

Department of comparative cognition
University of Veterinary Medicine Vienna
Messerli Institute
(Head: Univ.-Prof. Dr.rer.nat. Ludwig Huber)

Do dogs of cooperative and independent working breeds assess the reliability of human informants differently?



messerli
Forschungsinstitut

vetmeduni

Submitted by: Franziska Freudensprung, BSc

Vienna, 05/2023

Supervisor: Priv.-Doz. Zsófia Virányi, PhD. & Lucrezia Lonardo, MSc.

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

.....20.05.2023.....

date



.....

(signature)

STATEMENT FOR DATA PROTECTION

I confirm that all persons who can be seen in the photos included in this work have explicitly provided their consent for the use of the photos. The images will be utilized solely for scientific purposes, specifically for this master's thesis.

.....14.07.2023.....

date



.....

(signature)

Content

Zusammenfassung.....	8
Abstract.....	9
1. Introduction.....	8
1.1. Early domestication and later breed selection.....	8
1.2. Breed comparison studies.....	10
1.3. Reliability assessments.....	13
1.4. Can dogs assess the reliability of human informants?	15
1.5. Limitations of nonverbal paradigms so far.....	17
1.6. Current study.....	19
2. Material and Methods	21
2.1. Subjects	21
2.2. Experimenters	23
2.3. Experimental setup.....	24
2.4. Procedure.....	26
2.4.1. Warm-up.....	26
2.4.2. Training.....	26
2.4.3. Familiarization.....	27
2.4.4. Demonstration phase.....	27
2.4.5. Transfer tasks.....	32
2.4.5.1. Screen choice task:.....	32
2.4.5.2. Pointing task:.....	34
2.4.5.3. Begging task:.....	36
2.4.5.4. Unsolvable task:.....	37
2.5. Video scoring	40

2.6. Statistical analysis	41
2.6.1. Cue following model (Demonstration phase)	41
2.6.2. Cue following model - first exposure (Demonstration phase)	43
2.6.3. Hesitation model (Demonstration phase).....	44
2.6.4. Choice of informant (Transfer phase)	45
2.6.5. Proximity to informant (Transfer phase).....	46
3. Results	47
3.1. Demonstration phase	47
3.2. Transfer Tasks	55
4. Discussion	59
4.1. Reliability assessment.....	59
4.2. Comparison of breed groups	63
4.3. Novel cue following without ostensive communication	65
Conclusion.....	67
Acknowledgment	68
References	69

Zusammenfassung

Die vorliegende Studie befasste sich mit der Frage, ob Hunde die Zuverlässigkeit menschlicher Informanten beurteilen können und ob sich kooperativ und unabhängig arbeitende Hunde in dieser Hinsicht unterscheiden. Vergangene Studien haben gezeigt, dass sich verschiedene Rassen in unterschiedlichen Aufgaben aufgrund ihres Arbeitsstils unterscheiden: kooperative Rassen wurden ausgewählt, um in direktem Kontakt mit Menschen zu arbeiten, während von unabhängigen Rassen erwartet wird, dass sie eher allein arbeiten. Ob und wie sich diese Rassen in ihrer kognitiven Fähigkeit unterscheiden, wird auch noch heute diskutiert. Daher soll diese Arbeit durch die Überprüfung von Glaubenssätzen weitere Erkenntnisse liefern. Frühere Studien hatten gezeigt, dass Hunde von zwei menschlichen Informanten denjenigen wählen, dessen Hinweise zuverlässiger sind. Jedoch zeigen diese Studien nicht, ob Hunde die Zuverlässigkeit der Person an sich einschätzen können. In der aktuellen Studie haben wir ein Experiment mit drei neuartigen Hinweisen (Handlungen) entworfen, die keine gemeinsamen Komponenten hatten, um sicherzustellen, dass die Hunde kritische Urteile über die Informanten und nicht nur über deren Hinweise abgeben. In drei Demonstrationsphasen wurde untersucht, ob Hunde die Handlungen der Zuverlässigen und Unzuverlässigen generalisieren und im Laufe des Versuchs aufgeben den irreführenden Hinweisen der Unzuverlässigen Informantin zu folgen. Zusätzlich haben wir in einer Serie von vier verschiedenen Präferenztests untersucht, ob die Hunde in neuen Versuchsanordnungen eine Präferenz für die zuverlässige gegenüber der unzuverlässigen Informantin zeigen. Dies soll zeigen, ob Hunde ihre bisherigen Erfahrungen mit den Informanten auf völlig neue Kontexte übertragen können. Entgegen unseren Erwartungen fanden wir keine Hinweise darauf, dass Hunde zwischen verlässlichen und unzuverlässigen Informanten unterscheiden können. Außerdem wurden in dieser Studie keine Unterschiede zwischen den Rassengruppen festgestellt. Vielmehr folgten Hunde, unabhängig von ihrer Rasse, in etwa 80 % der Durchgänge den menschlichen Hinweisen, unabhängig von der Zuverlässigkeit des Informanten. Die Leistung der Hunde bei diesen Aufgaben wird in Bezug auf die mögliche Rolle von ostensiven Hinweisen und die Fähigkeit der Hunde zur Generalisierung auf neue Kontexte diskutiert.

