MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Direct reciprocity between familiar dogs?“

verfasst von / submitted by

Leona Fux, B.Sc.

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 UG2002
Betreut von / Supervisor:	Assoz.-Prof. PhD. Priv.-Doz. Friederike Range

Acknowledgement

I want to thank all the great people at the University of Vienna, the Veterinary University of Vienna, the Messerli Institute and beyond that helped me during my Master's thesis with their knowledge, constructive feedback and encouragement. Thank you, Friederike Range and Jim McGetrick, for your trust in me and for offering me the opportunity to study at the Clever Dog Lab under your supervision. And thank you, Thomas Bugnyar, for becoming my official mentor at the University of Vienna. My biggest thanks to you all, for inspiring me and showing me many ways to improve my work but also to be more self-confident about it.

Thanks also, Johannes Schullern, for our team work and funny jokes during stressful weeks between lockdowns and constantly changing safety regulations. Thank you, Karin Bayer, for always keeping a watchful eye on us and also thank you, Roger Mundry, for your help with any statistical questions. Of course, I also want to thank all the wonderful owners and their wonderful dogs to come with joy and motivation to every session and for being as curious as I am about canines' cognitive abilities.

In this way, I especially want to thank you, Alina Troniarsky and Kurt, for starting as my participants to turning into my precious friends. Huge thanks also to you, Johanna Gaudeck and Robert Kahlenberg, as well as to the rest of my friends and family for supporting me and constantly listening to my dog rambling. Last but not least, I want to thank you, Casca, for being a weird funky little dog that inspired me not only for my scientific interest and my career but also for my path in life.

cleverdoglab VIENNA messerli Foundation vetmeduni vienna

Table of Contents

1. Introduction

Darwin hypothesized that only behaviours that are useful to the individual's reproduction can be sustained during evolution. On the other hand, traits that do not benefit the individual lead to reduced fitness and therefore will not continue to spread in the population (Darwin, 1859). Cooperative behaviour can offer a mutual benefit to all parties and therefore can be explained by natural selection (Nowak, 2006). But natural selection failed to explain observed cooperative behaviours whereby only the receiving individual profits, while the other individual gets a seeming disadvantage. This is often called altruism and is defined as behaviour that either does not improve or even damages the reproductive success of the donor (Trivers, 1971).

So-called kin selection or inclusive fitness theory offered the first explanation for seemingly altruistic behaviour. By counting the survival of genes shared with relatives as an indirect way of increasing one's fitness, even cases in which it seems that helping another conspecific brings disadvantages to oneself, become plausible (Hamilton, 1964). But still, for almost another decade, the problem remained that many individuals cooperate altruistically although they are not close kin or sometimes not even from the same species (Trivers, 1971).

Reciprocal altruism added another part to the puzzle (Trivers, 1971). By starting an encounter as a cooperative individual but only helping those in the future that repaid the favour, stable forms of cooperation can be achieved. Simultaneously, the risk of exploitation by cheaters (uncooperative individuals) remains low, as one only offers help in the future if previous favours had been repaid (Axelrod & Hamilton, 1981). But difficulties in defining the term reciprocity remain to this day as some authors use different criteria and therefore come to different conclusions regarding the question of whether a non-human species is cognitively capable of reciprocity (Schweinfurth & Call, 2019). While strict limitations to only highly cognitively demanding processes of cooperation lead to the conclusion that reciprocity is rarely seen in species other than Homo sapiens, sticking to the more general definition of exchanging favours, reciprocity between animals is commonly observed (reviewed in Taborsky et al., 2016a).

In this thesis, it was assumed that reciprocity is defined as exchanging favours alternately (Schweinfurth & Call, 2019). According to this definition, three different forms of reciprocity (plus some additional special cases) are known which demand different levels of cognitive abilities (reviewed in Schweinfurth & Call, 2019). Generalised reciprocity is believed to be the simplest. Generalised reciprocity means that an individual that received help will be more likely to act helpfully but without constraining its favour specifically to the former donor. Without the need to keep track of every former encounter and without the need to identify separate individuals, this form is predicted to exist even in species with simpler cognitive abilities (Barta et al., 2011). Female but not male Norway rats (Rattus norvegicus) for example are known to reciprocate help in a generalized way (Rutte & Taborsky, 2007). Nonetheless, it is also found in Homo sapiens (Bartlett & DeSteno, 2006) the species that is believed to be the most advanced, cognitively (Melis & Semmann, 2010).

Secondly, direct reciprocity is defined as returning a favour to the same individual that offered help before (Trivers, 1971). Consequently, individuals must differentiate others to not confuse helpful with unhelpful partners and therefore minimize the risks of being exploited by cheaters (Pfeiffer et al., 2005). Additionally, it is often claimed that species have to keep track of every social interaction (reviewed in Carter, 2014). However, different cognitive mechanisms have been proposed which do not require an individual to recollect all former encounters. For example, by using the most recent experience, individuals might only build up an attitude instead of forming emotional bonds and without weighing offered help against the calculated amount they would expect in return. In many cases forming an attitude seems sufficient for direct reciprocity to occur while at the same time avoiding a large cognitive burden (Schweinfurth & Call, 2019). Even humans rely mostly on attitudinal mechanisms while using a more complex mechanism like weighing cost against gain only rarely with unfamiliar, unrelated individuals (Carter, 2014). Therefore, it is suggested that direct reciprocity is more common than previously thought (Schweinfurth & Call, 2019). So far, direct reciprocity is known to occur in some species, for example in Norway rats (Rutte & Taborsky, 2008; Zentall, 2016), cotton-top tamarins (Saguinus oedipus) (Hauser et al., 2003), Barbary macaques (Macaca sylvanus) (Molesti & Majolo, 2017) and common vampire bats (Desmodus rotundus) (Wilkinson et al., 2016).

The last form is indirect reciprocity. Without having personal experience, one decides to be cooperative by observing cooperative interactions between others and therefore evaluating how cooperative these potential social partners are (Alexander, 1987). Indirect reciprocity seems to be a rare phenomenon. So far, apart from humans, only a few examples like the cleaner wrasse (Labroides dimidiatus) (Bshary & D'Souza, 2005) and song sparrows (*Melospiza melodia*) (Akçay et al., 2010) are known to use it (Molesti & Majolo, 2017). But again, it is plausibly more widespread than currently thought. If it is enough to differentiate individuals and build up an attitude, it could be sufficient to only watch others' social encounters for forming an attitude and acting cooperatively according so (Schweinfurth & Call, 2019).

Studying reciprocity in different species helps to further enhance our understanding of cooperation in general. Interestingly though, only recently, scientists started to be interested in the cooperative abilities of humans' oldest known non-conspecific cooperative partner: dogs (Freedman & Wayne, 2017). Consequently, also reciprocity in dogs is a rather new scientific topic with only a few projects that have studied it directly so far. More studies have investigated behaviours that can be requirements for some forms of reciprocity. As reciprocity, regardless of its complexity, requires individuals to pay attention towards others, skills like social attention are very important so individuals can recognize cooperation opportunities (Schweinfurth & Call, 2019). Social attention means one has to pay attention towards others to gather information (Cohen, 1972). For dogs, gazing is often used to measure social attention and they are known to be very attentive towards conspecifics and even more so towards humans (Range et al., 2009). Furthermore, dogs are known to differentiate between conspecifics (Molnár et al., 2009) as well as to possess considerable cognitive skills like memory for socially learned tasks (Fugazza & Miklósi, 2014). Both, individual recognition and social memory, are necessary skills to reciprocate help not only in a generalized but also in a direct manner (Schweinfurth & Call, 2019). Also, dogs are known to act prosocially under certain conditions (Quervel-Chaumette et al., 2015, Dale et al., 2016) (although not all: see Dale et al., 2019). This means, dogs behave helpful towards conspecifics, which is an important prerequisite to test reciprocity in an experimental setting (McGetrick et al., 2021).

Consequently, this background knowledge about dogs' cognitive and social skills predicted direct and general reciprocity as theoretically possible (Gfrerer & Taborsky, 2017). According to this, two studies looked into reciprocal behaviour between working dogs of the Swiss army (Gfrerer & Taborsky, 2017, 2018) and one study combining two projects looked into reciprocity between pet dogs and humans (McGetrick et al., 2021). The results of the first two studies support the prediction. They also found, that even though military dogs can use direct reciprocity, they more often rely on the less cognitively demanding generalized reciprocity (Gfrerer & Taborsky, 2017). Additionally, military dogs can return a received favour by using alternative tasks (Gfrerer & Taborsky, 2018), which has been emphasised to be important in a natural setting for reciprocity to occur (Schweinfurth & Taborsky, 2017). If a military dog received a reward due to its conspecific pulling a rope, the individual reciprocated the favour on the next day by enabling the helpful individual access to food by pushing a lever. This controlled for the ability of the dogs to generalize the cooperative experience (Gfrerer & Taborsky, 2018).

Interestingly though, the most recent study regarding reciprocity in dogs did not find a correlation between the helpfulness of humans (measured in donated food to the dog) and the amount of food the dogs returned (McGetrick et al., 2021). Contrary to the previous two reciprocity studies, pet dogs instead of military dogs were used as subjects and humans instead of other dogs functioned as social partners. Therefore, instead of unfamiliar dogs, unfamiliar humans either provided the subjects with food by pushing a wireless button for a food dispenser or not pushing the button and therefore not donating food to the dog. During the test condition, the dogs were given the opportunity to repay the favour.

Apart from comprehension difficulties, a couple of other potential explanations were proposed for why reciprocity did not occur in pet dogs: One explanation focuses on the species difference (McGetrick et al., 2021). Although dogs have been shown to differentiate between helpful and unhelpful humans and tend to spend more time with helpful ones (Carballo et al., 2015), in former prosociality studies dogs acted prosocially towards conspecifics but ceased to do so towards humans (Quervel-Chaumette et al., 2016). In the mentioned studies with military working dogs by Gfrerer and Taborsky (2017, 2018) potential problems with species differences were avoided by using only conspecifics.

Another explanation by McGetrick et al. (2021) focuses on the influence of familiarity on dog-human interactions. This influence is called the familiarity effect, according to which being familiar with someone or something makes a difference in one's interactions and reactions (Gruman et al., 2016). Humans for example are known to rely heavily on social affiliation and tend to act more altruistically the more acquainted they are with another individual (Massen et al., 2010). Although the familiarity effect in dog-human interactions has not been well studied so far, McGetrick et al. (2021) point out, that familiarity could play an important role in dog-human interactions and therefore also in cooperative behaviour between the two species. For example, the familiarity effect can influence dogs' cognitive skills, which are specifically necessary during a test paradigm for direct reciprocity: It was discovered that it is more difficult for dogs to differentiate unfamiliar humans of the same sex (as were used during McGetrick et al.'s (2021) study) than familiar humans of the same sex (Carballo et al., 2015). Furthermore, it is known that regardless of their training level, dogs are better at interpreting directional pointing, gazing and head movement cues from a familiar than from an unfamiliar human (Cunningham & Ramos, 2014) and have been shown to prefer cues provided by their owner over cues from an unfamiliar social partner (Cook et al., 2014).

The familiarity effect between dog-dog interactions is also not well studied and so far mixed results were found. In military dogs familiarity between social partners was not necessary for reciprocity to occur (Gfrerer & Taborsky, 2017). Then again, this was not the case during prosocial tasks with pet dogs: Quervel-Chaumette et al. (2015) applied a bar-pulling paradigm by which the subjects got the choice to either provide another pet dog in a neighbouring enclosure with food or pull another bar by which none of them got a reward. The results showed that although pet dogs are capable of other-regarding behaviour they tend to help only familiar individuals while not providing unfamiliar ones with food. A follow-up study by Date et al. (2016) looked at whether this familiarity effect would hold up in another task. Instead of bar-pulling, a token choice paradigm, similar to those used with primates was conducted. During the experiment, the subjects could choose a token to deliver food to their conspecific in the adjacent enclosure. Although the familiarity effect remained, the experimenters could not rule out social facilitation as the underlying motive for prosocial behaviour. Therefore, Dale et al. (2016) discussed how influential even small changes in the setup and tasks can be and how important the only rarely included controls for social facilitation are.

As the current master thesis is a follow-up project to the reciprocity study by McGetrick et al. (2021) the discussed influence of familiarity and species differences on the behaviour of pet dogs were strongly taken into account. Therefore, the focus shifted from interactions between unfamiliar dogs and humans to interactions between familiar dogs, which means, all social partners were dogs, familiar with the subject. This project was further designed with several modifications to facilitate the understandability of the setup. For one thing, the food bowl and the box with the button were orientated in a way that the dogs stood parallel to each other, instead of facing each other directly. This not only reduced the risk for potential aggression (as confronting a conspecific could be uncomfortable for some dogs), but it also ensured that both individuals could witness the button being pressed and food falling into the bowl, instead of the box blocking the view for the dog in the neighbouring enclosure. For another thing, in accordance with the mentioned token choice paradigm by Dale et al. (2016), it was assumed that dogs are capable of learning a non-mechanical connection between a wireless button and food being delivered by a food dispenser. But by hiding as well as moving the food dispenser out of the enclosure and letting the food pieces fall through a tube into the bowl it was anticipated that the dogs would more easily understand the connection between reward and button pressing as they would not focus their attention on the food dispenser and could not try to physically manipulate it to obtain the reward.

1.1. Hypotheses and predictions

For the current study, it was hypothesized that dogs reciprocate help received from helpful individuals (Gfrerer & Taborsky, 2017, 2018). Therefore, it was predicted that during a food-donation scenario the subjects would increase the amount of food they donate to a familiar dog with whom they had a helpful experience compared to a familiar one who had not donated food for them. Additionally, as it was assumed that the subjects understood the task and the setup it was predicted that regardless of the previous experience dogs would neither push the button during the asocial nor the social facilitation control.

Furthermore, it was hypothesized that reciprocity would correlate with social attention. As understanding a social task requires social attention (Range et al., 2009; Törnqvist et al., 2015), it was predicted that the amount of gazing would have a positive influence on the dogs' performance. In other words, the more the subject watched their unhelpful partner during the experience phase the less they would press the button, while the more they watched their helpful partner, the more they would press the button for them during the test session.

To test the predictions the data was analysed according to five questions:

1. Question: Do the dogs reciprocate received help (i.e. do they provide food to the helpful partner more than the unhelpful partner and more than in control conditions?)

2. Question: Is the number of times the subject pressed the button in the test condition for the helpful partner influenced by the number of times the helpful partner pressed the button in the experience phase?

3. Question: Is the number of times the subjects pressed the button in the test condition influenced by the proportion of time they spent looking at the partner during the experience phase?

4. Question: Do the subjects spend different durations gazing at the two different partner types during the experience phase?

5. Question: Do the subjects spend different durations gazing at the two different partner types during the test condition?

2. Methods

The following experiment was discussed and approved by the institutional ethics and animal welfare committee in accordance with GSP guidelines and national legislation [ETK-82/05/2019; ETK-060/03/2020; 137/09/2020]. The owners had to sign consent forms to participate.

2.1. Subjects and social partners

For this study pet dogs (*Canis familiaris)* were used. 12 subjects and 24 social partners were included in the study in total. The sex balance was almost even with 7 out of 12 subjects and 13 out of 24 social partners being female (see Supplementary Table 1). All subjects and social partners were recruited via an online recruitment poster or through the already existing databank of the Clever Dog Lab of the Messerli Research Institute at the University of Veterinary Medicine, Vienna. The dogs needed to be socially compatible with other conspecifics and should be older than one year. All breeds, mixed-breeds, both sexes and intact as well as neutered individuals were accepted for the study, provided they could complete the training steps in an adequate amount of training sessions.

For each subject, two partner dogs that were familiar with the subject were needed. The subject had to be in weekly contact with the other dogs for the six months preceding the study to ensure a high familiarity level. To exclude the possibility of kinship differences affecting the results it was attempted to maintain the same degree of kinship regarding the subject and each partner. Ideally, the recruited dogs were to come from three-dog households, with no or the same degree of kinship. Recruitments whereby only one familiar partner originated from the same household as the subject were excluded, as there would have been a familiarity and relationship difference between the partner of the household and the additional recruit. Consequently, in the case of two-dog-households, both animals could only participate either as subjects, with two external partners respectively, or could be trained as social partners for another familiar subject out of another household. By testing each subject with two new social partners the current study generated a dataset of 12 completely independent samples and therefore avoided pseudoreplication. Pseudoreplication is a problem known for a long time (Hurlbert, 1984) and has been stated to be one of the most influential methodological problems in studies about ecology and animal behaviour (Schank, 2009). Pseudoreplication occurs when data (e.g. samples or experimental units) is analysed as independent when in reality it is not. During experimental studies, this can happen for example by sampling individuals multiple times throughout the same experiment. This leads to the impression that more data is available than there is and can alter the results (Garton et al., 2012). Unfortunately, pseudoreplication still happens quite often and rarely is recognised as such (Schank, 2009).

