameron	M	8	Border Collie
Roy	M	9	Terrier
Rigo	M	9	Mix
Mia	F	9	Smooth Collie
Lilly	F	10	Pomeranian
Flamme	M	13	Pyrenean Sheepdog
Emma	F	13	Mix
Bounty (excluded)	M	5	King Charles Spaniel

2.2.2 Experimental setup

The experiment took place at the outdoor test enclosure at the CDL situated within the University of Veterinary Medicine Vienna (Figure 1). In this outdoor enclosure there were two circles in which the different partners stood, which were 4 meters apart, and a center point between these circles which was 2 meters away from each circle. There was also a chair for the owner, which was 3 metrers away from the center point. A camera pointed at the owner filmed the whole study. The entrance to the enclosure was opposite the owner and was largely covered by a paravent.

We marked the points (labelled P1 on the left side and P2 on the right side) where people should stand and the cross, which marked the center point between P1 and P2, by spray painting the grass. P1 and P2 were marked, as mentioned before, two meters from the center point. The circle around P1 and P2 had a radius of 50 cm. A GoPro camera (Hero 4 Black) was placed on a tripod and recorded the whole experiment. At the entrance to the enclosure there was a paravent to prevent the dog from seeing the partners beforehand, leaving half of the entrance free so that partners could enter and leave. During the test trials the owner was sitting on a chair three meters from the marked points with the dog on the leash. The experimenter stood outside the enclosure and conducted the whole experiment.

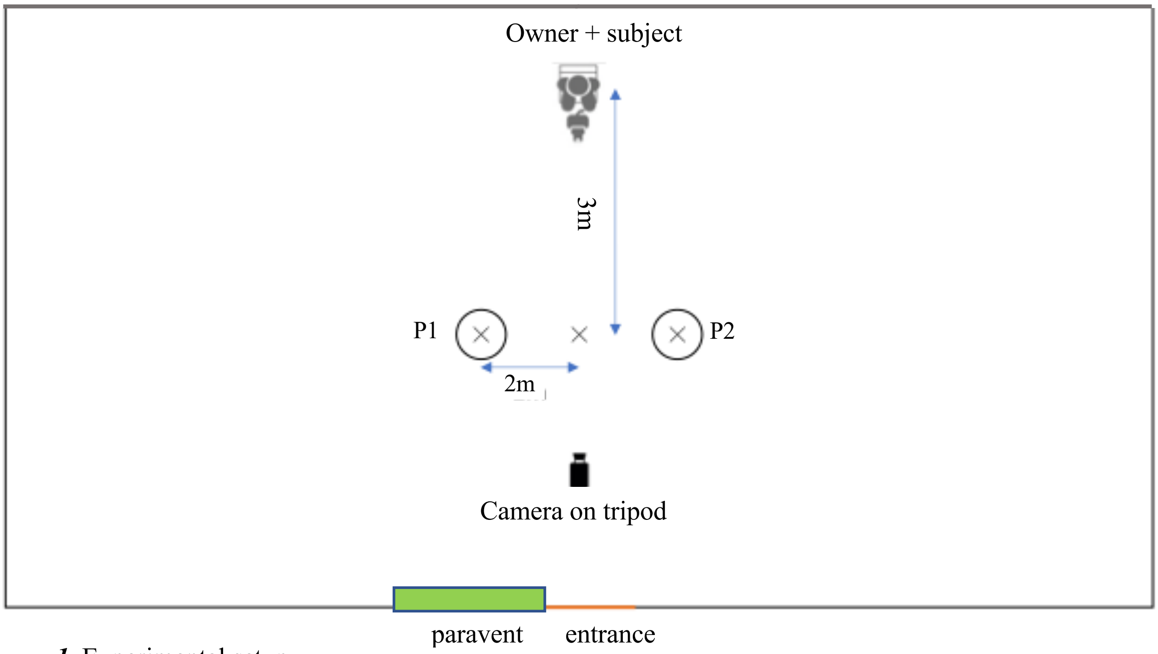


Figure 1. Experimental setup.

2.2.3 Experimental design

Four women, all of them unfamiliar with the dogs, acted as partners in this study. Two individuals always formed a pair of partners that was fixed for each subject. Also, one partner wore white clothes and the other wore black clothes throughout the sessions to help the animal distinguish between the generous and selfish roles. The partner's role and the color of their clothes were randomized and fixed within subjects, but counterbalanced between subjects. They each wore a dark blue hip bag that was facing backwards with approximately 30 sausages inside so that the dogs could not be influenced by the different intensities of smell.

2.2.4 Experimental procedure

The pilot study consisted of one session. Before the experiment, the dog could explore the enclosure freely for a few minutes while the experimenter explained the procedure to the owner. When the dog was comfortable with the environment, the experiment started. The owner put the leash on the dog and walked to their position (P1 or P2, counterbalanced across subjects).

The session was divided into two phases:

1. Experience-phase: The owner stood in a position with the dog on the leash (e.g., P2). The two partners were hidden behind a paravent. One partner entered the enclosure and stood on P1 with a piece of sausage in her right hand; the order of which partner entered first was counterbalanced across subjects. The owner walked up to the middle point and the dog could continue to approach the partner alone. When he/she approached the generous partner, she raised her hand to show the piece of sausage to the dog and said, “Du kannst das haben!” or “Here you go!” and fed the dog (Figure 2a). After the interaction, the owner walked back to P2 with the dog; when the dog turned away, the generous person left the enclosure and then the selfish person entered. The procedure was repeated with the selfish partner, except that the dog was not fed. When the partner raised her hand as the dog approached, she said, “Du kriegst das nicht!” or “You can’t have it!” and then crossed her arms, turned around and kept the food in her hand (Figure 2b). This happened four times, so the dog approached every partner two times from P2 and then the owner and the dog swapped positions, so they then stood on P1. The partners then entered and stood on P2 and the interactions happened four more times. Therefore, each partner interacted with the dog four times in total (Figure 4).

Figure 2a. Generous interaction with the dog.

(Picture taken by Kadisha Belfiore)

Figure 2b. Selfish interaction with the dog.

(Picture taken by Kadisha Belfiore)

2. Test phase: The owner sat on the chair and held his/her dog by the collar and was blindfolded so he/she could not influence the dog's behavior. Both partners entered the test enclosure and stood on P1 and P2, each holding a piece of sausage in their right hand. Their positions were randomized and counterbalanced across subjects. The experimenter called, "Okay!" to indicate to the owner to release the dog, and started the stopwatch. At the same moment, both partners raised their right hand. If the dog would not leave the owner, the owner could give him/her a nudge by touching the dog with both, hands or legs at the same time to not send him/her towards a particular direction. If the dog approached a partner, i.e., placing both front paws into or on the circle and looking at the partner, it was defined as choice. Some dogs were hesitant to approach the unfamiliar people. If the dog walked towards and gazed at a partner for two seconds, she reacted and this was considered to be a choice. When the dog chose a partner, she would act the same way as in the experience-phase, i.e., if the dog chose the generous partner, she would feed the dog (Figure 3a), and if the dog chose the selfish partner, she would not feed the dog (Figure 3b). As soon as this first choice was made, the second person moved her hands to her chest to show that she was no longer going to feed the dog. If the dog approached the second person after its first choice, this person ignored the dog. If the dog did not make a choice within 15 seconds, it was defined as a "no choice" response. After the trial, the experimenter called, "Stop!" and the owner called the dog back and held him/her by the collar. Some dogs were attached to a towing line to easily bring the dog back to the owner. If the generous partner was chosen, she would rebait herself, but both partners moved their hand to their hip bag so that the

dog would not react more strongly to the one person moving their hands and therefore choose that person (stimulus enhancement).

This procedure was repeated for a total of 30 trials (see Figure 4), and the partners' positions were semi-randomized for each trial so that they never stayed in the same position more than three times in a row.