Abstract

The present study addressed the questions whether dogs can assess the reliability of human informants and whether cooperative and independent working dogs differ in this regard. Former studies found differences in diverse contexts between dogs of different breeds characterized by different working styles: cooperative breeds have been selected to work in direct contact with humans whereas independent breeds are expected to work rather on their own. Whether and how these kinds of dogs differ in their cognition is still debated; here I proposed that testing their belief revision could provide us with further hints in this regard. Former research had suggested that dogs can select of two human informants the one whose cues have been more accurate but these studies do not tell us whether the dogs can assess the reliability of the persons themselves. In the current study, we designed an experiment with three novel cues (actions) that had no common components, thereby ensuring that dogs had to make critical judgements about the informants and not only about their cues. Across three consecutive stages of a demonstration phase, we investigated whether the dogs can generalize across the different actions of the reliable and unreliable informants and give up on following the uninformative cues of the unreliable informant as the experiment progresses. Additionally, in a series of four different preference tests, we have investigated whether the dogs show a preference for the Reliable over the Unreliable informant in new setups in order to see if dogs are capable to transfer their previous experiences with the informants to completely new contexts. Contrasting our expectations, we did not find evidence that dogs would differentiate between Reliable and Unreliable informants. Further, no difference between breed groups occurred in this study. Rather, dogs, independent of their breed type, followed the human cues in around 80 % of trials, irrespective of the informant's reliability. Dogs' performance in these tasks is discussed in relation to the potential role of ostensive cues and dogs' ability to generalise to novel contexts.

1. Introduction

1.1. Early domestication and later breed selection

Dogs (*Canis familiaris*) represent the domesticated species that has lived the longest in the vicinity of humans. Nowadays it is impossible to imagine human beings without dogs. Dogs play an important role in many areas: they serve as hunters, guard dogs, herd dogs, sled dogs or mainly simply as companions (family dogs). This long-lasting relationship started at least 13.000 years ago with wild wolves (Serpell, 2017).

Different theories exist how this relationship developed. Miklósi and colleagues (2018) summarized different hypotheses how domestication might have taken place. According to the first theory, domestication might have started with humans picking up wolf puppies and raising them in their camps. In contrast, the second theory of self-domestication has proposed that wolves were first pre-selected by eating humans' food remains and living close to humans. Third, others have argued against this scrounging hypothesis, and suggested that it was rather humans who were attracted to wolves and went out to hunt together with them. The truth might lie in between all these hypotheses (Miklósi et al., 2018). Also, in regard to the timing and location of domestication, different papers make different suggestions, even if they agree that domestication started a long time ago: at one point between 13.000 and 45.000 years ago (Serpell, 2017). Although every year new findings and methods are published about the domestication process, it is still unclear where, when and how exactly dogs' domestication took place (Frantz et al., 2020). Despite of this diversity, all hypotheses agree that dog domestication had several stages: first, ancient dogs emerged from wild wolves by natural selection favouring basic characteristics that allowed for living close to humans without conflicts. Only later on artificial selection for breeds started to shape dogs' capabilities and appearances (Parker et al., 2017).

The first important characteristics of dogs likely were being less fearful and less aggressive towards humans and/or being able to communicate and cooperate with them (Hare & Tomasello, 2005; Hare et al., 2012; Miklósi et al., 2018). As such, dogs were selected for their behaviour and cognitive abilities, rather than their appearance, at these early stages of domestication (Serpell, 2017). It has been suggested that, as part of this, dogs have evolved increased attentiveness towards us (Miklósi et al., 2003) and they see us as social partners (Gácsi, Gyoöri, et al., 2009). In addition, their general social competence towards humans have improved (Miklósi & Topál, 2013), including forming a close and dependent relationship

with us and reading our gestures. For example, dogs are better in following pointing gestures than wolves, even if both were reared under the same conditions (Virányi et al., 2008).

Such early selection suggests that domestication has prepared all dogs to interact, communicate and cooperate with humans. However, even if so, not all modern breeds cooperate with humans in the same way. There are breeds, developed during later stages of domestication, which were selected for closer work with humans, and, therefore, have a higher attentional state to their human partner. A visual contact with human is needed when dogs work as herding or gun dogs. In contrast, other dogs were selected to work independently: they can work without visual contact with humans, like earth dogs and guarding dogs: they need to work more independently and more goal-driven (Gácsi, McGreevy, et al., 2009). The history of such breeds started much later during the course of domestication: when humans began to breed dogs for diverse functions (Serpell, 2017).