The requirements for the subjects were met most of the time. Nine out of 12 three-dog teams came from the same household. In one of the nine cases, no better role assignment was possible than to allow the sibling of a subject as a social partner while the other social partner came also from the same household but was unrelated. The remaining three teams were assembled from two-dog-households plus one external dog. In two of these three cases, it was possible to use the dogs from the two-dog-households as social partners and add a familiar subject from another household. But in one case the team could only be used with one social partner originating from the same household as the subject, while the other social partner came from another household as the subject. Consequently, two of the 12 teams did not comply with both aspects: the same familiarity level and the same kinship level between the subject and both partners. Because of the lack of better-suited participants, the teams were still included in the study.

2.2. Setup

The setup consisted of two enclosures of wired mesh (1.5 m x 1.5 m) inside a room of the Clever Dog Lab (7.0 m x 6.0 m). The enclosures were separated by around 0.5 m. Two Plexiglas doors between the enclosures ensured that the dogs were able to see but could not have physical contact with each other, while at the same time could quickly be moved from one area to another (see Figure 1).

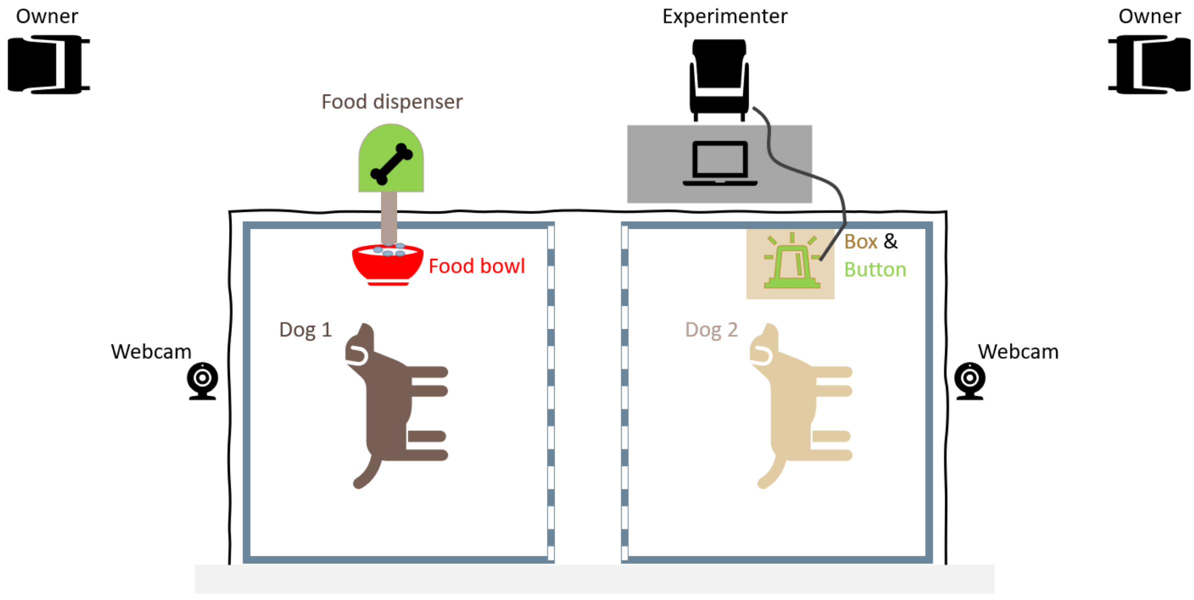


Figure 1: General setup. Two enclosures were separated with fences and Plexiglas doors. So although the dogs could see into both areas, they could not have physical contact with each other. Owner(s), experimenter and the food dispenser were hidden behind black curtains. The wireless button for the food dispenser was in one of the enclosures inside a wooden box which had a rope attached to it. By pulling the rope the experimenter could therefore

manually regulate the dog's access to the button. By pressing the button food fell out of the dispenser and through a tube into the food bowl of the neighbouring enclosure. Webcams on both sides of the enclosure as well as room cameras (not shown) recorded all training and test sessions.

One of the enclosures included a wooden box with a wireless button for a hidden food dispenser outside the neighbouring enclosure. In the neighbouring enclosure, a tube directed the food out of the dispenser through the fence and into a food bowl. By pulling a rope attached to the box the experimenter could manually regulate the dog's access to the button. Whenever the button was pressed an artificial sound would ring and a food reward would be released into the bowl of the other enclosure. The dispenser was always hidden behind a curtain, so only the food falling out of the tube was visible to the dogs. The sound of the rotating food dispenser was audible though.

The Experimenter and the owner(s) were present inside the room but hidden behind black curtains that were hung up on the enclosure's front, left and right sides. Webcams attached to the side of each enclosure ensured the monitoring of the dogs. Additionally, cameras on the walls of the room were used to provide an overview of the whole setup. Afterwards, the video material would be used for coding the animal's behaviour. The webcams gave the experimenters a live view of what was happening inside the enclosure. Additionally, the webcam recordings served as a backup in case the room cameras did not work correctly or a file was lost or broken.

2.3. Study outline

This study included three phases: the training phase, the experience phase and the testing phase (see Figure 2). For the training, a maximum of three sessions was aimed for, although some individuals needed extra training. Two experimenters (the author of this study included) conducted testing in a counterbalanced order, while only the author analysed the data.

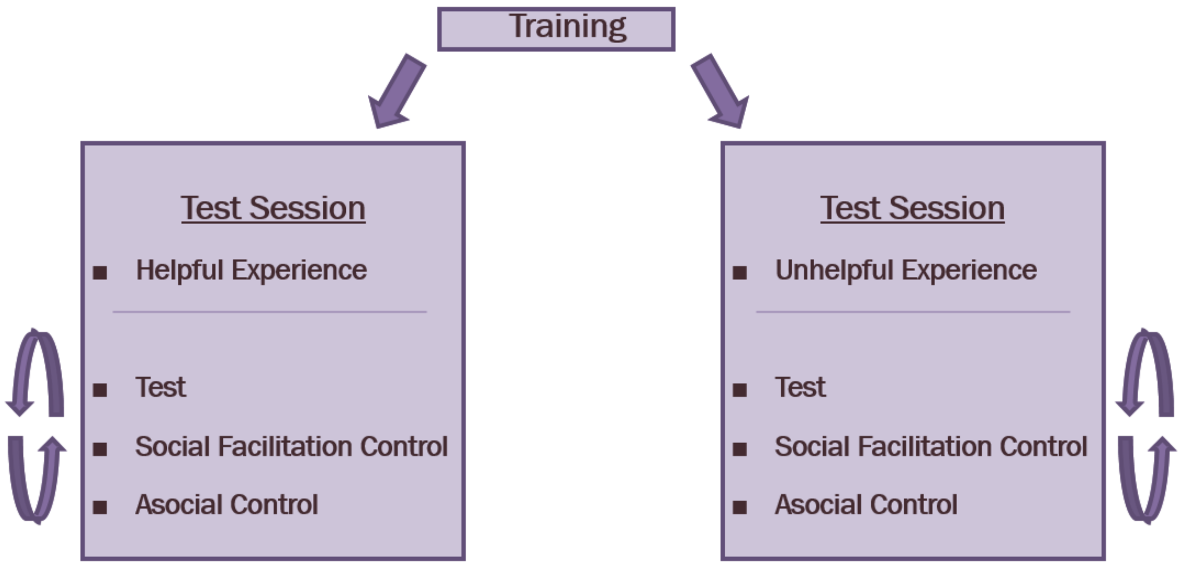


Figure 2: **Outline of the experiment.** First, each dog received training as either a helpful or unhelpful partner or as a subject. On two separate days, test sessions were conducted. Both test sessions began with an experience phase of 10 trials separated by a short break. Therefore, after 5 experience trials, a 2-minute break was included before the remaining 5 experience trials were done. After the experience phase, 5 motivational trials followed. Half of the subjects started with a helpful, the other with an unhelpful experience. After the experience phase, the test condition, social facilitation control and the asocial control were conducted in a partially randomized and counterbalanced manner. Each of the three conditions lasted for 20 trials with 5 motivational trials afterwards.

Each subject needed a total of two appointments for the tests, which took place on two different days. On the first of the two testing days, the subject either experienced a helpful or an unhelpful partner during the experience phase. After each experience phase, the test condition and two controls were conducted in a partially randomized and counterbalanced manner.

Each subject experienced six different condition/partner type combinations: Test, social facilitation and asocial control after the helpful experience, as well as test, social facilitation and asocial control after the unhelpful experience (see Figure 2).

2.4. Training phase

During the training, the future role of the dog was decided semi-randomly. If kinship or different familiarity levels did not predetermine the roles, dogs that were not interested or did not know how to press the button were preferred as unhelpful partners. While dogs that were very eager to press the buttons were preferred as helpful partners. Afterwards, the dogs had to complete different training steps according to their role before they could be tested or used as a social partner (see Figure 3).

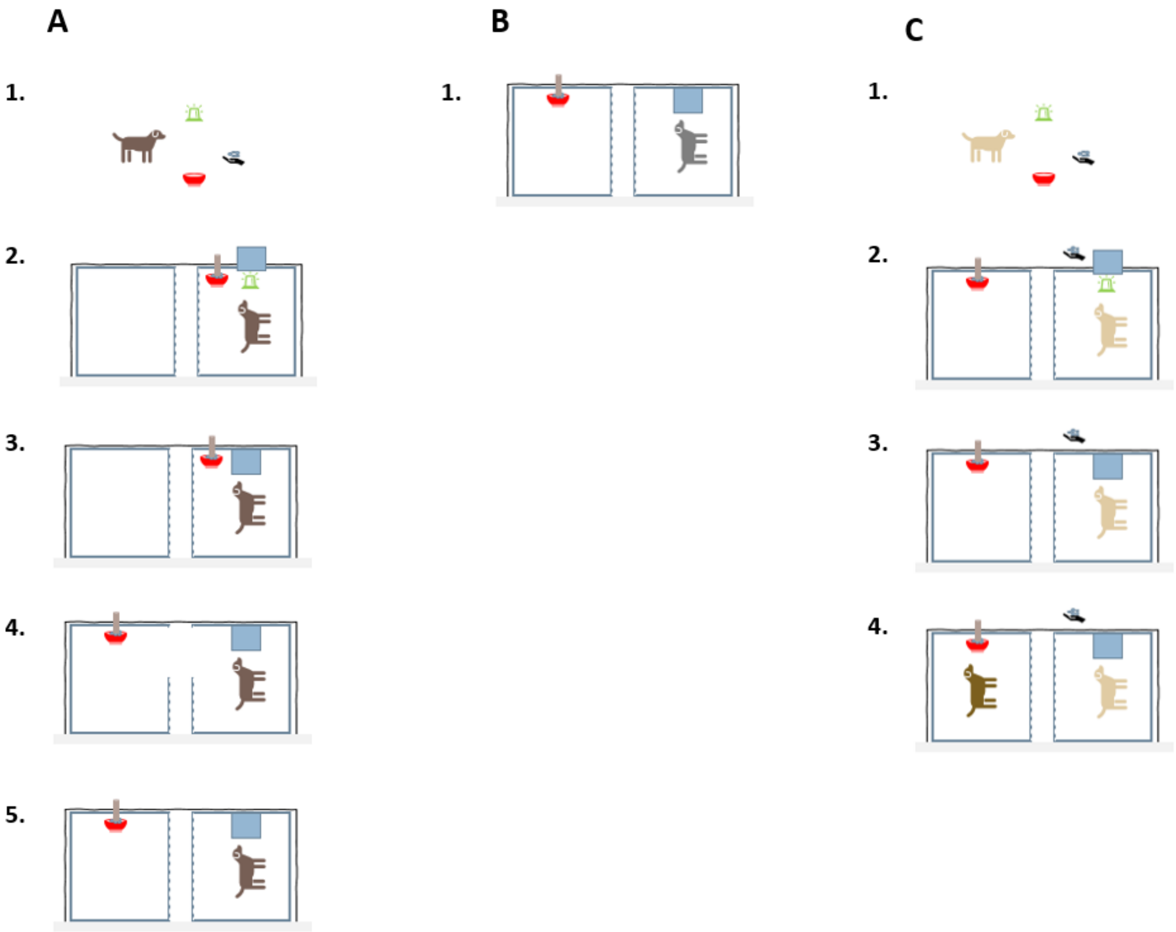


Figure 3: **Outline of the training of the subject and its social partners. A: Training of the subject**. (1) The subject learnt that by pressing the button food would be delivered into a bowl. (2) It was habituated to the enclosure and the open box. (3) It was habituated to the box opening and closing. (4) It could move between the enclosures to receive food. (5) It could no longer move between the enclosures. Five motivational trials, which resembled step 4 of training, followed step 5. **B: Training of the unhelpful partner**. (1) A naïve dog did not receive button training for their role as an unhelpful partner but it was habituated to the enclosure and the moving box. **C: Training of the helpful partner.** (1) The helpful partner learnt that by pressing the button food would be delivered. A reward was only delivered by hand. (2) It was habituated to the enclosure and the open box. A reward was thrown into the enclosure. (3) It was habituated to the box opening and closing. The intervals between the rewards were increased steadily. (4) The helpful partner was habituated to another dog in the neighbouring enclosure and eating the food while the helpful partner pressed the button three to five consecutive times before a reward was thrown into the enclosure.

At first, it was planned to exclude dogs (subjects and social partners) that did not manage to complete the training phase within three sessions. After the first few training sessions, this rule was loosened. Whether to continue training for subjects after three sessions was decided based on the previous training results and the willingness of the owner.

2.4.1. Training of the subject

1. The subject had to learn the non-mechanical connection between the button and the food bowl (see Figure 3 A 1.). Therefore, the button and the bowl were placed 0.5 meters apart

from each other. During this step, the dog was outside the enclosure. The dog had to press the button five consecutive times independently to move on to the next stage.

2. The subject had to get used to being in the enclosure and hearing the noise of the food dispenser and the food falling through the tube into the bowl (see Figure 3 A 2.). Therefore, the button was put inside the box in the enclosure and the bowl was placed 0.5 meters apart from it. During this stage, the box remained open all the time. The food dispenser outside of the enclosure was now introduced and would provide food through a tube if activated by pressing the button. The dog had to press the button ten consecutive times independently to move on to the next stage. Most dogs needed small intermediate stages whereby either the experimenter or the owner sat with them in the enclosure and moved the button step by step into the box.
3. The subject had to get used to the moving box inside the enclosure (see Figure 3 A 3.). Therefore, the box was now opened and closed with the attached rope by the experimenter. It remained open for 10 seconds before it closed for 4 seconds. If the dog was afraid of the movement the experimenter or the owner would sit with them, encourage them or move the box by hand at the beginning. Again, the dog had to press the button ten consecutive times independently to move on to the next stage. Only when the experimenter was no longer in the enclosure and the button was inside the box the counting for independent presses started.
4. The subject had to understand the non-mechanical connection of the button with the food over a longer distance and had to move between the enclosures (see Figure 3 A 4.). Therefore, the bowl was placed in the neighbouring fenced area and the Plexiglas doors were opened for the dog to move freely and press the button whenever the box was opened. This time, the dog had to press ten consecutive times to move on to the next stage. The motivational trials were conducted similarly to this training step.
5. The subject had to understand the inaccessibility of the food if the doors between the enclosures were closed (see Figure 3 A 5.). Therefore, the dog was still able to press the button if the box was opened but was not able to pass to the other enclosure and, consequently, was not able to eat the food in the bowl. The dog experienced ten trials in which the box was opened for 10 seconds before being closed for 4, but it did not matter how often the button was pressed.
6. The subject had to regain its motivation level (see Figure 3 A 6.). Therefore, five motivational trials followed whereby the doors between the enclosures were opened and the dog could access the food again.

2.4.2. Training of the helpful Partner

At first, it was planned to not have a specific training procedure for helpful partners as it was anticipated, that specific training would take many training sessions. Instead, very motivated dogs that pressed in eight to ten trials during the food-inaccessible step of the subject training (step 5) were used as helpful partners.