Figure 3a. Generous interaction with the dog after choice.
(Picture taken by Eva Martinelli)

Figure 3b. Selfish interaction with the dog after choice.
(Picture taken by Eva Martinelli)

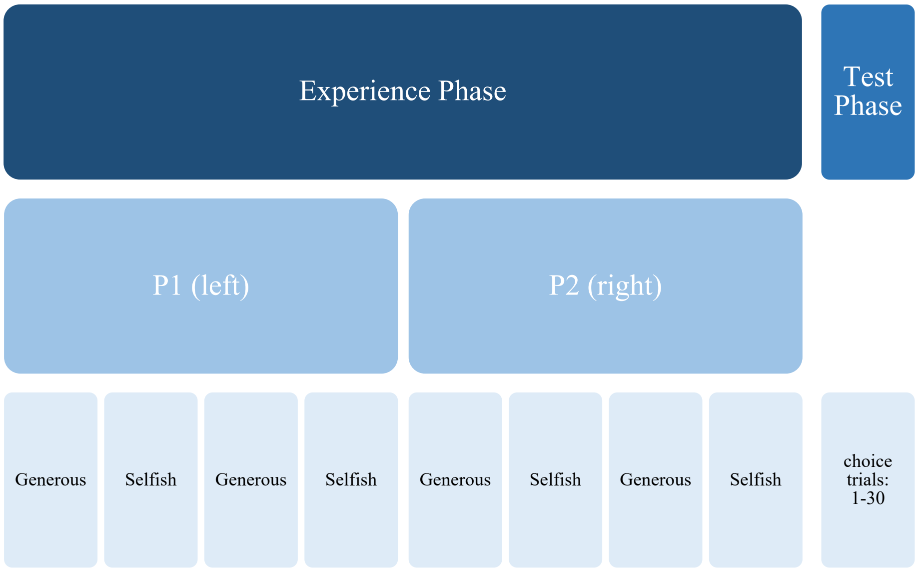


Figure 4. An example of the procedure of the pilot study.

2.2.5 Behavioral analysis

The videos were uploaded to Loopy by Loopbio, a software to store, label, and code experimental videos. We defined when the trial started and ended. The start of each trial was indicated by the "Okay!" said by the experimenter or the beep of the stopwatch, whichever came first. The end was indicated by the "Stop!" said by the experimenter or by the stopwatch beeping again, whichever came first. Then we coded for each trial whether the dog chose the generous or the selfish partner first, which was defined by whether the dog approached a partner, i.e., placing both front paws into the circle and looking at the partner. If there was no choice within the trial, we coded it as NA.

2.2.6 Statistical analysis

All statistical analyses were carried out using the program R (version 4.1.1 R Foundation for Statistical Computing 2021). To test whether the dogs significantly chose the generous person more than the chance level (0.5), we conducted one-sample t-tests. We excluded all NAs from the data and split the trials into blocks of five because we argued that in previous studies on reputation formation (Carballo et al., 2015, Carballo et al., 2017, Jim et al., 2021, Jim et al., 2022), the minimum number of trials was 6, which means that our lowest number of trials would have been six and in order to test whether the dogs chose the generous person above chance and at the same time to be able to make a reliable statement about it, dogs should choose the generous person 5 out of 6 times, as this very certainly indicates that dogs could form a reputation and still allow them to make a mistake. That is why we statistically analyzed each block of 5 separately until we came to one block that was significant.

2.2.7 Results

Dogs did not prefer the generous person above chance in Block 1 (the first five trials) (t(94) = 1.130, *p* = 0.261). However, Block 2 showed that the dogs significantly chose the generous person above chance between trials 6 to 10 (t(91) = 5.195, *p* = <.001).

2.3 Main study

The study was conducted from May to July 2021.

2.3.1 Study subjects

We recruited 44 new dogs of various breeds, which were between 1-12 years old and the sex of the dogs was counterbalanced (22 males and 22 females). They were all recruited from a database of owners at the CDL at the University of Veterinary Medicine in Vienna, who have volunteered to participate in behavioral studies (Table 2). Two dogs dropped out due to time constraints, one dog was excluded due to insufficient food motivation and one dog dropped out due to illness. Thus, the final sample was 40 dogs.

The dogs were classified into the following age categories: 'Early adulthood' (1-3 years), 'Middle age and late adulthood' (4-7 years), and 'Senior and geriatric' (8-12 years) in accordance with published studies: Wallis et al. (2020) defined 1- to 3-year-old dogs as early adulthood and 4-to-6-year-old dogs as middle age. Chapagain et al. (2017) stated that dogs between 6 and 7 years were in their late adulthood, seniors were 8-to-10-year-old dogs and geriatric were dogs older than 10. Based on our sample size, we categorized the age groups accordingly, so that the size of each group would be as balanced as possible. Without the dogs that dropped out or were excluded, there were 11 dogs in 'Early adulthood', 15 dogs in 'Middle age and late adulthood' and 14 dogs in 'Senior and geriatric'.

A total of 10 women acted as the partners in the study. We used different individuals in the two sessions but their roles were fixed within a session (e.g., individuals A and B in session 1, individuals C and D in session 2). The teams were never mixed, so individual A always stayed with individual B.

Table 2. Individual characteristics of dogs in the main study. The dogs that got excluded or dropped out are in red.

Name	Sex	Age	Age category	Breed	First session
Spaiky	M	1	Early adulthood	Cane Corso	Experience
Gandalf	M	1		Andalusian Hound	No experience
Dunni	F	1		Miniature Pinscher	No experience
Jazzie	F	1		Labrador	Experience
Timo	M	2		Mix	No experience
Crash	M	2		Australian Shepherd	Experience
Daytona	F	2		Australian Shepherd	Experience
Franzi	F	2		Rottweiler	Experience
Holly	F	2		Rottweiler	Experience
Zazu	F	2		Miniature Pinscher	No experience
Ravenna	F	3		American Staffordshire Terrier	No experience
Cody	M	4	Middle age and late adulthood	Siberian Husky	No experience
Sixtus	M	4		Petit Brabancon	No experience
Asha	F	4		Rhodesian Ridgeback	No experience
Ajani	F	4		Mix	No experience
Django	M	5		American Staffordshire Terrier	Experience
Chivas	M	5		Siberian Husky	Experience
Ellis	F	5		Portuguese Water Dog	Experience
Lilu	F	5		Mix	No experience
Lenny1	M	6		Canarian Warren Hound	No experience

Ozzy	M	6		Labrador	No experience
Kiki	F	6		Mix	Experience
Bailey	F	7		Pitbull Terrier	Experience
Capper	M	7		Andalusian Hound	No experience
Graf Aidan	M	7		Irish Setter	Experience
Rusty	M	7		Mix	No experience
Aaron	M	8	Senior and geriatric	Labrador	Experience
Lilly	F	8		Miniature Pinscher	Experience
Helena	F	9		Poodle	No experience
Shari	F	9		Rottweiler	Experience
Sally	F	9		Mix	Experience
Fiona	F	9		Rhodesian Ridgeback	No experience
Alois	M	9		Mix	Experience
Snoopy1	M	10		Shih Tzu	No experience
Jolie	F	10		Mix	No experience
Mozart	M	11		Labrador	No experience
Monty	M	11		Border Collie	Experience
Ultimo	M	11		Border Collie	No experience
Amy	F	11		Border Collie	Experience
Snoopy2	M	12		Beagle	Experience
Pauli	M	2	Early adulthood	Labrador	Experience
Anni	F	3		Mix	No experience
Fee	F	7	Middle age and late adulthood	Rottweiler	Experience

Lenny2	M	12	Senior and geriatric	Border Collie	Experience

2.3.2 Experimental setup

The experimental setup was the same as in the pilot study.

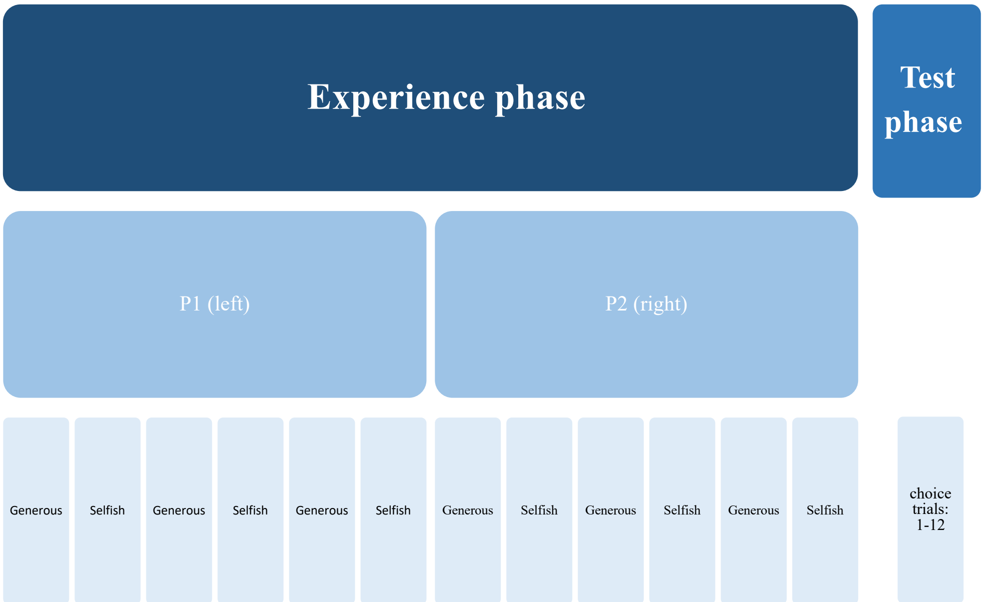
2.3.3 Experimental design

The dogs participated in two sessions. One session consisted of an experience-phase followed by twelve choice trials (experience session). The other session only consisted of twelve choice trials with no prior experience-phase (no experience session). The sessions were counterbalanced across subjects (i.e., half of the dogs did the experience session first, the other half did the no experience session first). (see Figure 5). Subsequently, we wanted to find out if there was a difference in performance between dogs after session 1 and after session 2, regardless of whether they did the experience-phase or the no experience-phase first.

Based on the results of the pilot study, in this main study, we increased the experience-phase from eight to twelve, because in the pilot study, after 8 interactions, they still needed 6-10 trials to significantly favor the generous person. We argued that an increased experience-phase would increase performance but if you increased it too much, dogs could lose attention and become bored. We also wanted to allow the dogs to form an opinion more quickly during the trials.

Furthermore, in Jim et al. (2021, 2022), the animals had six choice trials, which yielded non-significant results. Based on the results of the pilot study (Block 6-10 was significant), we gave them 10 trials, plus two more to allow them to make mistakes while choosing.

Session 1 (experience session):



After 5-21 days

Session 2 (no experience session):

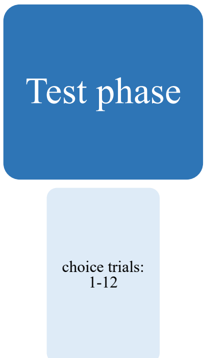


Figure 5. An example of the experimental procedure of the main study.

2.3.4 Experimental procedure

The procedure was very similar to the pilot study. The starting position of the owner (P1 or P2) was counterbalanced between subjects.

In the experience session, the procedure consisted of two parts:

1. Experience-phase: The experience-phase was identical to the Pilot Study, except the dog interacted with each partner three times on each side (three times on P1 and three times on P2) and had therefore six interactions per side and a total of 12 interactions.

2. Test phase: The procedure was the same as in the pilot study, but the dog experienced 12 trials.

In the no experience session, the procedure only consisted of the test phase.

2.4 Behavioral analysis

The videos were uploaded to Loopy by Loopbio, a software to store, label, and code experimental videos. We coded the wide-angle video footage of the enclosure, which had been recorded using a GoPro Hero 4 Black. We coded each trial for each dog separately, starting the trial when the experimenter said "Okay!" and stopping it when the experimenter said "Stop!" Inside each trial we counted different factors. First, we coded the first choice of the dog in each trial, which was binomial (generous/selfish). If the dog did not make a choice within 15 seconds, it was coded as NA (no choice). Furthermore, we coded the time dogs spent in proximity towards a partner (see Table 3 for definitions).

Table 3. Definitions of coded behaviors.

Behavior	Definition
Choice (Event)	• The dog's two paws were within or on the circle and it was looking at the partner. Some dogs were hesitant to approach the unfamiliar people, therefore if the dog walked towards and gazed at a partner for 2 seconds, she reacted and this was also considered a choice • If there was a reaction due to a lack of attention from the partners, even if the criteria were not fully met, it counted as a choice. If there was no reaction from the partner it did not count as a choice (NA).
Proximity (Duration)	• When the generous partner fed the dog or when the dog was touching the partner • When the dog's two paws were on or within the 50 cm circle it was in proximity • If the dog moved outside of the circle or only one paw stayed inside the circle after the choice but still stayed stationary near the partner until it moved

	• If the dog stepped out of the circle but continued to look at the partner until it looked away for more than a second. • When the dog started jumping and its paws left the circle, but it was clearly interacting with the partner • If the dog was only briefly passing the partner, possibly even in the circle, without looking at her, it did not count as proximity

2.5 Statistical analysis

To test whether the dogs of different ages learnt to choose the generous partner significantly more often than the selfish partner, we conducted a Generalized Linear Mixed-Effects Model (GLMM) with binomial error structure and logit link function (function glmer of the package lme4). Covariates (session and trial) were z-transformed for all models.

The response variable was first choice (generous/selfish). The test predictors were age category (factor with three levels: early, middle and old), session (1 or 2), and trial (1-12). Therefore, the full model included an interaction between age category × session × trial. Subject ID was added as a random intercept and an interaction between session and trial was added as a random slope within subject ID. To ease convergence, we changed the optimizer used by the function glmer to "bobyqa" (Jacobson, n.d.). Then, we compared the full model to the null model, which lacked the interaction between age category × session × trial. The model was stable (see the range of estimates in Table 4) and there was no collinearity (max VIF = 1.000).

Secondly, we tested if dogs spent more time close to the generous partner or to the selfish partner. We used a Generalized Linear Model (GLM) with beta error distribution and logit link function (function betareg of the package betareg). The response variable was proportion of time spent in proximity to the generous partner (values above 0.5 indicate bias towards generous person and values below 0.5 indicate bias towards the selfish person). The test predictors were again age category (factor with three levels: early, middle and old), session (1 or 2), and trial (1-12). Therefore, the full model included an interaction between age category × session × trial. Subject ID was added as a random intercept and an interaction between session and trial was added as a random slope within subject ID. We excluded the correlation between the random intercept and slope because it was estimated to be essentially 1, which is indicative of it not being identifiable (Matuschek et al., 2017). Then, we compared the full model to the null model, which lacked the interaction between age category × session × trial. The model was very stable (see the range of estimates in Table 5) and there was no collinearity (max VIF = 1.001).

3 Results

The likelihood ratio test of the binomial model comparing the full to null model shows that the interaction between the test predictors *age category x session x trial* had no significant effect on whether the dogs chose more often the generous or the selfish partner ($\chi^2 = 6.669$, $df = 11$, $p = 0.825$). The model was stable (see the range of estimates in Table 4) and there was no collinearity (max VIF = 1.000).

Table 4. Results of the binomial model (estimates, standard errors, confidence intervals CIs, z-value, *p*-value and the estimates obtained for model stability minimum and maximum).

Term	Estimate	SE	95% CI		z	*p*	Min	Max
			Upper	Lower				
Intercept	-0.122	0.158	0.221	-0.462	-0.733	0.439	-0.191	-0.033
Age category: middle[1]	-0.102	0.208	0.312	-0.530	-0.491	0.624	-0.189	-0.033
Age category: old[1]	-0.048	0.211	0.403	-0.471	-0.226	0.821	-0.145	0.021
z-transformed session	-0.115	0.131	0.168	-0.373	-0.881	0.378	-0.197	-0.043
z-transformed trial	-0.067	0.130	0.219	-0.337	-0.515	0.607	-0.116	-0.020
Age category: middle x z-transformed session	0.151	0.172	0.459	-0.235	0.877	0.380	0.079	0.234
Age category: old x z-transformed session	0.130	0.175	0.464	-0.203	0.742	0.458	0.059	0.213
Age category: middle x z-transformed trial	0.044	0.171	0.374	-0.295	0.257	0.798	-0.003	0.104

Age category: old x z-transformed trial	-0.002	0.174	0.354	-0.341	-0.011	0.991	-0.048	0.054
z-transformed session x z-transformed trial	-0.145	0.130	0.123	-0.444	-1.118	0.264	-0.190	-0.087
Age category: middle x z-transformed session x z-transformed trial	0.233	0.171	0.609	-0.104	1.363	0.173	0.175	0.276
Age category: old x z-transformed session x z-transformed trial	0.344	0.175	0.716	-0.007	1.969	0.049	0.267	0.393

Estimate, standard error, confidence intervals, results of significance tests (Wald's z approximation) and the range of estimates derived after excluding individuals one at a time

[1] Age category: early as reference level

The results of the beta regression model comparing full to null model with the interaction between the test predictors *age category x session x trial* show that the proportion of time spent with the generous partner was non-significant ($\chi^2 = 9.855$, $df = 11$, $p = 0.544$, Figure 6), so they did not significantly spend more time close to the generous partner than the selfish partner. The model was stable (see the range of estimates in Table 5) and there was no collinearity (max VIF = 1.001).