The inability to find such highly motivated dogs in every team and the observation that during the test some dogs ceased to press the button as soon as another dog was present, led to the establishment of a new helpful partner training. No strict trial numbers for the training steps were predefined as different dogs reacted very differently. Consequently, this change of training required also more than three sessions ab initio. If the owner still agreed to take part no limit was put on the number of extra sessions as long as the dog was willing to cooperate and showed improvements. The following steps represent a training example that was viable for most dogs. In some cases, additional individual adaptations were made, that are not shown.

1. The helpful partner had to understand the non-mechanical connection of the button with food (see Figure 3 C 1.). Therefore, the button was presented to the dog and it was rewarded by hand for pressing it. After it had pressed the button independently five consecutive times it moved on to the next stage.
2. The helpful partner had to get used to being in the enclosure. Therefore, the button was put inside the box in the enclosure. During this stage, the box remained open all the time. The food reward was thrown over the enclosure's fence by the experimenter. The dog had to press the button at least ten consecutive times independently to move on to the next stage. Most dogs needed small intermediate stages in which either the experimenter or the owner sat with them in the enclosure and moved the button step by step into the box. Only when the experimenter was no longer in the enclosure and the button was inside the box the counting for independent presses started.
3. The helpful partner had to get used to the moving box inside the enclosure (see Figure 3 C 4.). Therefore, the box was opened and closed through the attached rope by the experimenter. It remained open for 10 seconds, or until the dog pressed before it closed for 4 seconds. Each time by pressing the button food was thrown into the enclosure. Again, the dog had to press the button at least ten consecutive times independently to move on to the next stage.
4. The helpful partner had to get used to food coming into the inaccessible bowl in the other enclosure (see Figure 3 C 3.). Therefore, the food dispenser was activated and by pressing the button food fell through the tube into the bowl inside the other enclosure. But the dog

was not able to access it; instead, it was rewarded by throwing food into the enclosure for each press.

5. The helpful partner had to learn to press the button ideally five consecutive times (see Figure 3 C 2.). Therefore, it was no longer rewarded after just one press. By increasing the number of presses until the food was thrown the dog was trained to press the button more often.

6. The helpful partner had to get used to the presence of a conspecific that ate the donated food in the other enclosure (see Figure 3 C 7.). Therefore, another dog was placed inside the neighbouring fenced area. Whenever the helpful partner pressed the button it witnessed the neighbouring dog eating the incoming food. Again, the helpful partner had to press the button five times independently. In cases in which the helpful dog was very startled by the other dog's presence, they were again rewarded first for every press before the interval was slowly increased.

2.4.3. Training of the unhelpful Partner

1. Naïve dogs played the role of unhelpful individuals (see Figure 3 B 2.). If a dog was not interested in the button or did not know how to press it, no specific training for not pressing the button followed.

2. The unhelpful dog had to get used to the moving box inside the enclosure (see Figure 3 B 2.). Therefore, the box was opened and closed through the attached rope by the experimenter. It remained open for 10 seconds before it was closed for 4 seconds. If the dog was anxious about the movement the experimenter or the owner would sit with them, encourage them or move the box by hand at the beginning. Ideally, the dog did not even try to press. If they showed interest, they were not rewarded.

3. The unhelpful dog had to get used to being in the other enclosure, hearing the noise of the food dispenser and the food falling through the tube into the bowl (see Figure 3 A 2.). Therefore, the food dispenser was activated by hand by the experimenter and would provide food through a tube. This was performed until the dog showed no discomfort with the sound and the food falling through the tube anymore.

2.5. Experience Phase

The experience phase lasted for 10 trials, with a small 2-minute break after 5 trials. During the break, the subject was usually taken out of the room, while their social partner stayed (for exceptions see below).

2.5.1. Helpful Partner Condition

The subject was placed inside the enclosure with the food bowl while the helpful partner had access to the box with the button. Because of its training and/or huge motivation, the helpful partner pressed the button as soon as the box was opened and, therefore, provided the subject with food. During each trial, the box would close as soon as the helpful partner pressed the button and would remain closed for 4 seconds. After five trials the subject was led out of the room by the owner or the experimenter and the helpful dog was rewarded. Therefore, although the subject was able to watch its partner push the button, it could not see, hear or smell this behaviour being rewarded.

During the break, the helpful partner had motivational trials to ensure that they would continue pressing the button for the subject. The training sessions of the helpful partner helped them to become more familiar with the experimenter and the setup so they showed no discomfort when left alone. After the break, five more helpful trials followed.

2.5.2. Unhelpful Partner Condition

The subject was placed inside the enclosure with the food bowl while the unhelpful partner had access to the box with the button. Because it had not received specific training and therefore was naive in this task, the unhelpful partner did not press the button. During each trial, the box remained open for approximately 10 seconds before it stayed closed for another 4 seconds. Therefore, although the subject was able to watch its partner failing to push the button, it also did not see, hear or smell this behaviour being rewarded.

After five trials the subject was led out of the room by the owner while the unhelpful dog stayed in the room with the experimenter for two minutes to control for the time the helpful partner would have had to regain its motivation level. As most dogs came from the same household and shared their owner, only one of the two dogs could be handled by them during the short break. Some unhelpful dogs were anxious to be left alone with the experimenter, so in these cases, the subject remained in the room and their owner led out the unhelpful dog. As the subject knew the experimenters and the setup from their training sessions, they did not seem stressed to be left alone with them.

Five more unhelpful trials followed the break before the unhelpful dog was led out of the room. After the unhelpful dog had left, the subject received the same amount of food as it would have during a helpful condition. This was included to prevent the effect of different satiety levels influencing the behaviour of the subject (Zentall, 2016). Instead of getting the food through the tube inside the enclosure, it was fed into a second bowl outside of the enclosure which was only provided for this short feeding period.

2.6. Testing Phase

2.6.1. Test Condition

During the test condition, the positions of the subject and its social partner were switched so that the subject then had the chance to push the button and donate food to its conspecific. The test condition lasted for 20 trials. Every time the box was opened for a maximum of 10 seconds before it closed for 4 seconds. If the dog activated the button the box was closed immediately and stayed closed for 4 seconds. Five motivational trials and a 1-minute break were added for the subject before either the next session phase started or (in the case of both controls already having been completed) the session ended.

2.6.2. Social Facilitation Control

Because studies have shown that the mere presence of another individual can be enough to alter the behaviour of a subject, a social facilitation control was included (e.g. Boumans et al., 2018). This control avoided the possibility that any button pressing was motivated simply by the presence of the partner. During the social facilitation control, the social partner remained inside the enclosure, while the tube and bowl were removed, and the food dispenser was deactivated. The box with the button opened for the subject in the same way as during the test (20 trials, 4 seconds closed, 10 seconds open). It was predicted that if the subject had learnt to use the setup and by pressing the button aimed to reciprocate received help, it would not press the button if its partner was not able to access the reward. Five motivational trials and a 1-minute break were added for the subject before either the next phase started or (in the case of both controls already having been completed) the session ended.

2.6.3. Asocial Control

To control for the possibility that the dog only pressed the button in the test condition for its enjoyment or because they were exploring and therefore manipulating an interesting object, an asocial control was added. The social partner was absent while the box with the button opened for the subject the same way as during the test (20 trials, 4 seconds closed, 10 seconds open). The dispenser remained functional and, therefore, food fell through the tube into the food bowl. It was predicted that if the subject understood the setup and by pushing the button aimed to reciprocate received help, it would not press the button if no neighbouring dog was present to access the reward. Five motivational trials and a 1-minute break were added for the subject before either the next session phase started or (in the case that both controls had already been completed) the session ended.

2.7. Motivational Trials

To maintain the subject's motivation level, five motivational trials were added after each phase of the test sessions (experience, test, asocial and social facilitation control), as well as after the last two steps of a dog's training. During these trials, the partner (if one was present) would leave the enclosure and the doors between the fenced areas were opened so the subject could switch independently between both enclosures. Therefore, after pressing the button, the subject could walk to the other side and eat the reward inside the bowl itself. It had to press five consecutive times. If a dog no longer wanted to press during motivational trials in the test session (after the experience phase and between the test and control conditions) the owner was allowed to encourage it. If it stopped pressing during training the session was stopped for a 2-minute break to let the dog out of the enclosure. If afterwards it still refused to press five consecutive times the session was terminated and continued on another day.

2.8. Coding

The main focus of this study lay in the number of times the subjects pressed the button. Touching the green ring around the button with whatever body part was counted as pressing, although, the food dispenser might not have been activated. Additionally, it was attempted to analyse the head orientation of the subject during the experience and the test phase. This was used as an indicator of gaze and therefore attention of the subject to its social partner. Whenever the subject orientated its snout towards the other enclosure, between the right and left fence post, it was coded as looking towards its social partner. In the helpful experience phase also the duration of the feeding was recorded, as during feeding dogs could not orientate their heads towards their partners. Therefore, whenever the dog orientated its head inside the bowl was coded and subtracted from the whole duration of the helpful experience phase.

Although the number of presses in each condition was noted on a sheet of paper, predominantly the recordings of the room cameras were used to determine the number of presses and the head orientation during the session. The recordings of the two webcams were used as a backup in case the room cameras had not recorded correctly or a file was lost or damaged. The webcam videos were analysed in three out of 24 test sessions. In these three videos, the dogs were visible most of the time and therefore could also be analysed for their head orientation, although the whole enclosure was not visible in the recordings. For cases in which the dog exited the webcam's field of view, the duration they were not visible was recorded and subtracted from the duration of the whole specific condition. The video material was coded using Loopy (Loopbio Gmbh).

Although two experimenters conducted testing in a balanced order and number, only one experimenter coded all videos of the project. The second experimenter coded 20 % of the videos independently for interobserver reliability.

2.9. Statistics

The statistical analysis was conducted using R (R Core Team, 2020; version 1.4.1717). Five different models were fitted. Whenever the main response was the number of times the subjects had pressed the buttons a generalized linear mixed model with a binomial error distribution and the default logit link function was used (function "glmer" from the R package lme4 (Bates et al., 2015; version 1.1.27.1)). Whenever a proportion of time was the main response also a generalized linear mixed model was used but with a beta error distribution and the default logit link function (function "glmmTMB" from the R package glmmTMB (Brooks et al., 2017; version 1.1.2.2)). Functions kindly provided by Roger Mundry were used to check which random slopes needed to be included in the models, whether the models were overdispersed, what range the models' estimates spanned and how stable the models were. In the following subsection, the analysis will be described separately depending on the two different main response variables.

To test whether a predictor variable had a significant effect on a model's main response variable a full-null model comparison was performed. A similar model but without the targeted predictor variable was fitted. The fit of each of the two models was then compared using a likelihood ratio test. For this, the "anova" function of the basic R-package stats (R Core Team, 2020; version 4.0.3) was used with the chi-squared test as the test argument. If no significant difference between the two models could be found, it was concluded that the predictor variable did not have a significant influence on the model's fit. In other words, the predictor variable (e.g. how often the partner had pressed for the subject) did not influence the response variable (e.g. how often the subject pressed for the partner) in a significant way.

All plots were created with the R-package ggplot2 (Wickham, 2016; version 3.3.5), except the plots for the models' estimates. These were generated using a function written by Roger Mundry (Supplementary figure 1-5). Additionally to ggplot2, the R-package ggeffects (Lüdecke, 2018; version 1.1.1) was used to calculate models' prediction curves and visualize their 95 % confidence intervals (Figures 5 and 7).

2.9.1. Models with number of presses as main response

For the generalized linear mixed model with a binomial error distribution and a logit link function response variables (how often the dogs pressed and how often they did not) were included with the column bind function "cbind" of R (R Core Team, 2020). Depending on which question the model was designed to address different fixed and random effects were included.

The correlation between the random slopes and random effects was analysed using the "summary" function of R on each model (R Core Team, 2020). Also, the best linear unbiased predictors (BLUPs) and collinearity were assessed using variance inflation factors (VIFs). The histograms of the BLUPs were visually inspected to confirm that the BLUPs were normally distributed. For the VIFs, the function “vif” from the R-package car (Fox & Weisberg, 2019; version 4.0.5) was used. As the package car could not handle generalised linear mixed models, the generalized linear mixed model had to be reduced to a linear model (function “lm” of the basic R-package stats (R Core Team, 2020; version 4.0.3)) and therefore all interactions needed to be excluded. If the values of the according “vif” function were close to 1 a correlation could be ruled out. This was the case in all models.

To interpret the output of the model more easily the condition order and the day order were z-transformed and the conditions and the partner types were dummy coded and centred if they were included as random slopes.

1. Question: Do dogs reciprocate received help

This question focused on whether the subjects, in general, provided more food to the helpful compared to the unhelpful partner and therefore pressed significantly more often in the helpful test condition than in the four control conditions (helpful social facilitation control, helpful asocial control, unhelpful social facilitation control, unhelpful asocial control) as well as in the unhelpful test condition. The number of times the subjects pressed the button out of 20 trials was the main response.

The test predictor was the condition (test, social facilitation control and asocial control) as well as the partner type (helpful and unhelpful) and the interaction between the two. Because every subject experienced three conditions, with two different social partners each, and therefore participated in a total of six conditions a sum of 72 observations were made across all 12 subjects.

For the control predictors, the day and condition order were included. The subject ID and the ID of the observation were included as random intercepts. Additionally, the random slopes of condition, partner type, condition order and day order were included. The model as written in R, follows:

*full = glmer(cbind(numberofpresses, number_notpress)~condition*partnertype + z.partnertype_order + z.condition_order + (1 + condition.socialfacilitation + condition.test + partnertype.unhelpful + z.condition_order + z.partnertype_order|subject) + (1|observation), family = binomial, data = xdata, control = contr)*

2. Question: Is the number of times the subject pressed the button in the test condition for the helpful partner influenced by the number of times the helpful partner pressed the button in the experience phase?

The second question focused on the test and experience condition. It was investigated whether the number of times the helpful partner provided the subject with food during the experience phase was linked to the number of times the subject pressed the button in the test condition for the corresponding helpful partner. For this, only observations from the helpful test condition were used, making it one observation per subject and, therefore, 12 observations in total.

The number of times the subject pressed the button was used as the main response. The test predictor was the number of times the helpful partner pressed for the subject in the experience phase. The day and condition order were included as control predictors. The subject ID was added as a random intercept. As only one observation for each subject per test was made, no random effect of observation needed to be included. No random slopes needed to be included. The model as written in R, follows:

full = glmer(cbind(numberofpresses, number_notpress) ~ numberpresses_helpfulexperience + z.partnertype_order + z.condition_order + (1 | subject), family = binomial, data=ydata, control = contr)

3. Question: Is the number of times the subjects pressed the button in the test condition influenced by the proportion of time they spent looking at the partner during the experience phase?

This question evaluated whether the number of times the subject pressed the button during the test was linked to the proportion of time they had looked at their partner while they experienced them acting either helpfully or unhelpfully. As only observations from the test conditions (test condition after helpful and test condition after unhelpful partner) were used, 24 observations across 12 subjects were analysed.

The number of times the subject pressed the button in the test conditions was used as the main response with the interaction between the proportion of time the subjects spent gazing at their partners and the partner type as the test predictor. Day order and condition order functioned as control predictors. Subject ID and observation ID were included as random intercepts without any additional random slopes. The model as written in R, follows:

*full = glmer(cbind(numberofpresses, number_notpress) ~ prop_gaze*partnertype + z.partnertype_order + z.condition_order + (1 | subject) + (1|observation) , family = binomial, data = ydata, control = contr)*

2.9.2. Models with proportion of time as main response

By evaluating the gazing behaviour of the subjects, it was attempted to measure their attentiveness to their social partners. As attention is a hidden intrinsic state of the dog's mind, the proportion of time the subject had its head turned in its social partner's direction was used as an assessable proxy to estimate attention. The duration of looking at the social partner was measured to calculate the proportion of time the subject gazed at their partner dog. The equation as written in R, follows:

proportion of gazing = duration looking at the partner / duration of the condition

In the case of the helpful experience phase, the time the subject spent feeding was excluded from the duration of the condition. This was excluded because eating reduced the time the dog had to look at the partner without being distracted by the food. The equation as written in R, follows:

proportion of gazing = duration looking at the partner / (duration of condition – feeding time)

A generalized linear mixed model with a beta error distribution and the default logit link function was used whenever the main response variable of a model was a proportion of time. Depending on which question the model was designed to answer (see: 2.9.1. Five questions to analyse), different fixed effects were included. To interpret the output of the model more easily condition order and testing-day order were z-transformed. Random slopes were not needed. Collinearity was assessed via the VIFs by using the function "vif" from the R-package car (Fox & Weisberg, 2019; version 3.0.11). As the package car could not handle generalised linear mixed models, the generalized linear mixed model had to be reduced to a linear model (function "lm" of the basic R-package stats (R Core Team, 2020; version 4.0.3)) and therefore all interactions needed to be excluded.

4. Question: Do the subjects spend different durations gazing at the two different partner types during the experience phase?

This question centred around whether there was a difference in the proportion of time the subjects were looking at their two different partner types. As only the observation during the experience phases was needed, only 24 observations across 12 subjects were analysed.

The proportion of time the subjects spent looking at their partners was used as the main response, while the partner type was the main test predictor and day order functioned as a control predictor. The subject ID was added as a random intercept but no random sloped needed to be included. The model as written in R, follows:

full = glmmTMB(prop_gaze ~ partnertype + z.partnertype_order + (1 | subject), family = beta_family(link="logit"), data = ydata)

5. Question: Do the subjects spend different durations gazing at the two different partner types during the test condition?

The last question analysed whether the subject spend a different amount of time looking at the two partner types during the test condition. Again only two observations per subject were analysed making it 24 observations in total.

The proportion of time the subjects spent looking during the test condition was the main response, with the partner type functioning as a test predictor. Day and condition order were added as control predictors and only the random intercept of subject ID was included, while no random slopes were added. The model as written in R, follows:

full = glmmTMB(prop_gaze ~ partnertype + z.partnertype_order + z.condition_order + (1|subject), family = beta_family(link="logit"), data = ydata)

2.9.3. Interobserver reliability

The function “icc” from the R package “irr” (Gamer et al., 2019; version 0.84.1) was used to evaluate the interobserver reliability by calculating the intraclass correlation coefficient. With regards to the structure of the data, the two-way model and the consistency type settings of the “icc” function were used. Although five statistical models had been analysed, only three measured variables (number of presses, duration of gazing during the experience phase and duration of gazing during the test phase) were evaluated for their interobserver reliability as all five statical models used a variation of the same variables. For all three variables, a statistically excellent interobserver reliability (Regier et al., 2013) of at least 86 % was measured (Number of presses: ICC = 0.986, $n_{subjects}$ = 15, $n_{observers}$ = 2, $p < 0.001$; duration of gazing during the experience phase: ICC = 0.865, $n_{subjects}$ = 5, $n_{observers}$ = 2, $p = 0.013$; duration of gazing during the testing phase: ICC = 0.935, $n_{subjects}$ = 5, $n_{observers}$ = 2, $p = 0.003$).

3. Results

3.1. Influence of condition and partner type on pressing behaviour of the subject?

Model 1: Do the dogs reciprocate received help?

The full-null model comparison revealed neither a significant influence of partner type alone nor of the interaction of the partner type with the condition type on the number of times the subject pressed the button during the test. This implies, that no significant difference in the number of times the subject pressed the button depending on the previously experienced cooperativeness, or the lack thereof, was found (full-null comparison: $\chi^2 = 1.76$, $df = 5$, $p = 0.881$) (see Figure 4). For a detailed output of the model see Supplementary Figure 6.

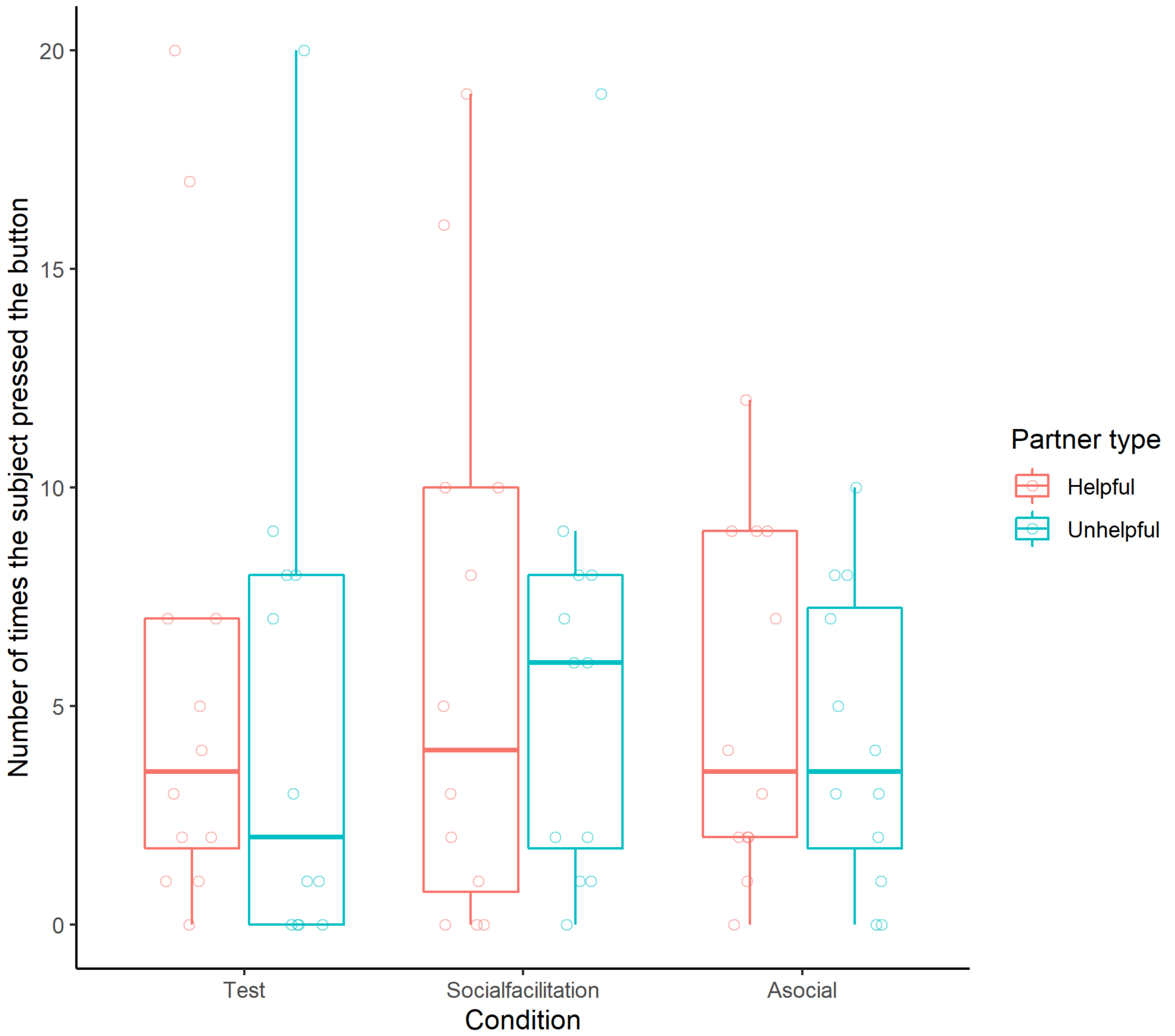


Figure 4: Boxplots showing the number of times the subjects had pressed the button during different conditions (test, social facilitation and asocial control) as well as with different partner types (helpful and unhelpful partner). Circles represent data points, while boxes display the interquartile range (IQR) which contains 50 % of all data points. The thickened horizontal bar of each box shows the median, the whiskers extent from the lower (and upper) hinge to the smallest (and largest) value point but no further than 1.5 times the IQR.

3.2. Influence of pressing count of the helpful partner on pressing behaviour of the subject?

Model 2: Is the number of times the subject pressed the button in the test condition for the helpful partner influenced by the number of times the helpful partner pressed the button in the experience phase?

The full-null model comparison revealed no significant influence of the number of times the helpful social partner pressed the button for the subject on the subject's pressing behaviour during the test. This implies that regardless of how often the helpful partner had pressed during the experience phase, no significant difference in the number of times the subject pressed the button for the helpful partner during the test was found (full-null comparison: $\chi^2 = 0.47$, $df = 1$, $p = 0.492$) (see Figure 5). For a detailed output of the model see Supplementary Figure 7.

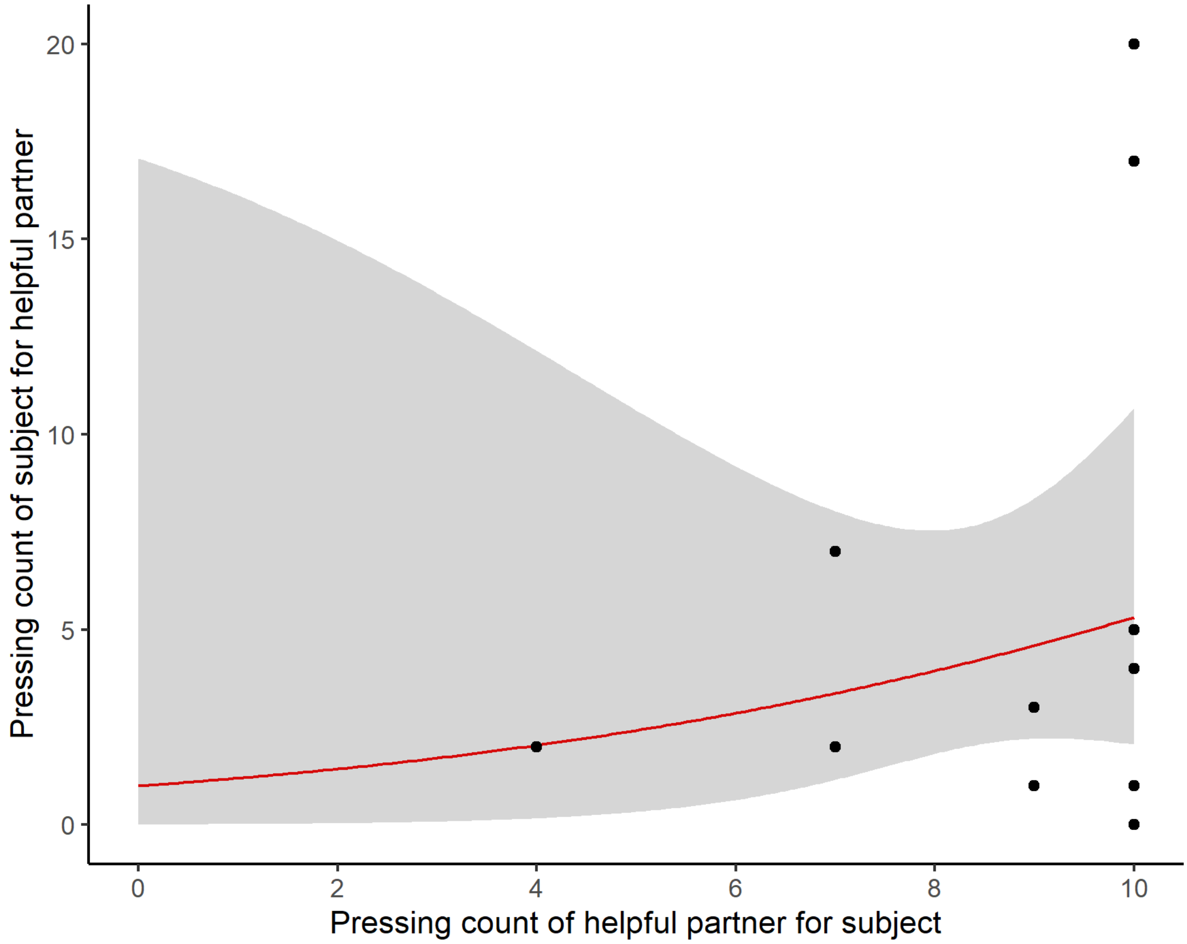


Figure 5: Plot showing the number of times the subjects pressed the button during the test condition in regards to how often their helpful partner had pressed for them during the helpful experience phase. The red line plots the model, while the grey area indicates the 95 % confidence interval.

3.3. Influence of gazing during experience phase on pressing behaviour of the subject?

Model 3: Is the number of times the subjects pressed the button in the test condition influenced by the proportion of time they spent looking at the partner during the experience phase?

The full-null model comparison revealed a significant influence of the interaction between partner type and the subject's proportion of gazing at their partner during the experience phase on the number of times the subject pressed the button during the test (full-null comparison: $\chi 2 = 6.80$, df = 2, $p = 0.033$) (see Figure 6). By looking at the model summary, it was revealed that the longer the subject had looked at their helpful partner during the experience phase, the more they pressed during the test. While for the helpful partner it seemed the opposite relationship emerged, the longer the subject looked, the fewer times they pressed. For a detailed output of the model see Supplementary Figure 8.

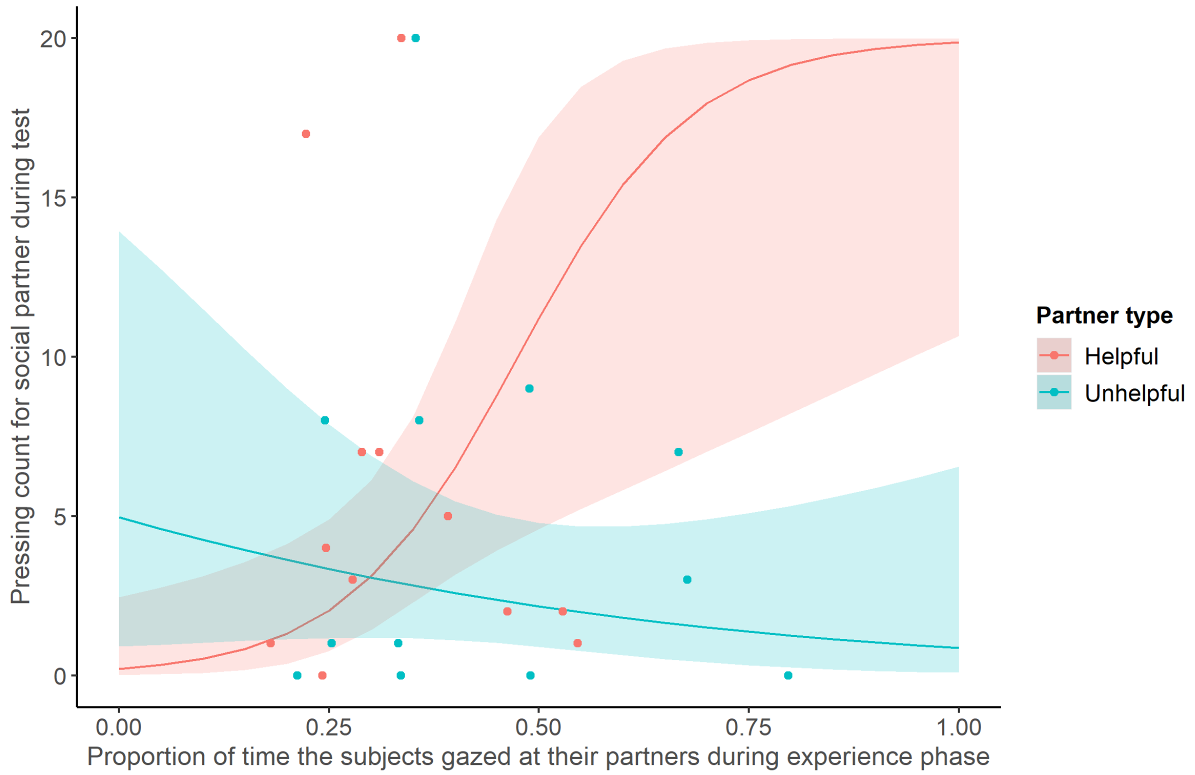


Figure 6: Number of times the subject pressed the button for its social partner (helpful vs. unhelpful) during the test in regards to the subject's proportion of time they had spent gazing at their partner during the experience phase (where the partner did or did not press the button for the subject). The coloured lines plot the model, while the coloured areas indicate the 95% confidence intervals.

3.4. Gazing behaviour of the subject during the experience phase?

Model 4: Do the subjects spend different durations gazing at the two different partner types during the experience phase?

The full-null model comparison revealed no significant influence of the partner type on the subject's gazing behaviour in the experience phase. This implies, that there was no significant difference in the proportion of time the subject looked at the helpful compared to the unhelpful social partner during the experience phase (full-null comparison: $\chi^2 = 2.85$, $df = 2$, $p = 0.241$) (see Figure 7). For a detailed output of the model see Supplementary Figure 8.