Table 5. Results of the beta regression model (estimates, standard errors, confidence intervals Cis, z-value, *p*-value and the estimates obtained for model stability minimum and maximum).

Term	**Estimate**	**SE**	**95% CI**		**z**	***p***	**Min**	**Max**
			Upper	**Lower**				
Intercept	0.246	0.109	0.430	0.020	2.261	0.024	0.187	0.304
Age category: middle[1]	0.139	0.143	0.399	-0.122	0.977	0.328	0.078	0.195

Age category: old[1]	0.039	0.145	0.320	-0.245	0.265	0.791	-0.022	0.115
z-transformed session	0.106	0.092	0.283	-0.076	1.151	0.250	0.078	0.134
z-transformed trial	0.061	0.091	0.232	-0.110	0.667	0.505	0.037	0.095
Age category: middle x z-transformed session	0.039	0.121	0.257	-0.188	0.326	0.745	-0.003	0.085
Age category: old x z-transformed session	-0.000	0.124	0.246	-0.242	-0.001	1.000	-0.029	0.031
Age category: middle x z-transformed trial	-0.000	0.121	0.234	-0.242	0.000	1.000	-0.043	0.023
Age category: old x z-transformed trial	0.006	0.123	0.247	-0.235	0.048	0.962	-0.028	0.035
z-transformed session x z-transformed trial	0.050	0.091	0.220	-0.113	0.547	0.584	0.008	0.074
Age category: middle x z-transformed session x z-transformed trial	-0.086	0.122	0.139	-0.318	0.704	0.482	-0.109	-0.044
Age category: old x z-transformed session x z-transformed trial	-0.108	0.123	0.121	-0.341	-0.873	0.383	-0.136	-0.066

Estimate, standard error, confidence intervals, results of significance tests (Wald's z approximation) and the range of estimates derived after excluding individuals one at a time

[1] Age category: early as reference level

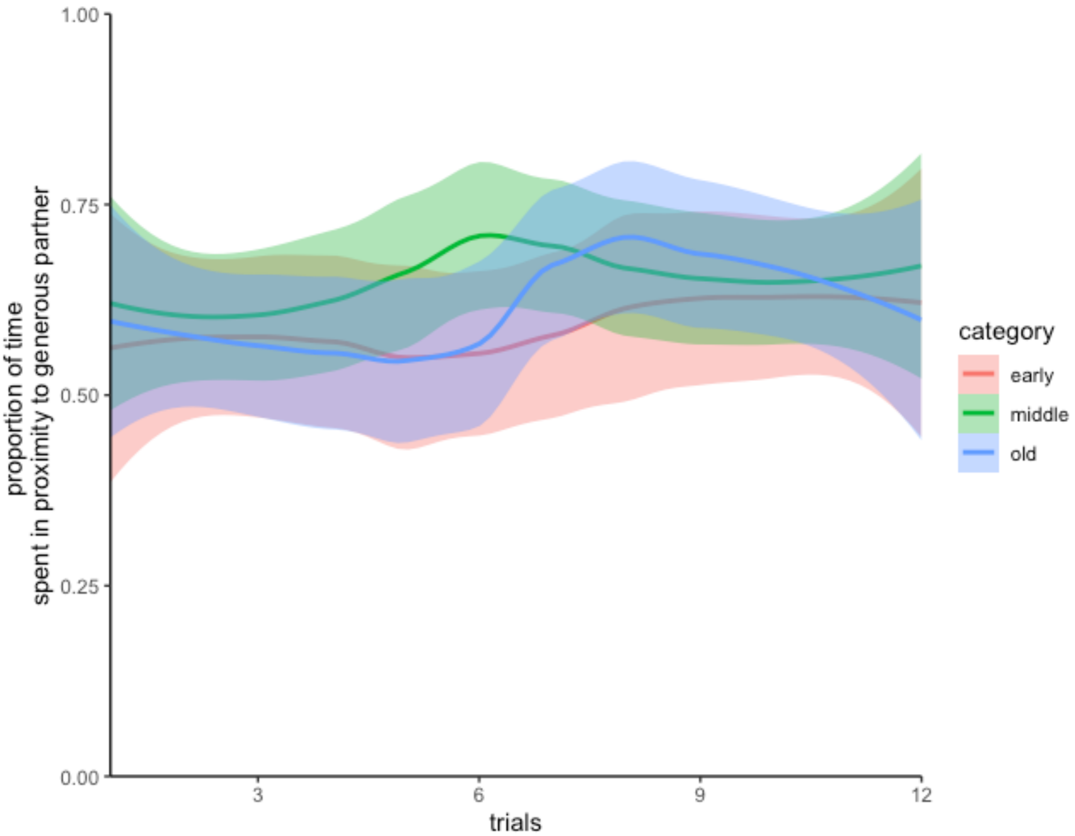


Figure 6. Proportion of time spent in proximity to the generous person (within 15 seconds) in relation to the age categories (early, middle, old) in trials 1 to 12.

As the results of the GLMMs were not significant, we furthermore conducted exact binomial tests to test whether the dogs showed a side or color bias. For the side bias, we counted the number of left choices as success. The partners' role and color of clothes were randomized between and fixed within subjects, so we counted the number of times the dogs chose the black dressed person as success for color bias. 0.5 was the hypothesized probability of success for both.

The results showed that 19 out of 40 dogs showed a significant side bias for the left side (see Table 6). Only five out of 40 dogs showed a color bias (see Table 6), thus they did not show a color bias overall.

Table 6. Color and side bias of all study dogs.

		Color bias		Side bias	
Subject	**Number of trials**	**Number of successes (black)**	***p***	**Number of successes (left)**	***p***
Aaron	24	13	0.839	2	**< 0.001**
Aidan	27	18	0.122	13	1
Ajani	27	14	1	9	0.122
Alois	25	13	1	2	**< 0.001**
Amy	24	17	0.064	18	**0.023**
Asha	27	15	0.701	13	1
Bailey	27	8	0.052	8	0.052
Capper	27	11	0.442	8	0.052
Chivas	19	10	1	3	**0.004**
Cody	26	9	0.169	12	0.845
Crash	25	14	0.690	3	**< 0.001**
Daytona	24	16	0.152	7	0.064
Django	24	11	0.839	6	**0.023**
Dunni	26	9	0.169	4	**0.001**
Ellis	27	12	0.701	7	**0.019**
Fiona	26	16	0.327	4	**0.001**
Franzi	26	13	1	7	**0.029**
Gandalf	27	10	0.248	14	1
Helena	27	13	1	0	**< 0.001**
Holly	27	16	0.442	21	**0.006**
Jazzie	27	17	0.248	4	**< 0.001**
Jolie	26	12	0.845	18	0.076
Kiki	20	15	**0.041**	11	0.824
Lenny	26	7	**0.029**	11	0.557
Lilly	27	11	0.442	6	**0.006**
Lilu	27	10	0.248	18	0.122
Monty	27	10	0.248	11	0.442
Mozart	21	7	0.804	13	0.383
Ozzy	27	14	1	26	**< 0.001**
Ravenna	27	13	1	12	0.701
Rusty	27	9	0.122	6	**0.006**
Sally	26	11	0.557	14	0.845
Shari	21	7	**0.029**	13	0.383
Sixtus	25	14	0.690	10	0.424
Snoopy1	25	9	0.230	4	**0.001**
Snoopy2	25	7	**0.043**	10	0.424
Spaiky	23	12	1	14	0.405
Timo	22	9	0.524	17	**0.017**

Ultimo	26	14	0.845	24	**< 0.001**
Zazu	26	6	**0.009**	10	0.327

In Table 7 and Figure 7, you can see how often the dogs with a side bias selected the left or right circle and how often those without a side bias selected the left or right circle.