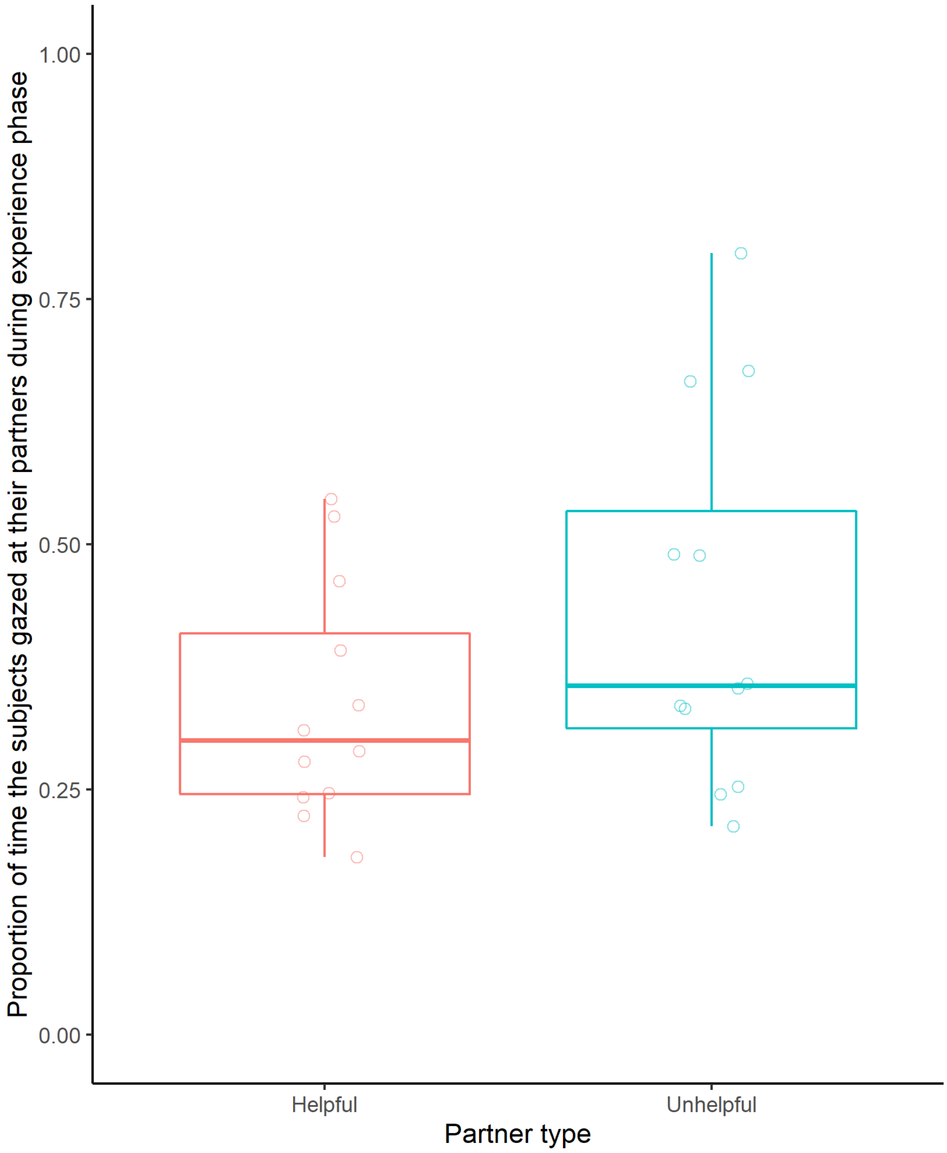


Figure 7: Boxplots showing the proportions of time the subjects spent gazing at their two different partner types during the experience phase. Circles represent data points, while boxes display the interquartile range (IQR) which contains 50 % of all data points. The thickened horizontal bar of each box shows the median, the whiskers extent from the lower (and upper) hinge to the smallest (and largest) value point but no further than 1.5 times the IQR.

3.5. Gazing behaviour of the subject during the test condition?

Model 5: Did the subjects spend different durations gazing at the two different partner types during the test condition?

The full-null model comparison revealed no significant influence of the partner type on the subject's proportion of gazing during the test condition (full-null comparison: $\chi^2 = 2.03$, $df = 1$, $p = 0.155$) (see Figure 8). For a detailed output of the model see Supplementary Figure 10.

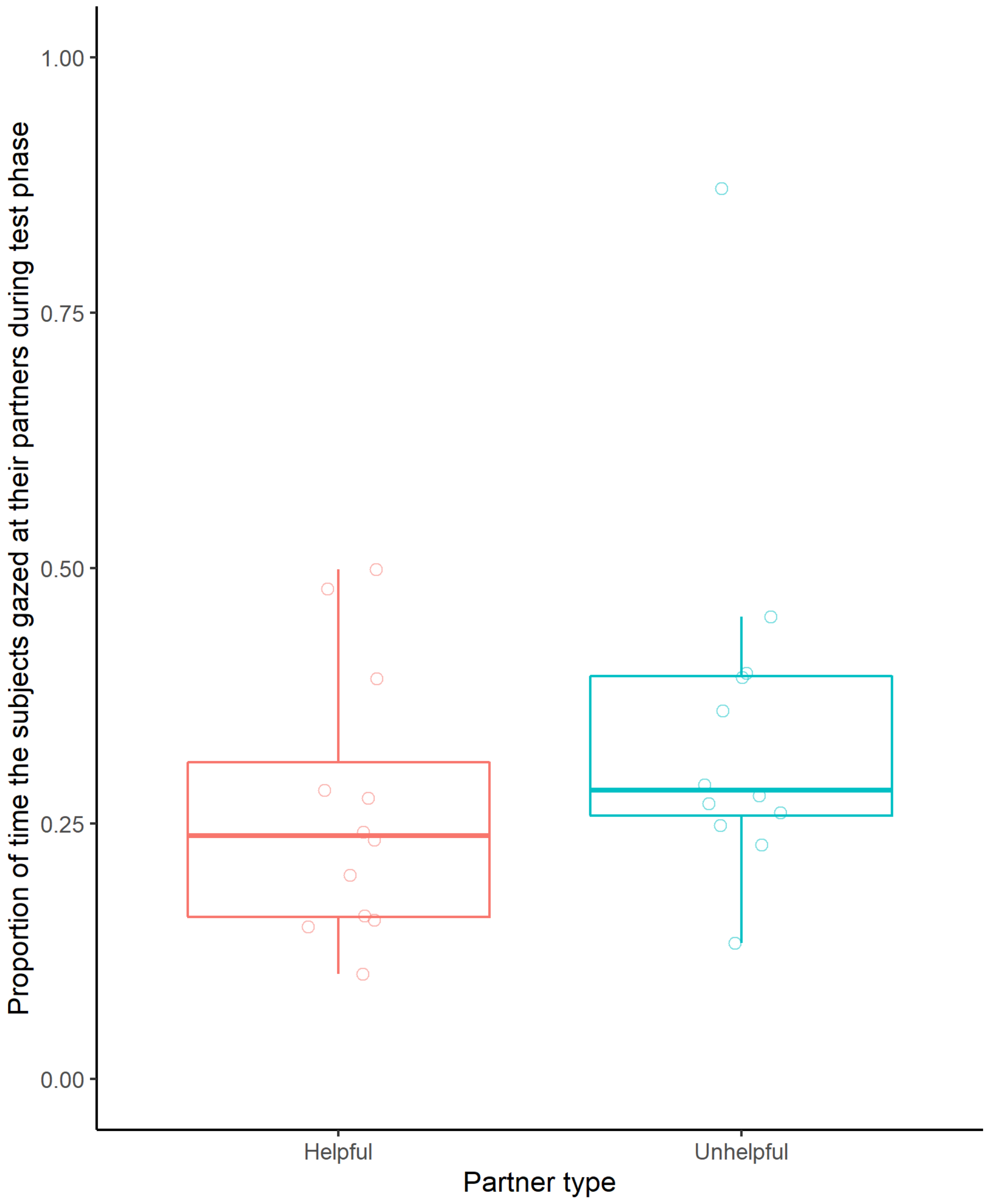


Figure 8: Boxplots showing the proportions of time the subjects spent gazing at their two different partner types during the test condition. Circles represent data points, while boxes display the interquartile range (IQR) which contains 50 % of all data points. The thickened horizontal bar of each box shows the median, the whiskers extent from the lower (and upper) hinge to the smallest (and largest) value point but no further than 1.5 times the IQR.

4. Discussion

Based on the results of the mere pressing behaviour of the subjects, it was concluded that no reciprocity occurred in this study. It had been predicted that differences would be found in the amount of food the subjects donate between test and control conditions, as well as between different partner type conditions after the dogs had experienced one partner donating food, while another did not. But neither the partner type nor the condition type nor how often the helpful partner donated food seemed to influence the pressing response of the subject during the test.

Contrary to this, the head movements of the subject which were analysed as a proxy for gazing behaviour and therefore social attention (Range et al. 2009; Törnqvist et al., 2015, 2020) had a significant influence on the pressing behaviour of the subjects. The more the subject had gazed at their helpful partner during the experience phase, the more they pressed for the helpful partner in the test. And the more they had looked at their unhelpful partner, the less they pressed for the unhelpful partner. Although one cannot measure whether the dogs actually paid attention to their partners' actions or were focused on something else in the general vicinity of the partner, these results fit the idea of comprehending one's partner's role better if one watches them more closely and that the dogs were behaving according to reciprocity. Otherwise, no significant correlation with gazing behaviour was found. The subjects neither differed in the proportion of time they spent looking at their partner during the experience nor the test phase.

But the significant result should be interpreted with caution and highlights the importance of multiple replication attempts during times in which common practices to gain and publish experimental results have led to a replication crisis in science (Ioannidis, 2005; Pashler & Harris, 2012; Baker, 2016). Regardless of the statistical significance, by looking at the model's stability plot, the validity of the result becomes questionable (Supplementary Figure 4). Also, by visually inspecting the model's plot the broad confidence interval of the model fit as well as the scattered data points hint that these results are most likely not reproducible (Figure 6). Baker (2016) famously surveyed over 1.500 scientists from different research fields. The majority of these scientists had at least once failed to reproduce their own work or studies by another researcher. Furthermore, replication attempts, especially failed ones, were said to be harder to publish. In the following, potential explanations why pet dogs fail to cooperate in a reciprocating manner will be discussed and therefore, why previous studies about reciprocity in dogs could not be replicated in the current study.

On the one hand, difficulties in comprehending the setup seem unlikely, as during a study with a token exchange paradigm dogs were able to learn the connection between non-mechanically attached objects (Dale et al., 2016). This seems to make it unlikely that the non-mechanical connection between the button and the food dispenser in the current study was a problem. Furthermore, all current subjects were able to use the food dispenser to feed themselves during the training and the motivational trials. On the other hand, our experiment is also partly similar to a prosociality experiment, because regardless of the experience phase and the role exchange, our subjects had the opportunity to donate food towards their familiar conspecifics. But the subject of the current study did not act according to prosociality, as they did not differ in the amount of food they donated to any social partners. Compared with other prosocial studies that mostly found evidence for prosocial behaviour between familiar dogs (Quervel-Chaumette et al., 2015; Quervel-Chaumette et al., 2016; Dale et al., 2016) (but see: Dale et al. 2019), this could be a hint for comprehension problems of the setup.

Problems with comprehension of the other dog’s role could offer another plausible explanation in the current experiment. Especially the experience phase might have been too short for the dogs to comprehend the social act of the other dog. Bräuer (2015) highlighted that compared to humans and chimpanzees, dogs often seem to have comprehension problems about whether the behaviour of others is goal-directed which therefore hinders cooperation. Gfrerer and Taborsky (2017, 2018) used much longer experience phases, with sessions over many different days which might have eased the dogs’ understanding. Moreover, the role exchange was included in the training during these two studies. Because of concerns about the extent of the pre-experimental training in Gfrerer and Taborsky’s studies (2017, 2018), a lot less training and also fewer experience trials were carried out in the current study. Therefore, although it was aimed to strike a balance between enough training to understand the setup while still not altering the dogs' natural behaviour, dogs' cognitive abilities and flexibility to adapt to a new situation might have been overestimated.

Furthermore, potentially the subjects did not comprehend their partners as cooperative as their social acts were faked. While the unhelpful dogs were always completely naïve to the task, the helpful partners were either trained to press the button repeatedly and wait for their delayed reward or were very motivated to press the button regardless of the reward. Both partner types consequently neither chose to be cooperative or uncooperative. Therefore, also important communication cues like begging to initiate cooperation might have never occurred. Gfrerer and Taborsky (2017, 2018) faked the cooperative act similar to the current study. But they trained the role exchange with both the subjects and the helpful partners, which means, both parties experienced that another dog could donate food for them. This might have facilitated the understanding and enhanced communication cues. In their second reciprocity study, Gfrerer and Taborsky (2018) analyzed the friendly and begging behaviour of the social partners and found that these communication cues correlated with the propensity of the subject to donate food. Therefore in the future, it might be worth looking more at communication cues.

Aside from problems of comprehension, food rewards might have hindered the occurrence of reciprocity in the current study. Perhaps dry food was not motivating enough for dogs to show higher cognitive abilities. Studies with mice have shown that motivation increases timing precision and/or accuracy during a self-feeding task, as well as other temporal cognitive functions (Avlar et al., 2015). Consequently, the use of more motivating rewards might increase the cognitive performance of an animal. During the four different prosociality studies with dogs, the standard reward was sausage pieces (Quervel-Chaumette et al., 2015; Quervel-Chaumette et al., 2016; Dale et al., 2016; Dale et al., 2019). This is usually a highly valued reward for dogs and likely increased their motivation. Compared with wolves, dogs were also observed to be less tolerant of food sharing with conspecifics (Dale et al., 2017), which has been discussed as an important factor for prosociality (Dale et al., 2019) and therefore might also influence reciprocity. Furthermore, the asymmetrical manner in which food is usually shared with the owner might influence whether dogs share it with conspecifics (Carballo, et al., 2020a), as pet dogs might never experience situations in which offering food to another partner (human or dog) would be advantageous. But the critique on food rewards leads to the question of why such strong evidence for reciprocity was found in military dogs despite the use of dry food (Gfrerer & Taborsky, 2017, 2018).

The subject sample is therefore a potential reason for the contradictory results in previous reciprocity studies. Breed but also height, weight and sex were shown to influence dogs' behaviour significantly (e.g. sociability, playfulness, aggression, curiosity, fearfulness) (Stone et al., 2016). Gfrerer and Taborsky (2017, 2018) used only male intact Belgian Malinois military dogs, while McGetrick et al. (2021) and the current study included a wide range of pet dogs. Therefore, it is possible that reciprocity could not be found in the diverse pet dog sample but only in the homogeneous sample of military dogs. Military dogs are trained to fulfil tasks during a variety of situations and under different conditions. Gfrerer and Taborsky (2017) argue that, because of their intense training background with different humans, military dogs might also be better at solving problems and generalizing social experiences with conspecifics. This is supported by a comparative study between untrained pet dogs, trained pet dogs and assistance and therapy working dogs living with human families. Assistance and therapy dogs were shown to solve a task with food hidden in a box more independently and with a higher success rate than even the best-trained pet dogs (Carballo, et al., 2020b). Carballo et al. (2020b) consequently argue that the training type and training methods of working dogs as well as their life experiences changed their behaviour compared to pet dogs. But Carballo et al. (2020b) also review, that the experimental evidence regarding the influence of training is mixed which fits observations from the current study: In the current study, many participating pet dogs were excessively trained for dog sport, man- or pet trailing, obedience and so on. Still, these individuals did not seem to differ in their performance compared to the dogs with less training experience, though this was not explicitly tested for.

Perhaps more important than the training background is the differing intensity of the pre-experimental training. In Gfrerer and Taborsky‘s studies (2017, 2018), in addition to learning how to handle the apparatuses, the role exchange between the helpful partner and the subject was trained. Similar to the current study, the dogs were trained at first to handle the apparatus alone and receive a reward for themselves. But contrary to the current study, they learnt afterwards to donate food to a partner dog in a neighbouring enclosure. For this, they were not rewarded by the experimenters but instead, the dogs' roles were reversed and after each successful donation, the previous donor would get food from the previous receiver. In the following training days, the number of trials until the roles were switched was prolonged until the dogs donated food unrewarded seven times in a row for the other. It took between 14 to 19 days with two training sessions per day for all military dogs to reach this criterion. This intense training might have been problematic as the dogs could have been unintentionally trained to show reciprocal behaviour, as they had learnt they would receive a delayed reward for themselves after the role exchange.