Table 7. Side biased and non-side biased dogs and their side choices.

	Left	**Right**	**Total number of trials**
Side bias (N = 19)	142	283	425
No side bias (N = 21)	218	258	476

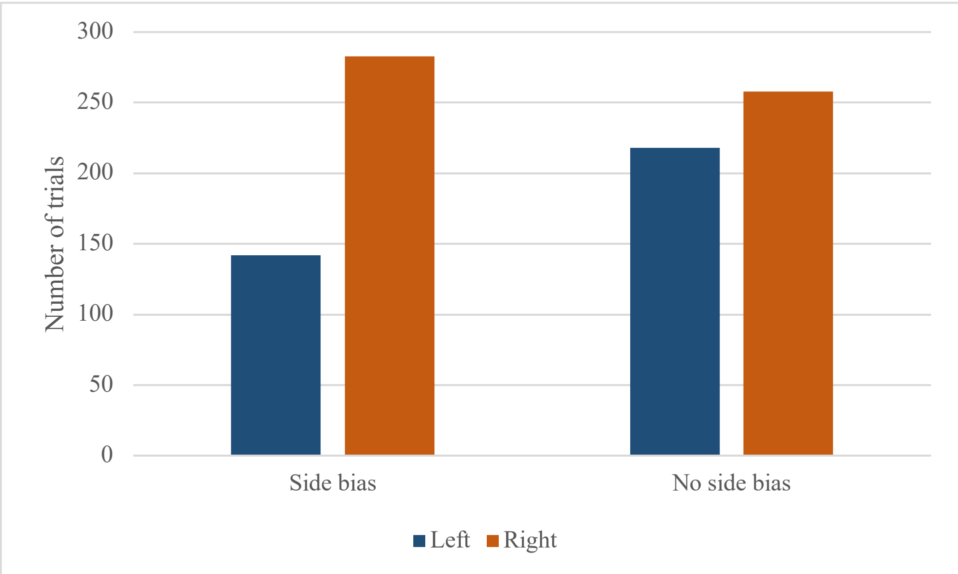


Figure 7. Side choices of side biased dogs and non-side biased dogs in experience and no experience-phase.

Therefore, we excluded the dogs with a side bias and ran the GLMMs again. For the binomial model again, the comparison of the full to the null model with the same test predictors as before was not significant ($\chi^2 = 10.423$, $df = 11$, $p = 0.493$). The model was stable (see the range of estimates in Table 8) and there was no collinearity (max VIF = 1.001).

Table 8. Results of the non-side biased dogs' binomial model (estimates, standard errors, confidence intervals CIs, z-value, *p*-value and the estimates obtained for model stability minimum and maximum).

Term	**Estimate**	**SE**	**95% *CI***		***z***	***p***	**Min**	**Max**
			Upper	**Lower**				
Intercept	-0.342	0.285	0.305	-1.017	-1.198	0.231	-0.479	-0.172
Age category: middle[1]	0.078	0.348	0.831	-0.659	0.225	0.822	-0.088	0.213
Age category: old[1]	0.185	0.389	1.051	-0.617	0.477	0.634	-0.098	0.379
z-transformed session	-0.148	0.206	0.279	-0.638	-0.720	0.472	-0.293	-0.037
z-transformed trial	-0.060	0.208	0.342	-0.518	-0.291	0.771	-0.184	0.065
Age category: middle x z-transformed session	0.081	0.251	0.631	-0.404	0.322	0.748	-0.033	0.224
Age category: old x z-transformed session	0.116	0.282	0.707	-0.469	0.410	0.682	-0.057	0.260
Age category: middle x z-transformed trial	0.037	0.253	0.641	-0.463	0.145	0.885	-0.087	0.156
Age category: old x z-transformed trial	-0.203	0.287	0.398	-0.802	-0.705	0.481	-0.327	-0.077
z-transformed session x z-transformed trial	-0.284	0.207	0.153	-0.837	-1.369	0.171	-0.444	-0.162

Age category: middle x z-transformed session x z-transformed trial	0.419	0.253	1.098	-0.122	1.658	0.097	0.300	0.575
Age category: old x z-transformed session x z-transformed trial	0.687	0.287	1.447	-0.198	2.392	0.017	0.508	0.843

Estimate, standard error, confidence intervals, results of significance tests (Wald's z approximation) and the range of estimates derived after excluding individuals one at a time

[1] Age category: early as reference level

Also, the beta regression model for the non-side biased dogs was non-significant, as shown in the following table (Table 9) (χ^2 = 18.551, df = 11, p = 0.070). The model was stable (see the range of estimates in Table 9) and there was no collinearity (max VIF = 1.003).

Table 9. Results of the non-side biased dogs' beta regression model (estimates, standard errors, confidence intervals CIs, z-value, *p*-value and the estimates obtained for model stability minimum and maximum).

Term	**Estimate**	**SE**	**95% CI**		**z**	***p***	**Min**	**Max**
			Upper	**Lower**				
Intercept	0.296	0.164	0.605	-0.035	1.805	0,071	0.150	0.381
Age category: middle[1]	0.153	0.202	0.554	-0.218	0.758	0.449	0.075	0.293
Age category: old[1]	-0.062	0.225	0.366	-0.503	-0.283	0.777	-0.180	0.111
z-transformed session	0.143	0.133	0.409	-0.123	1.073	0.283	0.081	0.221
z-transformed trial	0.058	0.130	0.305	-0.184	0.448	0.654	0.001	0.143
Age category: middle x z-transformed session	0.167	0.164	0.475	-0.161	1.016	0.310	0.086	0.233

Age category: old x z-transformed session	0.016	0.184	0.384	-0.348	0.086	0.932	-0.064	0.080
Age category: middle x z-transformed trial	0.035	0.163	0.341	-0.277	0.216	0.829	-0.050	0.093
Age category: old x z-transformed trial	0.157	0.183	0.487	-0.186	0.856	0.392	0.071	0.216
z-transformed session x z-transformed trial	0.084	0.135	0.334	-0.175	0.642	0.521	-0.014	0.149
Age category: middle x z-transformed session x z-transformed trial	-0.111	0.163	0.183	-0.429	-0.682	0.495	-0.177	-0.014
Age category: old x z-transformed session x z-transformed trial	-0.229	0.183	0.127	-0.591	-1.252	0.211	-0.299	-0.132

Estimate, standard error, confidence intervals, results of significance tests (Wald's z approximation) and the range of estimates derived after excluding individuals one at a time

[1] Age category: early as reference level

4 Discussion

The aim of the study was to find out whether pet dogs could form reputations of people after direct interaction regarding their social role in a begging situation, where the generous person fed the dog and the selfish person withheld food from the dog. Our main finding is that dogs did not choose the generous partner more often than the selfish partner. Even with more experience, it was not possible for dogs to discriminate between two people. Additionally, they did not spend more time with the generous partner than the selfish one. Therefore, we failed to find evidence for our hypothesis that dogs can form direct reputations. Furthermore, we wanted to find out if older dogs (with more experience interacting with humans) would perform better than less experienced dogs in a begging situation, but we did not find any significant results. Our results are quite surprising because it is advantageous for animals to predict others' future behavior.

We predicted that especially dogs that live close to, or with humans, should be able to form reputation-like interferences about humans because they depend on humans for food and other resources. Additionally, we predicted that they would choose the generous partner more often than the selfish one, as previously demonstrated in Carballo et al. (2015). We may not have found this result because the dogs did not really get to know the two partners in our study. Although the partners in Carballo et al. (2015) were also unknown to the dogs, it could be argued that they were able to learn more about the partners during the study. This might be due to the fact that Carballo et al.'s (2015) study consisted of two training sessions, which in each case included six interactions with the generous partner and six interactions with the selfish partner, which sums up to 12 interactions per partner in total. After the first interaction the dogs had one choice-test and after the second interaction too. In our study, we only had 12 interactions in total, which means only six per partner, which is just half of what Carballo et al. (2015) did, and they were followed by many choice-tests based on trial and error. Carballo et al. (2015) found that even though the dogs did not prefer the generous person in the first choice-test they had a preference in the second choice-test. This could be because the dogs need more direct interactions to be able to form a reputation because they learn more from direct interactions than from choice trials. However, this statement contrasts with the fact that the 19 dogs in our pilot study were able to form an opinion after only eight interactions, four per partner, and after six to ten trials they significantly preferred the generous person. Since the pilot study was meant to be a predecessor of the main study and to give us insights into the adaptations of the study, it would not make sense here to compare the pilot and the main study directly with each other. Also because of the different experimental design, with more direct interactions in the main study, but more choice-tests in the pilot study and two instead of only one session in the main study. Finally, not the entire pilot study was significant, only the block consisting of trials 6 to 10 was significant, suggesting that dogs cannot maintain the ability to form opinions over all 30 trials. Since we included this number in the main study, it could have been assumed that dogs could maintain the ability for this time, but this was not confirmed here.