Still, a delayed reward would not explain why the subject ceased to activate the food mechanisms with an unhelpful partner present. Social facilitation controls in this context are particularly important as the mere presence of another individual can facilitate and modulate behaviour (Zajonc, 1965). But Gfrerer and Taborsky's studies (2017 & 2018) lacked social facilitation controls, which could have been used to determine whether the subjects had learnt the purpose of the mechanisms and had the intention of feeding their partner. Still, social facilitation alone also cannot explain the differences between the results of Gfrerer and Taborsky's (2017 & 2018) studies and the current one. Because, if social facilitation was the mere reason, a difference between the asocial control and the social facilitation control should have occurred in the current study. Conversely, this was not the case, as pet dogs did not cease pressing in any condition.

Amidst all the possible explanations for why no reciprocity was found compared to previous studies, the current results could also be a consequence of the small sample size. Small sample sizes are more easily influenced by outliers and biases. A breed bias likely influenced the current data, as 21 out of 36 dogs (58 % of the total sample) were herding dogs (Amerikan Kennel Club) (Supplementary Table 1). Herding dogs are a group of specialised dog breeds that were originally used to shepherd other domesticated animals. Therefore, the preferred traits of these breeds include high fitness, a strong working drive, as well as a general willingness to cooperate with their owners (Ridgway, 2021). These dog breeds were not actively sought for during recruitment but as the majority of dog owners that wanted to participate and could offer three familiar dogs owned such breeds, the sample became biased.

Likewise, the participating owners contributed a bias, as many took part in dog sports, were well educated in dog behaviour or were even dog trainers themselves. Although, as discussed above, so far only a few studies have looked into the influence of training background on cognition tests. Varied and partly contradictory results of previous studies hint that apart from the training intensity also the specific type of training (agility, obedience, human rescue etc.) has influence (reviewed in Carballo, 2020b).

Also, a sex bias cannot be ruled out in the current study. Sex differences in the cognitive abilities of dogs are not a well-studied topic but, so far, studies indicate that sex, as well as sterilization status, can have significant effects on cognitive experiments (reviewed in Junttila et al., 2021). While the military dogs of the previous studies were all uncastrated males (Gfrerer & Taborsky, 2017, 2018), the current study used intact as well as castrated dogs of every sex. Although the sex was nearly balanced with 58.3 % of all subjects and 54.2 % of all social partners being female, castrated versus uncastrated individuals was less evenly balanced. Altogether, only three dog teams where all dogs had been intact were tested (see Supplementary Table 1). Therefore, our sample was much more diverse regarding their sex and sterilization status but could have been influenced by sex differences.

All in all, the current study offers some interesting questions for the future that still need clarification. From an evolutionary point of view and regarding the domestication history of dogs, a comparison with wolves and their capability to act reciprocally would be interesting to explore. Reciprocity is an effective way to facilitate cooperation in a group, which in turn can increase the survival rate of all individuals (Nowak, 2012). Feral dogs were observed to live in less stable packs than wolves, rear their pups alone and also to rely more on individual foraging instead of pack hunting to survive (Marshall-Pescini et al., 2017). This lifestyle perhaps lifted the selective pressure on reciprocity. A recent comparison study on prosociality of wolves and dogs during a touch screen task support this hypothesis, as they found that wolves acted prosocial, while dogs did not (Dale et al., 2019). As the question of whether dogs were selected to cooperate with humans or whether they inherited all necessary traits from wolves continues (e.g. Burkart et al., 2014; Marshall-Pescini et al., 2017; Range et al., 2019), the possibility of reciprocity in wolves might also shine a light on their domestication history.

In the nearest future, the discussed potential problems of the current study should be looked at. A follow-up study with the same setup but with a mechanical connection between the food and the donation mechanism is advised, as this could rule out the possibility that the non-mechanical connection was difficult to understand for the pet dogs. Additionally, the possibility of comprehension problems for the partner dogs' roles should be considered. Therefore, another role exchange during the test session could increase the likelihood of the dogs understanding the task. Also, a bigger and more diverse sample size might have brought different results. Therefore, it seems appropriate to test more subjects with a more diverse breed and training background. Certain recruitment restrictions, like a fixed maximum of certain breeds or more open street recruitments of people that do not know about the Clever Dog Lab, could help to broaden the current sample and consequently increase its representation of the wide and very diverse pet dog population. Lastly, rescuing behaviour, grooming and social support are perhaps ecologically more valid behaviours than food sharing. While grooming and social support seem to be hard to test (as they often occur in unpredicted manners), paradigms for dogs to rescue humans have already been established (Carballo et al. 2020a; Dzik et al., 2021; Van Bourg et al., 2020). In these experiments, a human pretends to be trapped and the dogs can release them. This basic idea could help to develop paradigms between dogs and their conspecifics as well. But even so one has to consider, that by faking the helpful act communication cues could be missing. It might therefore be worth developing paradigms, with which faking the partner's role can be avoided.

5. Conclusion

The current thesis concludes that no reciprocity can be found between familiar pet dogs. Although methodological influences and comprehension difficulties cannot be ruled out, by providing the first data about a reciprocity paradigm between familiar pet dogs this thesis started to fill a gap in the literature on reciprocity that has not been explored before and also highlights once more the importance of replicating studies. Follow-up studies should take the discussed comprehension problems and the potential importance of communication cues more into account and also look more into different training backgrounds and training intensities. Altogether, a bigger sample size with a more accurate representation of the global dog population is needed to answer whether dogs generally can reciprocate favours. A comparison with wolves might also help to shine more light on the evolution of canines' cognitive abilities.

6. Literature

Amerikan Kennel Club. (n.d.). *Herding Group*. https://www.akc.org/dog-breeds/herding/

Akçay, Ç., Reed, V. A., Campbell, S. E., Templeton, C. N., & Beecher, M. D. (2010). Indirect reciprocity: Song sparrows distrust aggressive neighbours based on eavesdropping. *Animal Behaviour*, *80*(6), 1041–1047. https://doi.org/10.1016/j.anbehav.2010.09.009

Alexander, R. D. (1987). *Foundations of human behavior. The biology of moral systems.* Aldine de Gruyter. https://psycnet.apa.org/record/1987-97634-000

Avlar, B., Kahn, J. B., Jensen, G., Kandel, E. R., Simpson, E. H., & Balsam, P. D. (2015). Improving temporal cognition by enhancing motivation. *Behavioral Neuroscience*, *129*(5), 576–588. https://doi.org/10.1037/bne0000083

Axelrod, R., & Hamilton, W. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. https://doi.org/10.1126/science.7466396

House, B., Silk J., Lambeth S., Schapiro S. (2014) Task Design Influences Prosociality in Captive Chimpanzees (Pan troglodytes). PLoS One 9: e103422. https://doi.org/10.1371/journal.pone.0103422 PMID: 25191860

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature, 533(7604), 452–454. https://doi.org/10.1038/533452a

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Barta, Z., McNamara, J. M., Huszár, D. B., & Taborsky, M. (2011). Cooperation among non-relatives evolves by state-dependent generalized reciprocity. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1707), 843–848. https://doi.org/10.1098/rspb.2010.1634

Bartlett, M. Y., & DeSteno, D. (2006). Gratitude and prosocial behavior : Helping when it costs you. *Psychological Science*, *17*(4), 319–325. https://doi.org/10.1111/j.1467-9280.2006.01705.x

Booksmythe, I., Hayes, C., Jennions, M. D., & Backwell, P. R. Y. (2011). The effects of neighbor familiarity and size on cooperative defense of fiddler crab territories. *Behavioral Ecology*, *23*(2), 285–289. https://doi.org/10.1093/beheco/arr184

Boumans, I. J. M. M., de Boer, I. J. M., Hofstede, G. J., & Bokkers, E. A. M. (2018). How social factors and behavioural strategies affect feeding and social interaction patterns in pigs. *Physiology and Behavior*, *194*, 23–40. https://doi.org/10.1016/j.physbeh.2018.04.032

Bräuer, J. (2015). I do not understand but I care. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *16*(3), 341–360. https://doi.org/10.1075/is.16.3.01bra

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., & Maechler, Martin Bolker, B. M. (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378–400. https://journal.r-project.org/archive/2017/RJ-2017-066/index.html

Bshary, R., & D'Souza, A. (2005). Cooperation in communication networks: Indirect reciprocity in interactions between cleaner fish and client reef fish. In *Animal Communication Networks* (pp. 521–539). Cambridge University Press. https://doi.org/10.1017/CBO9780511610363.027

Burkart, J. M., Allon, O., Amici, F., Fichtel, C., Finkenwirth, C., Heschl, A., Huber, J., Isler, K., Kosonen, Z. K., Martins, E., Meulman, E. J., Richiger, R., Rueth, K., Spillmann, B., Wiesendanger, S., & van Schaik, C. P. (2014). The evolutionary origin of human hyper-cooperation. *Nature Communications*, *5*(1), 4747. https://doi.org/10.1038/ncomms5747

Carballo, F., Dzik, V., Freidin, E., Damián, J. P., Casanave, E. B., & Bentosela, M. (2020a). Do dogs rescue their owners from a stressful situation? A behavioral and physiological assessment. *Animal Cognition*, *23*(2), 389–403. https://doi.org/10.1007/s10071-019-01343-5

Carballo, F., Cavalli, C. M., Gácsi, M., Miklósi, Á., & Kubinyi, E. (2020b). Assistance and Therapy Dogs Are Better Problem Solvers Than Both Trained and Untrained Family Dogs. *Frontiers in Veterinary Science*, *7*, 164. https://doi.org/10.3389/fvets.2020.00164

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*, *10*(2), e0116314. https://doi.org/10.1371/journal.pone.0116314

Carter, G. (2014). The Reciprocity Controversy. *Animal Behavior and Cognition*, *1*(3), 368. https://doi.org/10.12966/abc.08.11.2014

Cohen, L. B. (1972) . Attention-getting and attention-holding processes of infant visual preferences. Child Dev. 43:869–875. [PubMed: 5056611] DOI???

Cook, A., Arter, J., & Jacobs, L. F. (2014). My owner, right or wrong: The effect of familiarity on the domestic dog's behavior in a food-choice task. *Animal Cognition*, *17*(2), 461–470. https://doi.org/10.1007/s10071-013-0677-0

Cunningham, C. L., & Ramos, M. F. (2014). Effect of training and familiarity on responsiveness to human cues in domestic dogs (Canis familiaris). *Animal Cognition*, *17*(3), 805–814. https://doi.org/10.1007/s10071-013-0714-z

Dale, R., Palma-Jacinto, S., Marshall-Pescini, S., & Range, F. (2019). Wolves, but not dogs, are prosocial in a touch screen task. *PLoS ONE*, *14*(5), e0215444. https://doi.org/10.1371/journal.pone.0215444

Dale, R., Quervel-Chaumette, M., Huber, L., Range, F., & Marshall-Pescini, S. (2016). Task differences and prosociality; investigating pet dogs' prosocial preferences in a token choice paradigm. *PLoS ONE*, *11*(12). https://doi.org/10.1371/journal.pone.0167750

Dale, R., Range, F., Stott, L., Kotrschal, K., & Marshall-Pescini, S. (2017). The influence of social relationship on food tolerance in wolves and dogs. *Behavioral Ecology and Sociobiology*, *71*(7), 107. https://doi.org/10.1007/s00265-017-2339-8

Darwin, C. (1859). The origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life. In *The origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life.* London : John Murray. https://doi.org/10.5962/bhl.title.87916

Dzik, M. V., Carballo, F., Casanave, E., & Bentosela, M. (2021). Effects of oxytocin administration and the dog–owner bond on dogs' rescue behavior. *Animal Cognition*, *24*(6), 1191–1204. https://doi.org/10.1007/s10071-021-01512-5

Fox, J., & Weisberg, S. (2019). *An {R} Companion to Applied Regression*. Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Freedman, A. H., Wayne, R. K. (2017) Deciphering the Origin of Dogs: From Fossils to Genomes. Annu Rev Anim Biosci. Feb 8;5:281-307. https://doi.org/10.1146/annurev-animal-022114-110937. PMID: 27912242.

Fugazza, C., & Miklósi, Á. (2014). Deferred imitation and declarative memory in domestic dogs. *Animal Cognition*, *17*(2), 237–247. https://doi.org/10.1007/s10071-013-0656-5

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement*. https://cran.r-project.org/package=irr

Garton, E., Horne, J. S., Aycrigg, J. L., & Ratti, J. T. (2012). Chapter 1: Research and experimental design. In Techniques for Wildlife Investigations and Management (7th ed., Vol. 1). Wildlife Research Methods.

Gfrerer, N., & Taborsky, M. (2017). Working dogs cooperate among one another by generalised reciprocity. *Scientific Reports*, *7*. https://doi.org/10.1038/srep43867

Gfrerer, N., & Taborsky, M. (2018). Working dogs transfer different tasks in reciprocal cooperation. *Biology Letters*, *14*(2). https://doi.org/10.1098/rsbl.2017.0460

Gruman, J. A., Schneider, F. W., & Coutts, L. M. (2016). *Applied social psychology : understanding and addressing social and practical problems* (3rd ed.).

Hamilton, W. D. (1964). The genetical evolution of social behaviour. I. *Journal of Theoretical Biology*, *7*(1), 1–16. https://doi.org/10.1016/0022-5193(64)90038-4

Hauser, M. D., Chen, M. K., Chen, F., & Chuang, E. (2003). Give unto others: Genetically unrelated cotton-top tamarin monkeys preferentially give food to those who altruistically give food back. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1531), 2363–2370. https://doi.org/10.1098/rspb.2003.2509

Hurlbert, S. H. (1984). Pseudoreplication and the Design of Ecological Field Experiments. Ecological Monographs, 54(2), 187–211. https://doi.org/10.2307/1942661

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. PLoS Medicine, 2(8), e124. https://doi.org/10.1371/journal.pmed.0020124

Junttila, S., Huohvanainen, S., & Tiira, K. (2021). Effect of Sex and Reproductive Status on Inhibitory Control and Social Cognition in the Domestic Dog (Canis familiaris). *Animals*, *11*(8), 2448. https://doi.org/10.3390/ani11082448

Lehmann, L., & Keller, L. (2006). The evolution of cooperation and altruism – a general framework and a classification of models. *Journal of Evolutionary Biology*, *19*(5), 1365–1376. https://doi.org/10.1111/j.1420-9101.2006.01119.x

Loopbio Gmbh. (n.d.). *Loopy*. http://loopb.io

Lüdecke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Marshall-Pescini, S., Cafazzo, S., Virányi, Z., & Range, F. (2017). Integrating social ecology in explanations of wolf–dog behavioral differences. In *Current Opinion in Behavioral Sciences* (Vol. 16, pp. 80–86). Elsevier Ltd. https://doi.org/10.1016/j.cobeha.2017.05.002

Massen, J., Sterck, E., & de Vos, H. (2010). Close social associations in animals and humans: functions and mechanisms of friendship. *Behaviour*, *147*(11), 1379–1412. https://doi.org/10.1163/000579510X528224

McGetrick, J., Poncet, L., Amann, M., Schullern-Schrattenhofen, J., Fux, L., Martínez, M., & Range, F. (2021). Dogs fail to reciprocate the receipt of food from a human in a food-giving task. *PLOS ONE*, *16*(7), e0253277. https://doi.org/10.1371/journal.pone.0253277

Melis, A. P., & Semmann, D. (2010). How is human cooperation different? In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 365, Issue 1553, pp. 2663–2674). Royal Society. https://doi.org/10.1098/rstb.2010.0157

Molesti, S., & Majolo, B. (2017). Evidence of direct reciprocity, but not of indirect and generalized reciprocity, in the grooming exchanges of wild Barbary macaques (*Macaca sylvanus*). *American Journal of Primatology*, *79*(9), e22679. https://doi.org/10.1002/ajp.22679

Molnár, C., Pongrácz, P., Faragó, T., Dóka, A., & Miklósi, Á. (2009). Dogs discriminate between barks: The effect of context and identity of the caller. *Behavioural Processes*, *82*(2), 198–201. https://doi.org/10.1016/j.beproc.2009.06.011

Nowak, M. A. (2006). Five rules for the evolution of cooperation. In *Science* (Vol. 314, Issue 5805, pp. 1560–1563). NIH Public Access. https://doi.org/10.1126/science.1133755

Nowak, M. A. (2012). Evolving cooperation. *Journal of Theoretical Biology*, *299*, 1–8. https://doi.org/10.1016/j.jtbi.2012.01.014

Pashler, H., & Harris, C. R. (2012). Is the Replicability Crisis Overblown? Three Arguments Examined. Perspectives on Psychological Science, 7(6), 531–536. https://doi.org/10.1177/1745691612463401

Pfeiffer, T., Rutte, C., Killingback, T., Taborsky, M., & Bonhoeffer, S. (2005). Evolution of cooperation by generalized reciprocity. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1568), 1115–1120. https://doi.org/10.1098/rspb.2004.2988

Quervel-Chaumette, Mylene, Dale, R., Marshall-Pescini, S., & Range, F. (2015). Familiarity affects other-regarding preferences in pet dogs. *Scientific Reports*, *5*. https://doi.org/10.1038/srep18102

Quervel-Chaumette, Mylène, Mainix, G., Range, F., & Marshall-Pescini, S. (2016). Dogs do not show pro-social preferences towards humans. *Frontiers in Psychology*, *7*(OCT), 1416. https://doi.org/10.3389/fpsyg.2016.01416

R Core Team. (2020). *R: A language and environment for statistical computing*. https://www.r-project.org/

Range, F., Horn, L., Bugnyar, T., Gajdon, G. K., & Huber, L. (2009). Social attention in keas, dogs, and human children. *Animal Cognition*, *12*(1), 181–192. https://doi.org/10.1007/s10071-008-0181-0

Range, F., & Huber, L. (2007). Attention in common marmosets: implications for social-learning experiments. *Animal Behaviour*, *73*(6), 1033–1041. https://doi.org/10.1016/j.anbehav.2006.07.015

Range, F., Marshall-Pescini, S., Kratz, C., & Virányi, Z. (2019). Wolves lead and dogs follow, but they both cooperate with humans. *Scientific Reports*, *9*(1), 3796. https://doi.org/10.1038/s41598-019-40468-y

Regier, Darrel A.; Narrow, William E.; Clarke, Diana E.; Kraemer, Helena C.; Kuramoto, S. Janet; Kuhl, Emily A.; Kupfer, David J. (2013). DSM-5 Field Trials in the United States and Canada, Part II: Test-Retest Reliability of Selected Categorical Diagnoses. American Journal of Psychiatry. 170 (1): 59–70. https://doi.org/10.1176/appi.ajp.2012.12070999

Ridgway, M. (2021). Herding Dogs. *Veterinary Clinics of North America: Small Animal Practice*, *51*(4), 975–984. https://doi.org/10.1016/j.cvsm.2021.04.013

Rutte, C., & Taborsky, M. (2007). Generalized reciprocity in rats. *PLoS Biology*, *5*(7), 1421–1425. https://doi.org/10.1371/journal.pbio.0050196

Rutte, C., & Taborsky, M. (2008). The influence of social experience on cooperative behaviour of rats (Rattus norvegicus): Direct vs generalised reciprocity. *Behavioral Ecology and Sociobiology*, *62*(4), 499–505. https://doi.org/10.1007/s00265-007-0474-3

Schank, J. C., & Koehnle, T. J. (2009). Pseudoreplication is a pseudoproblem. Journal of Comparative Psychology, 123(4), 421–433. https://doi.org/10.1037/a0013579

Schweinfurth, M. K., & Call, J. (2019). Reciprocity: Different behavioural strategies, cognitive mechanisms and psychological processes. In *Learning and Behavior* (Vol. 47, Issue 4, pp. 284–301). Springer. https://doi.org/10.3758/s13420-019-00394-5

Schweinfurth, M. K., & Taborsky, M. (2017). The transfer of alternative tasks in reciprocal cooperation. *Animal Behaviour*, *131*, 35–41. https://doi.org/10.1016/j.anbehav.2017.07.007

Stone, H. R., McGreevy, P. D., Starling, M. J., & Forkman, B. (2016). Associations between Domestic-Dog Morphology and Behaviour Scores in the Dog Mentality Assessment. *PLOS ONE*, *11*(2), e0149403. https://doi.org/10.1371/journal.pone.0149403

Taborsky, M., Frommen, J. G., & Riehl, C. (2016a). Correlated pay-offs are key to cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1687), 20150084. https://doi.org/10.1098/rstb.2015.0084

Taborsky, M., Frommen, J. G., & Riehl, C. (2016b). Correlated pay-offs are key to cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1687), 20150084. https://doi.org/10.1098/rstb.2015.0084

Tiira, K., Tikkanen, A., & Vainio, O. (2020). Inhibitory control – Important trait for explosive detection performance in police dogs? *Applied Animal Behaviour Science*, *224*, 104942. https://doi.org/10.1016/j.applanim.2020.104942

Törnqvist, H., Somppi, S., Koskela, A., Krause, C. M., Vainio, O., & Kujala, M. V. (2015). Comparison of dogs and humans in visual scanning of social interaction. *Royal Society Open Science*, *2*(9). https://doi.org/10.1098/rsos.150341

Törnqvist, H., Somppi, S., Kujala, M. V., & Vainio, O. (2020). Observing animals and humans: Dogs target their gaze to the biological information in natural scenes. *PeerJ*, *8*. https://doi.org/10.7717/peerj.10341

Van Bourg, J., Patterson, J. E., & Wynne, C. D. L. (2020). Pet dogs (Canis lupus familiaris) release their trapped and distressed owners: Individual variation and evidence of emotional contagion. *PLOS ONE*, *15*(4), e0231742. https://doi.org/10.1371/journal.pone.0231742

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology*, *46*(1), 35–57. https://doi.org/10.1086/406755

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. https://ggplot2.tidyverse.org

Wilkinson, G. S., Carter, G. G., Bohn, K. M., & Adams, D. M. (2016). Non-kin cooperation in bats. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 371, Issue 1687). Royal Society of London. https://doi.org/10.1098/rstb.2015.0095

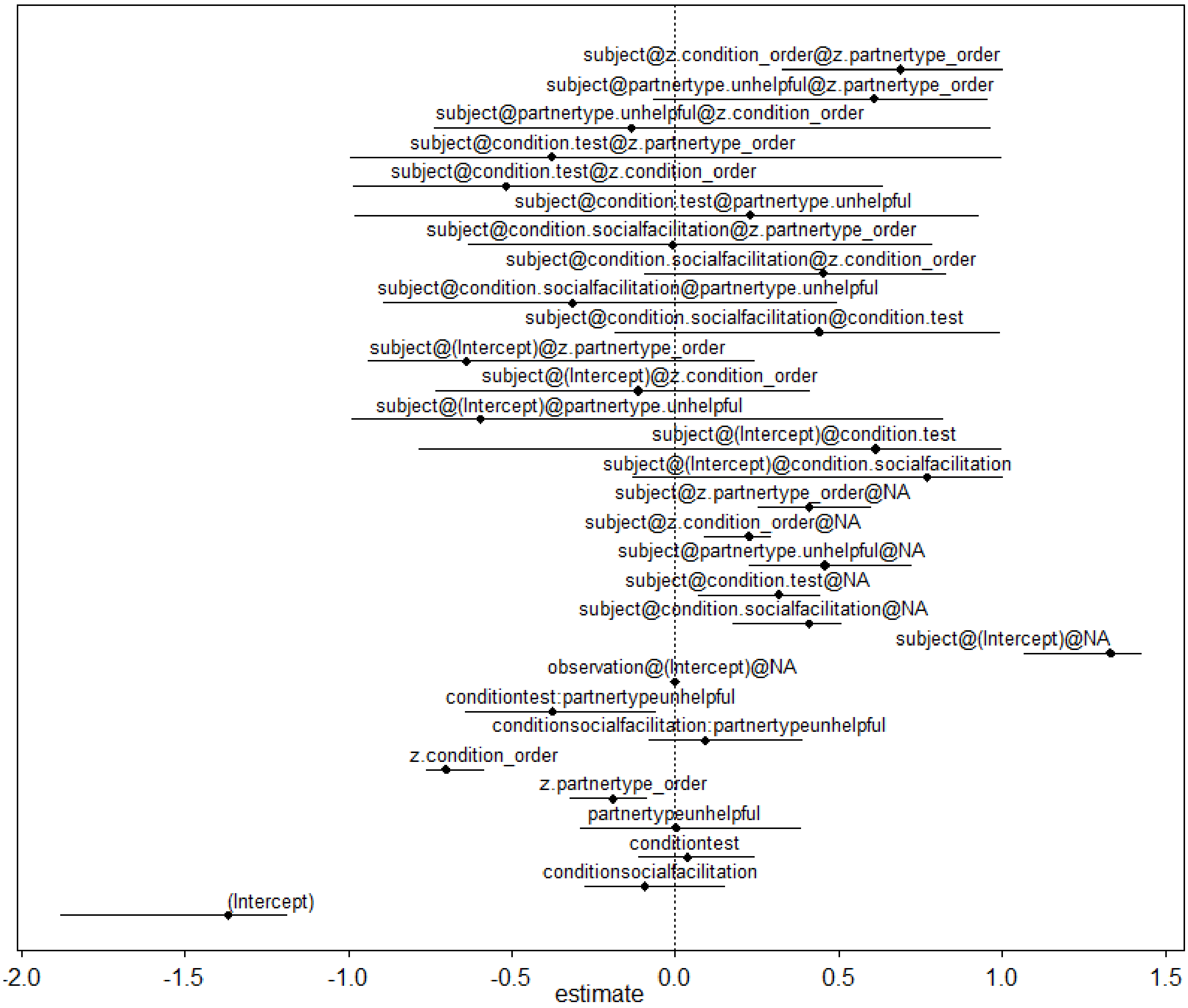
Zajonc, R. B. (1965). Social Facilitation. *Science*, *149*(3681), 269–274. https://doi.org/10.1126/science.149.3681.269

Zentall, T. R. (2016). Reciprocal altruism in rats: Why does it occur? *Learning and Behavior*, *44*(1), 7–8. https://doi.org/10.3758/s13420-015-0201-2

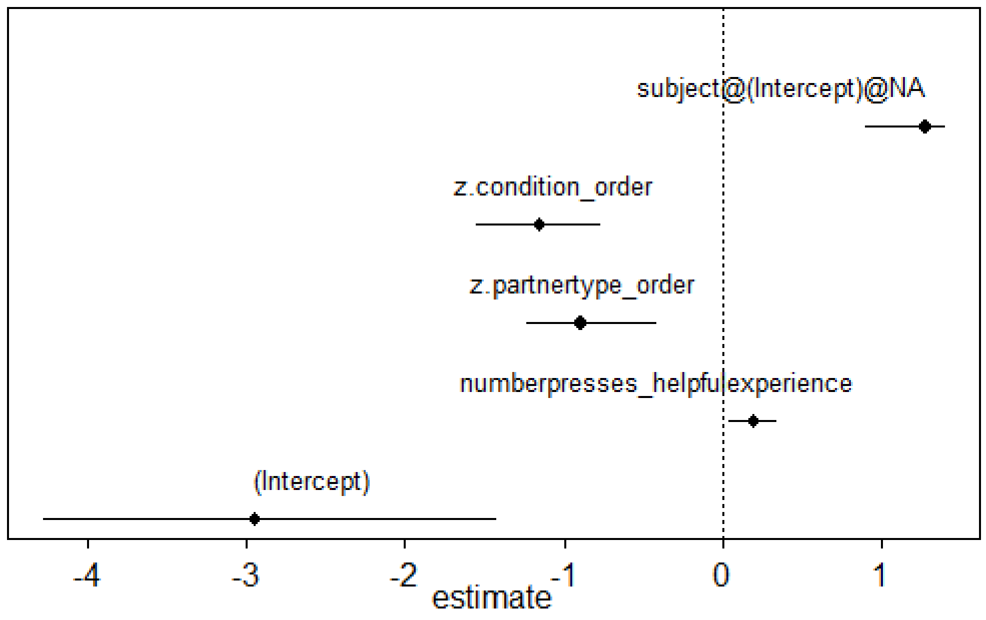
Supplementary material

Supplementary table 1: Sex, age and breed of the 12 subjects plus the corresponding 24 social partners. Castrated individuals are marked with a star (*) behind their sex.

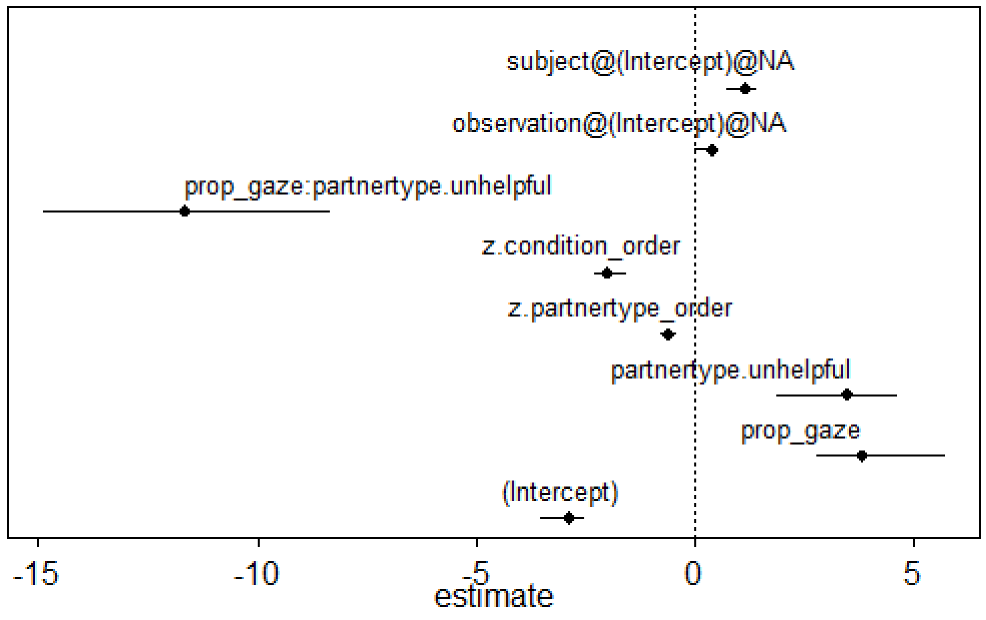
	Role	Sex	Age	Breed
1	**Subject**	**female**	**5,5**	**Shetland Sheepdog**
	Helpful Partner	female	7,5	Shetland Sheepdog
	Unhelpful Partner	female*	3,5	Shetland Sheepdog
2	**Subject**	**female***	**7,0**	**Australian Shepherd**
	Helpful Partner	female	2,0	Australian Shepherd
	Unhelpful Partner	male	5,0	Australian Shepherd
3	**Subject**	**male**	**2,5**	**Australian Shepherd**
	Helpful Partner	female	4,5	Australian Shepherd
	Unhelpful Partner	male	5,0	Australian Shepherd
4	**Subject**	**female***	**12,0**	**Pyrenean Sheepdog**
	Helpful Partner	male*	4,5	Pyrenean Sheepdog
	Unhelpful Partner	female*	1,0	Mix
5	**Subject**	**female**	**6,5**	**Whippet**
	Helpful Partner	female	9,5	Whippet
	Unhelpful Partner	female	6,5	Whippet
6	**Subject**	**male**	**8,5**	**Shetland Sheepdog**
	Helpful Partner	male	9,0	Australian Shepherd
	Unhelpful Partner	female	1,0	Mix
7	**Subject**	**female**	**1,5**	**Mountain Spitz**
	Helpful Partner	female*	10,5	Mix
	Unhelpful Partner	male*	7,5	Mountain Spitz
8	**Subject**	**male**	**6,5**	**Australian Shepherd**
	Helpful Partner	female*	10,5	Australian Shepherd
	Unhelpful Partner	male*	13,0	Australian Shepherd
9	**Subject**	**male***	**6,0**	**Canarian Warren Hound**
	Helpful Partner	male*	11,0	Border Collie
	Unhelpful Partner	male*	1,0	Andalusian Hound
10	**Subject**	**female**	**2,0**	**Australian Shepherd**
	Helpful Partner	male*	6,0	Nova Scotia Duck Tolling Retriever
	Unhelpful Partner	male	9,0	Nova Scotia Duck Tolling Retriever
11	**Subject**	**female**	**1,5**	**Australian Shepherd**
	Helpful Partner	female	3,0	Australian Shepherd
	Unhelpful Partner	female	4,5	Australian Shepherd
12	**Subject**	**male***	**2,0**	**Mountain Spitz**
	Helpful Partner	male*	6,5	Mix
	Unhelpful Partner	female*	2,0	Mountain Spitz



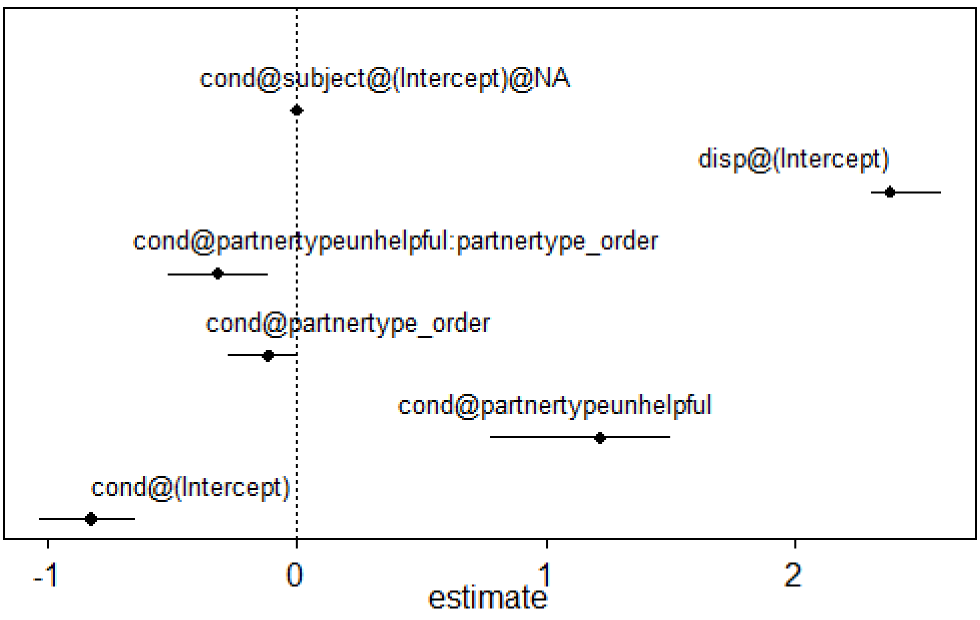
Supplementary figure 1: Stability plot for the generalized linear mixed model relating to question 1. Shown is the range of the estimates for each term of the model when the levels of the random effects are taken out one after another. Random effects are labelled with “@”.