Why a similar setting was significant in the pilot study and not in the main study could be due to the individuality of the dogs but also to the weather. The pilot study was conducted in spring, and half of the dogs were tested in pouring rain with a temperature of about 12 degrees Celsius, whereas the main study took place in high summer, when it was sometimes 30 degrees Celsius or even more. Most of the enclosure had no shade, and only near one of the two circles was there shade due to a tree. According to a study by Hall et al. (2021), which investigated the activity level and also a dogs will to exercise, dogs like the cold and ice less than rain and are less active in hot weather than in cooler weather. However, this depends on gender, the weather, breed, continent, daily activity and much more, and it is not generalizable. As for the different sides, we found that 19 out of 40 dogs showed a clear preference for one of the two sides, i.e., for the right side. There could be the same reason as for why they significantly preferred the generous person in the pilot study but not in the main study, which was the fact that there was a tree closer to P2, which provided shade on hot days and protection from the rain on rainy days.

Another possible reason for the side preference might be what Pongrácz et al. (2003) found out. When dogs do not get a demonstration, they tend to choose randomly one side which to approach, and this initial experience influences their following choices. This might be especially the case for the dogs that started with the no experience session.

Subiaul et al. (2008) conducted an experiment with chimpanzees, where seven chimpanzees received extensive training with two already familiar partners. They were tested in blocks of 8 criterion trials and this was repeated until they managed to choose the generous person in 7 out of 8 cases in two consecutive blocks. The four chimpanzees that reached this criterion were presented with five pairs of unknown partners in succession, each of which had 8 trials again. The aim was to see whether the chimpanzees that had learned to form opinions could also apply this to new situations. This leads to the question of whether dogs should first be accustomed to the type of test and the reaction of people by more extensive training and a specific criterion in a first session, in order to then be able to transfer this experience to the second session, or a new situation. As it turns out, reputation formation is not an easy task and may need to be learnt by the individual before it can be successfully applied. In our study, the dogs did not improve on the second set of trials, which could also be due to the fact that in contrast to Subiaul et al.'s study, we did not have a similar criterion at which the dogs could advance to the next session.

Dogs might have been more successful, if we would have provided more cues to them, such as looking at the dogs, talking to the dogs in different pitches or smiling at the dogs according to their role outside of a situation where the dogs had to decide or were interacting because generally, outside the direct experience or the dog's choices, we tried to eliminate at least most of the auditory and visual cues to create a setting as neutral as possible, when in real life a generous person would also act according to their role when the dog is only close. While the dog had direct experience with the partners, they looked at the dog when it approached, reacted with different body language and also said different sentences in different pitches. The same happened when approaching during a choice-test. Maybe these short signals were not enough to show the dog which role the people had and often the voice pitches were too similar. In addition, the partners also wore clothes with contrasting colors, which were counterbalanced between subjects, which means that the generous person wore black half of the time and white the other half. The study was mainly conducted during summer, so the partners often wore black shorts, which they brought themselves, or long white trousers, which were provided by us. This could also have confused the dogs and caused them not to consistently select the same color during the sessions because the black color was less noticeable on the subjects with white skin color and so there may not have been much contrast between black and white. Despite these aggravations, we did not find a significant preference for one color in total, but the dogs might have found it more difficult to distinguish between the two colors. Given that it has been found that dogs can distinguish between human faces expressing different emotional states (Müller et al., 2015), recognize familiar and non-familiar faces (Adachi et al., 2007), and also refer to faces as an important source of information in socially relevant situations (Gácsi et al., 2004), one could conclude that dogs may not pay too much attention to people's clothes or bodies, but much more to their faces, although there is not enough evidence for this yet. The importance of the faces could also be explained by the fact that when the dogs were near the partners, they often looked them directly in the face and thus demanded treats. One reason why the dogs could not distinguish between the faces could be that all partners were white females who might have looked relatively similar to the dogs.

Another factor could be that dogs learned to live with humans and therefore, they formed a specialization for communication with them (Bräuer et al., 2006, Reid, 2009). It is possible, that dogs do not care for the role of the partner in such a context, but just want to interact with them, which could lead to our results that dogs cannot differentiate between people. Also, dogs' lack of attention during the experience-phase might have led to their inability to distinguish between the two partners in the choice trials. Unfortunately, this was not coded during testing, which is why it can only count as hypothetical.

Since we had to rely on volunteers to act as partners, we had to use 5 different pairs of partners, which could have influenced the dogs' choice, since there were 10 individuals with different smells and they could have preferred certain individuals. This could have been avoided if we had always had the same two individuals as partners in the experience-phase, i.e., partners A and B, and also in the no experience-phase, i.e., partners C and D. Additionally, this would have given us a possibility to control for any individual differences, such as the consistent preference of one of the two individuals regardless of their role by the dogs. Another view would be to use a new team of partners per dog to eliminate individual preferences as much as possible.

One could argue that the sample size was too small. Carballo et al. (2017) tested 47 dogs to achieve significant results, but also Carballo et al. (2015), which tested only 19 dogs, found that dogs could distinguish between generous and selfish partners. Heberlein et al. (2016) tested only 13 dogs and eight wolves and had significant results, which assumes discrimination between individuals. Since there are differences in the number of the dogs to be tested in the literature, it cannot be clearly stated whether an increase in the sample size would have led to a significant result, i.e., that our sample size was too small to draw relevant conclusions.

It could also be that the dogs tested in the main study were not capable to distinguish between the two partners after only six demonstrations each or that 12 trials with these partners and 12 trials with different partners in another session were still too few for them to enable a distinction to be made. Nevertheless, an increased experience-phase, where the dogs could interact with the generous and selfish partner more times could have led to a significant outcome and also to a reputation formation that is maintained during all trials. In Carballo et al. (2015), as mentioned above, a double number of interactions, i.e., 24 interactions, during the experience-phase led to significant results. However, one could also assume that it is very difficult for animals to form an opinion because, as Subiaul et al. (2008) found out, five out of seven chimpanzees needed 15-75 experiences to show a preference for the generous one and still one of them could not maintain that preference. This also means that the range of the number of experiences is very large and varies greatly from individual to individual, which is why it is very difficult to make general statements here.

Another possible reason for the dogs not to choose the generous person more often could be that they only paid attention to the food in the hand of the partner. Nitzschner et al. (2012) tested reputation formation in dogs and did not find a preference for the generous person in a pilot study and argued that the dogs were not focused on the partners' actions when food was involved. If the dogs only paid attention to the hand with the food in it, then they could not distinguish the partners in the test phase in our study because both the generous and selfish partners had food in their hands. Carballo et al. (2015) argued that hands were apparently an insufficient cue to discriminate experimenters in their choice-test because for dogs that only paid attention to the hand, it could have overshadowed the other features that would have been really helpful to distinguish. In the future, it would be interesting to also record what the dogs' attention is focused on, but if they don't lose their focus on the food, it might be really difficult for them to notice what is going on around them.

Additionally, we wanted to find out if the dogs spent more time close to the generous partner than the selfish partner. Nitzschner et al. (2012) found out that dogs spent more time with the generous partner than the selfish partner, but we could not replicate this finding. This might be because after choosing the generous person, they tried to get another treat, so they approached the selfish person. Another reason could be that we conducted the study in the outside enclosure of the CDL and the weather sometimes was really hot, so dogs did not spend time with anyone because they preferred to lie in the shade and unfortunately, a major part of both circles was in the sun. Furthermore, they could explore the enclosure during the test trials rather than interact with the partners. We could see that some dogs sniffed the whole enclosure during the trials and were more interested in their surroundings than the humans. Nevertheless, before turning their attention to their surroundings, the dogs tried to get a treat and approached one of the partners, which explains the choices.

4.1 Age effect

We wanted to know whether dogs of different age groups differed in their ability to form direct reputations. We hypothesized that younger dogs would perform worse than older dogs due to their lack of experience with humans. Carballo et al. (2017) found that adult dogs, both family dogs and shelter dogs, preferred the generous person, whereas puppies did not. However we found that no age group formed reputations. Udell et al. (2008) and Gácsi et al. (2009) showed that wolves, from ages 3 to 9, and pet dogs, from ages 0.5 to 15, with comparable conditions outperformed shelter dogs, from ages 0.5 to 3.5, which suggests that experience with humans and socialization is important. One could argue that shelter dogs had a disadvantage due to the lack of contact with humans and their younger age compared to pet dogs and wolves.

4.2 Conclusion and outlook of the future

To conclude, it must be said that even though we did not find any significant results this time, it is highly unlikely that dogs cannot form reputations of humans after directly interacting with them, as many different studies show that they can (e.g.,Nitzschner et al., 2012, Heberlein et al., 2016, Carballo et al., 2015, Carballo et al., 2017). Therefore, the results of this study are surprising. However there are some studies that did not find that wolves or dogs could form direct reputations (e.g., Piotti et al., 2017, McGetrick et al., 2021, Jim et al., 2022). Now the question arises if dogs can even form direct reputations or if it is hard for them, and it is just as possible after many repeated interactions. The studies in which dogs were able to form a reputation differed greatly in the number of trials the dogs had. For example, in Nitzschner et al. (2012) dogs participated in seven experience trials per partner and four experimental trials, where they had a choice. In Heberlein et al. (2016), the dogs had three training sessions within three days. Each consisted of three interactions per partner. Furthermore, they had a single preference test and two preference test blocks consisting of four trials each. In addition, they had two test phases, consisting of two tests per partner. Furthermore, in Carballo et al. (2015), the dogs had two training blocks with 12 trial sessions (six per partner) and two choice-tests and in Carballo et al. (2017) they had two pre-training trials, two training sessions, each consisting of six interactions per partner and two choice-tests in total, so one after each training session. In the studies in which it was not possible for dogs to form an opinion, Piotti et al. (2017) used four blocks of two demonstration types (skillful vs. unskillful) followed by one trial each in their first study and modified their procedure using only one type of demonstration (skillful or unskillful) and one unsolvable task trial in study 2. In McGetrick et al. (2021) the dogs experienced two experience-phase sessions, each consisting of five sessions and each session consisted of ten trials. Afterwards, they had a test phase with three conditions, which each consisted of 20 trials. The dogs and wolves in Jim et al. (2022) had an experience-phase, where they interacted four times with each partner, followed by six test trials. So it is apparent that our pilot study was very similar to the design of Jim et al. (2022), except that we had 30 test trials, but showed that dogs are very capable of forming reputations. Unfortunately, this is no longer true for the main study.

Since the weather definitely played a role in this study, an indoor lab would therefore also be more appropriate. Especially too hot weather limited the dogs' activity. Unfortunately, we did not look at the differences in performance of the dogs tested in pleasant weather versus those tested in hot weather, this could be explored further down the line.

This study only tested for a positivity bias but some studies (Hamlin et al. 2007, Hamlin et al. 2010), especially on humans, used a neutral agent to distinguish between negativity and positivity bias. Chijiiwa et al. (2015) even found a negativity bias on dogs in an eavesdropping condition. In an improved version of this study, where discrimination may even be visible, it would be interesting to include a neutral agent to see if dogs approach it even though it has not interacted with them. As can be seen from the pilot study, dogs are capable of discriminating against people based on their role.

Therefore, it is important to continue testing under different circumstances. First, the sample size should be much bigger than ours and secondly, the testing area should be without external disturbances and without weather influence, which is why the study could be repeated in an indoor lab. Another aspect that could be changed in future studies is that the partners should either always be the same for all dogs to control for individual differences or a different pair for each dog to exclude individual differences in large part.

5 References

Abdai, J., Miklósi, Á. (2016): “The origin of social evaluation, social eavesdropping, reputation formation, image scoring or what you will.” *Frontiers in Psychology*. 7(1772. DOI:10.3389/fpsyg.2016.01772.

Adachi, I., Kuwahata, H., Fujita, K. (2007): “Dogs recall their owner’s face upon hearing their owner’s voice.” *Animal Cognition.* 10:17–21. DOI:10.1007/s10071-006-0025-8.

Anderson, J.R., Kuroshima, H., Takimoto, A., Fujita, K. (2013): ”Third-party social evaluation of humans by monkeys.” *Nature Communications*. 4. DOI:10.1038/ncomms2495.

Boesch, C., Boesch-Achermann, H. (2000): “The chimpanzees of the Taï-Forest: behavioral ecology and evolution.” Oxford University Press.

Botigué, L.R., Song, S., Scheu, A., Gopalan, S., Pendleton, A.L., Oetjens, M., et al. (2017): “Ancient european dog genomes reveal continuity since the early Neolithic.“ *Nature communications*. 8. DOI:10.1038/ncomms16082.

Bräuer, J., Kaminski, J., Riedel, J., Call, J., Tomasello, M. (2006): “Making inferences about the location of hidden food: social dog, causal ape.“ *Journal of Comparative Psychology*. 120(1):38-47. DOI:10.1037/0735-7036.120.1.38.

Carballo, F., Freidin, E., Putrino, N., Shimabukuro, C., Casanave, E., Bentosela, M. (2015): “Dog’s discrimination of human selfish and generous attitudes: the role of individual recognition, experience, and experimenters’ gender.“ *PloS ONE*. DOI:10.1371/journal.pone.0116314.

Carballo, F., Freidin, E., Casanave, E.B., Bentosela, M. (2017): “Dogs’ recognition of human selfish and generous attitudes requires little but critical experience with people.” *PloS ONE*. DOI:10.1371/journal.pone.0185696.

Chapagain, D., Range, F., Huber, L., Virányi, Z. (2017): “Cognitive aging in dogs.“ *Gerontology*. 64:165 - 171. DOI: 10.1159/000481621.

Chijiiwa, H., Kuroshima, H., Hori, Y., Anderson, J.R., Fujita, K. (2015): “Dogs avoid people who behave negatively to their owner: third-party affective evaluation.” *Animal Behaviour*. 106:123–127. DOI:10.1016/j.anbehav.2015.05.018.

Chijiiwa, H., Takagi, S., Arahori, M., Hori, Y., Saito, A., Kuroshima, H., Fujita, K. (2020): “Dogs and cats prioritize human action: choosing a now-empty instead of a still-baited container.” *Animal Cognition*. 24:65–73. DOI:10.1007/s10071-020-01416-w.

Dahl, A., Schuck, R.K., and Campos, J.J. (2013): “Do young toddlers act on their social preferences?“ *Dev. Psychol.* 49:1964–1970. DOI: 10.1037/a0031460.

de Waal, F. (1996): *Good Natured.* Cambridge (Massachusetts): Harvard University Press. 296.

Gácsi, M., Miklósi, Á., Varga, O., Tópal, J., Csányi, V. (2004): “Are readers of our face readers of our minds? Dogs (Canis familiaris) show situation-dependent recognition of human’s attention.” *Animal Cognition*. 7:144–153. DOI:10.1007/s10071-003-0205-8.

Gácsi, M., Györi, B., Virányi, Z., Kubinyi, E., Range, F.,Belényi, B. & Miklósi, Á. (2009): “Explaining dog wolf differences in utilizing human pointing gestures: selection for synergistic shifts in the development of some social skills.“ *Plos ONE.* 4. DOI: 10.1371/journal.pone.0006584.

Goodall, J. (1986): *The chimpanzees of Gombe*. Belknap Press, Cambridge.

Hall, E.J., Carter, A.J., Farnworth, M.J. (2021): “Exploring owner perceptions of the impacts of seasonal weather variations on canine activity and potential consequences for human-canine relationships.” *Animals.* 11(3302). DOI:.10.3390/ani11113302

Hamlin, J.K., Wynn, K., Bloom, P. (2007): “Social evaluation by preverbal infants.“ *Nature.* 450:557–559. DOI: 10.1038/nature06288.