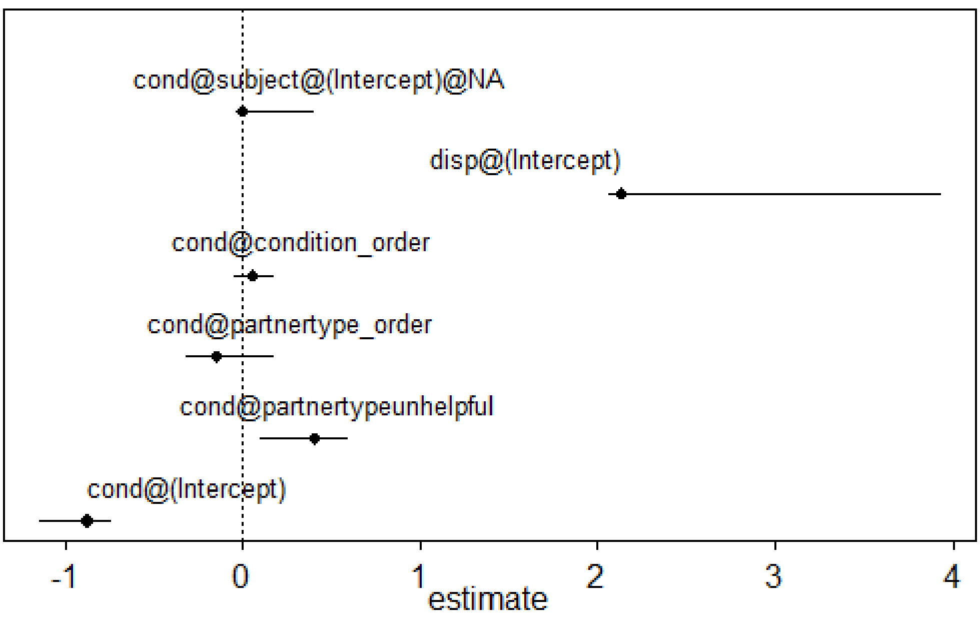
Supplementary figure 2: Stability plot for the generalized linear mixed model relating to question 2. Shown is the range of the estimates for each term of the model when the levels of the random effects are taken out one after another. Random effects are labelled with "@".



Supplementary figure 3: Stability plot for the generalized linear mixed model relating to question 3. Shown is the range of the estimates for each term of the model when the levels of the random effects are taken out one after another. Random effects are labelled with "@".



Supplementary figure 4: Stability plot for the generalized linear mixed model relating to question 4. Shown is the range of the estimates for each term of the model when the levels of the random effects are taken out one after another. Random effects are labelled with "@". The mean part of the model is labelled with "cond", while the dispersion part is labelled with "disp".



Supplementary figure 5: Stability plot for the generalized linear mixed model relating to question 5. Shown is the range of the estimates for each term of the model when the levels of the random effects are taken out one after another. Random effects are labelled with "@". The mean part of the model is labelled with "cond", while the dispersion part is labelled with "disp".

```
> summary (full)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: cbind(numberofpresses, number_notpress) ~ condition * partnertype +
    z.partnertype_order + z.condition_order + (1 + condition.socialfacilitation +
    condition.test + partnertype.unhelpful + z.condition_order +
    z.partnertype_order | subject) + (1 | observation)
   Data: xdata
Control: contr

     AIC      BIC   logLik deviance df.resid
   367.1    435.4   -153.5    307.1       42

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.64129 -0.69389 -0.00928  0.49018  2.54824

Random effects:
 Groups      Name                        Variance Std.Dev. Corr
 observation (Intercept)                 0.00000  0.0000
 subject     (Intercept)                 1.77539  1.3324
             condition.socialfacilitation 0.16865  0.4107    0.77
             condition.test              0.09964  0.3157    0.61  0.44
             partnertype.unhelpful       0.21109  0.4594   -0.59 -0.31  0.23
             z.condition_order           0.05131  0.2265   -0.11  0.45 -0.52 -0.13
             z.partnertype_order         0.16849  0.4105   -0.64 -0.01 -0.38  0.61  0.69
Number of obs: 72, groups:  observation, 72; subject, 12

Fixed effects:
                                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                                  -1.367982   0.450350  -3.038  0.00238 **
conditionsocialfacilitation                  -0.091844   0.320727  -0.286  0.77460
conditiontest                                 0.038319   0.320017   0.120  0.90469
partnertypeunhelpful                          0.001986   0.337051   0.006  0.99530
z.partnertype_order                          -0.191659   0.158620  -1.208  0.22694
z.condition_order                            -0.701355   0.129354  -5.422 5.89e-08 ***
conditionsocialfacilitation:partnertypeunhelpful  0.091693   0.355953   0.258  0.79672
conditiontest:partnertypeunhelpful           -0.375612   0.389198  -0.965  0.33450
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) cndtns cndtnt prtnrt z.prt_ z.cnd_ cndtns:
cndtnsclfcl -0.099
conditintst -0.230  0.517
prtnrtypnhl -0.595  0.280  0.353
z.prtnrtyp_ -0.497  0.004 -0.081  0.293
z.cndtn_rdr -0.020  0.201 -0.166 -0.005  0.309
cndtnsclfc:  0.185 -0.555 -0.283 -0.482 -0.012  0.014
cndtntst:pr  0.179 -0.222 -0.518 -0.484 -0.019  0.064  0.459
optimizer (bobyqa) convergence code: 0 (OK)
boundary (singular) fit: see ?isSingular
```

Supplementary figure 6: Output of the „summary“ function of R for the generalized linear mixed model relating to question 1. The testing-day order of the partner types (z.partnertype_order), as well as the order of the conditions during the testing phase (z.condition_order), were both z-transformed to a mean of 1 with a standard deviation of 1.

```
> summary(full)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: cbind(numberofpresses, number_notpress) ~ numberpresses_helpfulexperience +
    z.partnertype_order + z.condition_order + (1 | subject)
   Data: ydata
Control: contr

     AIC      BIC   logLik deviance df.resid
    68.7     71.1    -29.3     58.7        7

Scaled residuals:
     Min       1Q   Median       3Q      Max
-0.91006 -0.33279  0.05569  0.33166  1.04384

Random effects:
 Groups  Name        Variance Std.Dev.
 subject (Intercept) 1.607    1.268
Number of obs: 12, groups:  subject, 12

Fixed effects:
                                Estimate Std. Error z value Pr(>|z|)
(Intercept)                      -2.9892     2.4685  -1.211   0.2259
numberpresses_helpfulexperience   0.1958     0.2807   0.698   0.4854
z.partnertype_order              -0.9136     0.4956  -1.843   0.0653 .
z.condition_order                -1.1986     0.4838  -2.477   0.0132 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) nmbrp_ z.prt_
nmbrprsss_h -0.985
z.prtnrtyp_  0.386 -0.378
z.cndtn_rdr -0.340  0.357 -0.045
```

Supplementary figure 7: Output of the „summary“ function of R for the generalized linear mixed model relating to question 2. The testing-day order of the partner types (z.partnertype_order), as well as the order of the conditions during the testing phase (z.condition_order), were both z-transformed to a mean of 1 with a standard deviation of 1.

```
> summary(full)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: cbind(numberofpresses, number_notpress) ~ prop_gaze * partnertype +
    z.partnertype_order + z.condition_order + (1 | subject) +      (1 | observation)
   Data: ydata
Control: contr

     AIC      BIC   logLik deviance df.resid
   114.9    124.3    -49.4     98.9       16

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.00625 -0.51463  0.07246  0.46278  1.31826

Random effects:
 Groups      Name        Variance Std.Dev.
 observation (Intercept) 0.1733   0.4163
 subject     (Intercept) 1.3769   1.1734
Number of obs: 24, groups:  observation, 24; subject, 12

Fixed effects:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                    -4.5937     1.3385  -3.432 0.000599 ***
prop_gaze                       9.6711     3.7574   2.574 0.010057 *
partnertypeunhelpful            3.4850     1.5951   2.185 0.028899 *
z.partnertype_order            -0.6017     0.2008  -2.996 0.002737 **
z.condition_order              -1.9783     0.4667  -4.239 2.24e-05 ***
prop_gaze:partnertypeunhelpful -11.6658     4.5208  -2.580 0.009867 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:
            (Intr) prp_gz prtnrt z.prt_ z.cnd_
prop_gaze   -0.949
prtnrtypnhl -0.785  0.806
z.prtnrtyp_ -0.029  0.057 -0.041
z.cndtn_rdr  0.484 -0.498 -0.362  0.054
prp_gz:prtn  0.849 -0.900 -0.962  0.007  0.436
```

Supplementary figure 8: Output of the „summary“ function of R for the generalized linear mixed model relating to question 3. The order of the partner types (z.partnertype_order), as well as the order of the conditions during the testing phase (z.condition_order), were both z-transformed to a mean of 1 with a standard deviation of 1.

```
> summary(full)
 Family: beta  ( logit )
Formula:          prop_gaze ~ partnertype + z.partnertype_order + (1 | subject)
Data: ydata

     AIC      BIC   logLik deviance df.resid
   -15.8     -9.9     12.9    -25.8       19

Random effects:

Conditional model:
 Groups  Name        Variance  Std.Dev.
 subject (Intercept) 1.666e-10 1.291e-05
Number of obs: 24, groups:  subject, 12

Dispersion parameter for beta family (): 10.3

Conditional model:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)           -0.6511     0.1800  -3.617 0.000298 ***
partnertypeunhelpful   0.3981     0.2485   1.602 0.109167
z.partnertype_order   -0.1411     0.1250  -1.129 0.259104
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Supplementary figure 9: Output of the „summary“ function of R for the generalized linear mixed model relating to question 4. The order of the partner types (z.partnertype_order was z-transformed to a mean of 1 with a standard deviation of 1.

```
> summary(full)
 Family: beta  ( logit )
Formula:          prop_gaze ~ partnertype + z.partnertype_order + z.condition_order +      (1 | subject)
Data: ydata

     AIC      BIC   logLik deviance df.resid
   -12.8     -5.8     12.4    -24.8       18

Random effects:

Conditional model:
 Groups  Name        Variance  Std.Dev.
 subject (Intercept) 2.939e-10 1.714e-05
Number of obs: 24, groups:  subject, 12

Dispersion parameter for beta family (): 8.45

Conditional model:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)          -0.97703    0.20782  -4.701 2.58e-06 ***
partnertypeunhelpful  0.40791    0.28132   1.450    0.147
z.partnertype_order  -0.07462    0.14326  -0.521    0.602
z.condition_order     0.04935    0.14446   0.342    0.733
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Supplementary figure 10 Output of the „summary“ function of R for the generalized linear mixed model relating to question 5. The order of the partner types (z.partnertype_order), as well as the order of the conditions during the testing phase (z.condition_order), were both z-transformed to a mean of 1 with a standard deviation of 1.

Zusammenfassung/Abstract

German:

Reziprozität ist ein Konzept, das erklärt, warum nicht-verwandte Individuen ohne einen sofortigen beidseitigen Vorteil miteinander kooperieren (Trivers, 1971). Aktuell haben erst drei Studien Reziprozität in Hunden (*Canis familiaris*) untersucht und fanden widersprüchliche Ergebnisse. Speziesunterschiede, der Familiaritäts-Effekt und Verständnisprobleme wurden als potentielle Ursachen vorgeschlagen.

Diese Masterarbeit ist ein Follow-Up Projekt, um mehr Licht auf die Frage zu werfen, warum Reziprozität in Militärhunden, nicht aber in Haushunden gefunden werden konnte. Die Hypothese wurde aufgestellt, dass prinzipiell auch Haushunde Gefallen von hilfsbereiten, bekannten Artgenossen erwidern würden. Deshalb wurde prognostiziert, dass in einem Futter-spende-Szenario Hunde mehr Futter an hilfsbereite, als an nicht-hilfsbereite bekannte Artgenossen spenden würden.

Dafür wurden 12 Haushunde als Testsubjekte zusammen mit 24 Sozialpartnern getestet. Das Setup bestand aus zwei getrennten Abteilen, mit einem Futterautomaten im einen und einem kabellosen Auslöser im anderen Abteil. Nach einer hilfsbereiten oder nicht-hilfsbereiten Erfahrung mit einem Sozialpartner folgte ein Rollentausch und das Testsubjekt bekam die Möglichkeiten, die Futterspende zu erwidern.

Basierend auf den Ergebnissen wurde geschlossen, dass keine Reziprozität in dieser Studie stattgefunden hatte. Sowohl der Trainingshintergrund, als auch die Trainingsintensität, die Diversität der Stichprobe und/oder fehlende Kommunikationssignale könnten die Hunde daran gehindert haben, reziprokes Verhalten zu zeigen. Daher bleibt die Frage, ob Hunde reziprok kooperieren, weiter bestehen und zukünftige Studien sollten beispielsweise mit modifizierten Setups und einer veränderten Trainingsroutine die Debatte weiterführen.

English:

Reciprocity is a concept that helps to explain why unrelated individuals cooperate without an imminent mutual benefit (Trivers, 1971). So far, only three studies have looked into reciprocity in dogs (*Canis familiaris)* and found partially contradictory results (Gfrerer & Taborsky, 2017, 2018; McGetrick et al., 2021). Species differences, the familiarity effect and problems in comprehension were proposed as possible explanations.

The current thesis is a follow-up project to shed more light on why no reciprocity was found in pet dogs compared to military dogs. It was hypothesized that pet dogs reciprocate help received from helpful unrelated familiar conspecifics. It was predicted that during a food-

donation scenario dogs would donate more food to a helpful compared to an unhelpful familiar partner.

12 subject pet dogs plus 24 social partners were tested. The setup consisted of two separate enclosures with a food dispenser in one and a wireless button in the other enclosure. After a helpful or unhelpful experience with a partner, a role exchange followed and the subject had the opportunity to repay the food donation.

Based on the results, no reciprocity occurred in this study. Training background and intensity, sample diversity and/or missing communication cues perhaps hindered the dogs' ability to show reciprocal behaviour. As this contradicts two previous studies with military dogs, the question of whether all dogs cooperate reciprocally remains. Future studies with for example a modified setup and different training routines could help to clarify this.