Hamlin, J. K., Wynn, K., Bloom, P. (2010): “Three-month-olds show a negativity bias in their social evaluations.“ *Dev. Sci.* 13:923–929. DOI: 10.1111/j.1467- 7687.2010.00951.x.

Hamlin, J. K., Wynn, K. (2011): “Young infants prefer prosocial to antisocial others.“ *Cogn. Dev.* 26:30–39. DOI: 10.1016/j.cogdev.2010.09.001.

Hamlin, J. K., Wynn, K., Bloom, P., Mahajan, N. (2011): “How infants and toddlers react to antisocial others.“ *Proc. Natl. Acad. Sci. U.S.A.* 108:19931–19936. DOI: 10.1073/pnas.1110306108.

Hamlin, J.K. (2013): “Failed attempts to help and harm: intention versus outcome in preverbal infants’ social evaluations.“ *Cognition.* 128:451–474. DOI: 10.1016/j.cognition.2013.04.004.

Hamlin, J.K. (2014): “Context-dependent social evaluation in 4.5-month-old human infants: the role of domain-general versus domain-specific processes in the development of social evaluation.” *Frontiers in Psychology.* 5(614). DOI:10.3389/fpsyg.2014.00614.

Heberlein, M.T.E., Turner, D.C., Range, F., Virányi, Z. (2016): “A comparison between wolves, *Canis lupus*, and dogs, *Canis familiaris*, in showing behavior towards humans.“ *Animal Behavior.* 122:59–66. DOI:10.1016/j.anbehav.2016.09.023.

Heberlein, M.T.E., Manser, M.B., Turner, D.C. (2017): “Deceptive-like behaviour in dogs *(Canis familiaris)*.” *Animal Cognition*. 20(3):511–20. DOI: 10.1007/s10071-017-1078-6.

Herrmann, E., Keupp, S., Hare, B., Vaish, A. & Tomasello, M. (2013): “Direct and indirect reputation formation in nonhuman great apes (*Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pymaeus)* and human children *(Homo sapiens)*.“ *Journal of Comparative Psychology*. 127:63-75. DOI:10.1037/a0028929.

Jim, H.-L., Range, F., Marshall-Pescini, S., Dale, R., Plotnik, J.M. (2021): “Investigating indirect and direct reputation formation in Asian elephants *(Elephas maximus)*.” *Frontiers in Psychology.* 11(604372). DOI:10.3389/fpsyg.2020.604372.

Jim, H.-L., Plohovich, M., Marshall-Pescini, S., Range, F. (2022): “Wolves and dogs fail to form reputations of humans after indirect and direct experience in a food-giving situation.” *PLoS ONE.* 17(8). DOI:10.1371/journal.pone.0271590.

Leete, J., Vonk, J., Oriani, S., Eaton, T., Lieb, J. (2020): “Do domestic cats *(Felis silvestris catus)* infer reputation in humans after direct and indirect experience?” *Human-Animal Interaction Bulletin*. 8(3):35–53.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., Bates, D. (2017): “Balancing type I error and power in linear mixed models.” *Journal of Memory and Language*. 94:305–315. DOI:10.1016/j.jml.2017.01.001.

McGetrick, J., Poncet, L., Amann, M., Schullern-Schrattenhofen, J., Fux, L., Martínez, M., et al. (2021): “Dogs fail to reciprocate the receipt of food from a human in a food-giving task.” *PloS ONE.* 16(7). DOI:10.1371/journal.pone.0253277.

Muller, M.N., Mitani J.C. (2005): “Conflict and cooperation in wild chimpanzees.“ In: Slater, P.J.B., Rosenblatt, J., Snowdon, C., Roper, T., Naguib, M., editors: *Advances in the study of behavior*. New York: Elsevier. 275–331.

Müller, C.A., Schmitt, K., Barber, A.L., Huber, L. (2015): “Dogs can discriminate emotional expressions of human faces.” *CurrBiol.* 25(5):601–605. DOI:10.1016/j.cub.2014.12.055.

Nitzschner, M., Melis, A.P., Kaminski, J., Tomasello, M. (2012): “Dogs *(Canis familiaris)* evaluate humans on the basis of direct experiences only.” *PloS ONE*. 7(10). DOI:10.1371/journal.pone.0046880.

O’Connell, S.M. (1995): *Empathy in chimpanzees: Evidence for a theory of mind?* Primates 36: 397–410.

Piotti, P., Spooner, R.M., Jim, H.-L., Kaminski, J. (2017): “Who to ask for help? Do dogs form an opinion on humans based on skilfulness?” *Applied Animal Behaviour Science*. 195:93–102. DOI:10.1016/j.applanim.2017.05.024.

Pongrácz, P., Miklósi, A., Timár-Geng, K., Csányi, V. (2003): “Preference for copying unambiguous demonstrations in dogs.” *Journal of Comparative Psychology.* 117(3):337–343. DOI:10.1037/0735-7036.117.3.337

Preston, S., de Waal, F. (2002): “Empathy: Its ultimate and proximate bases.“ *Behav Brain Sci.* 25: 1–72.

R Core Team (2020): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rault, J.-L. (2019): “Be kind to others: prosocial behaviours and their implications for animal welfare.” *Applied Animal Behaviour Science.* 210:113–123. DOI:10.1016/j.applanim.2018.10.015.

Reid, P.J. (2009): “Adapting to the human world: dogs’ responsiveness to our social cues.” Behavioural Processes. 80(3):325–333. DOI:10.1016/j.beproc.2008.11.002.

Russell, Y.I., Call, J. & Dunbar, R.I. (2007): “Image scoring in great apes.“ *Behavioural Processes.* 78:108–111. DOI: 10.1016/j.beproc.2007.10.009.

Scarf, D., Imuta, K., Colombo, M., Hayne, H. (2012): “Social evaluation or simple association? Simple associations may explain moral reasoning in infants.” *PloS ONE.* 7(8). DOI:10.1371/journal.pone.0042698.

Schmuckler, M.A. (2001): “What is ecological validity? A dimensional analysis.” *Infancy*. 4(2):419–436. DOI:10.1207/S15327078IN0204_02.

Subiaul, F., Vonk, J., Okamoto-Barth, S., Barth, J. (2008): “Do chimpanzees learn reputation by observation? Evidence from direct and indirect experience with generous and selfish strangers.” *Animal Cognition.* 11:611–623. DOI:10.1007/s10071-008-0151-6.

Trösch, M., Bertin, E., Calandreau, L., Nowak, R., Lansade, L. (2020): “Unwilling or willing but unable: can horses interpret human actions as goal directed?” *Animal Cognition.* 23:1035–1040. DOI:10.1007/s10071-020-01396-x.

Udell, M.A.R., Dorey, N.R., Wynne, C.D. (2008): “Wolves outperform dogs in following human social cues.“ *Animal Behaviour*. 76:1767-1773. DOI:10.1016/j.anbehav.2008.07.028.

Wallis, L.J., Iotchev, I.B., Kubinyi, E. (2020): “Assertive, trainable and older dogs are perceived as more dominant in multi-dog households.” *PloS ONE.* 15(3). DOI:10.1371/journal.pone.0227253.

Watts, D., Mitani, J. (2002): “Hunting behavior of chimpanzees at Ngogo, Kibale National Park, Uganda.” *International Journal of Primatology.* 23(1). DOI:10.1023/A:1013270606320.

6 Appendix

6.1 Experimental adjustments and exceptions

There were special arrangements for some subjects. In the beginning of the study, we realized that not all dogs were comfortable with getting close to unknown persons. There were only ten dogs out of the 59 that completed the participation, which put two paws inside the circle when choosing a partner. This might be due to the fact that they are not comfortable with getting closer to an unfamiliar human. The other dogs stayed outside of the circle, and looked at the partner. This is why we made adjustments and defined also standing outside and looking at a partner as a choice. But only in cases where it was clear that the dog made a choice.

With the smaller dogs especially, we asked the partners to kneel in order to be closer to the dog. Some owners knew that their dogs found it easier to do the choice-tests or even the experience-phase off-leash. So we allowed that decision to be made by the owners as we assumed they knew their dogs best.