The earliest appearance of dog breeds in Europe can be dated to 1300, but it is quite possible that they existed much earlier. Hunting is likely one of the earliest functions of dogs, even though sighthounds and mastiffs also belong to the oldest breeds (Serpell, 2017). Only later on, in the middle of the 19th century, people began to breed dogs for their appearance and not for their working functions. As the Victorians gave the basis of the modern breeds and highlighted the parental lineage of a dog and how they must look like, the function of the breed was often not prioritised here (Serpell, 2017). However, breeds that we know today were further selected 200 years ago to achieve certain looks and embed desired behaviours (Parker et al., 2017). Today the Fédération Cynologique Internationale (FCI) lists 356 breeds which can be divided into ten groups, like Group 1 (Sheepdogs and Cattle dogs) or Group 3 (Terriers) (Fédération Cynologique Internationale, 2022). Modern dogs are now kept because of their look in the first place, for the colour, fur and size they have. Nevertheless, even today breeds can strongly differ in their behaviour. Border collies do not only herd sheep because they grew up on a farm: it is also genetically anchored in them, which is due to the selection of the different breeds for their working purpose (Parker et al., 2010). Indeed, several genetic and behavioural studies have shown that certain behaviours of different dog breeds are heritable.

Dutrow and his colleagues (2022) analysed genetic and owner reported behavioural data from different dog breeds and from feral dogs. This study identified ten different genetic lines which are connected with breed typical behaviours. Terriers scored high on prey chasing behaviour, which supports their goal-directed behaviour and independent work from humans. Similarly, scent hound lines scored low on trainability, which also indicates that these dogs work more independently and without relying on human-given cues. These breed linked

behaviours were most significant for herding, terrier and scent hound lines, in respect to their non-social fear. As these breeds are more sensitive to their surroundings, they might be more prone to movements in the environment. Herding dogs are sensitive to detect behavioural changes in farm animals, and terriers are focused on detecting prey under the ground (Dutrow et al., 2022).

One recent paper from Junttila et al. (2022) focusing on dogs' cognitive abilities also demonstrated differences between certain breeds. With several tests (such as following human communicative as well as misleading gestures, spatial problem-solving ability in a V-detour task, inhibitory control in a cylinder test, and persistence and human-directed behaviour during an unsolvable task), they found breed differences especially in inhibitory control and social cognition. To compare their inhibitory control, dogs had a demonstration how to get food from a lying opaque cylinder by reaching inside at its ends. Later the cylinder was presented in a transparent form so that the dogs could see the food; now the question was whether dogs want to access directly through the plexiglass or from the sides, as learnt before. The study found that herding dogs (like Border Collies and Australian Shepherds) fared better in inhibitory control than guard dogs (like German Shepherd or Malinois). The authors suggested that this reflected the special functions the breeds were selected for: as herding dogs need to suppress their reaction to potential prey they had better scores in inhibitory control than guard dogs whose job is to attack and scare away potential intruders (Junttila et al., 2022).

These studies as well as dog breeding practices show that dog behaviour can be selected for, which means that there is a genetic basis for certain behavioural traits (Dutrow et al., 2022; Parker et al., 2017). Even though the today diversity of numerous modern breeds does not have a long history, it is quite likely that the selection of dogs for a few different functions started a long time ago. As dogs differ from wolves primarily in behaviours that facilitate living and working together with people, it is likely that selection for these skills was one of the earliest sources of diversification of dog breeds. If so, these behavioural traits may have crystallized more and more for specific tasks later on, leading to a diversity of different breeds that work either in close contact with humans or more independently.

1.2. Breed comparison studies

Gácsi and colleagues (2009) were the first ones who proposed to divide dogs in two main categories which describe dogs' working style and can be named as independent and

cooperative breed groups. Cooperative workers, such as herding dogs for instance, were bred to work close with the humans, follow their instructions and always check back to the human to get new tasks. These dogs can work well when having an ongoing visual contact with humans. In contrast, independent workers, such as scent hunting dogs, are more goal-driven and independent while chasing a fox for instance, and they need less visual contact to humans (Gácsi, McGreevy, et al., 2009). As the different working tasks require different levels of human contact, these different working functions predict different abilities to read human-given cues and signals. Following up on this distinction and its relatively clear predictions, several studies have compared dogs of cooperative and independent working breeds in cognitive and behavioural tasks.

As such, (Gácsi, McGreevy, et al., 2009) and (Wobber et al., 2009) found that breeds used for closer work with humans are better in using human communicative cues (but see Pongrácz et al., 2005 for contrasting findings). Pongrácz and colleagues (2020) showed that cooperative breeds are more dependent on their owner when they are left alone during a separation test than independent breeds. Cooperative breeds barked more when their owner was absent and they had a higher intensity to look for their owner than independent breeds (Pongrácz et al., 2020). On the other hand, there are also studies that found no difference between cooperative and independent working breeds in this regard. Lenkei and her colleagues (2021) showed that there was no significant result in bonding behaviour between the two different working groups in the Strange Situation Test.

Regarding inequity aversion, McGetrick and his colleagues (2020) tested dogs in three different pawing conditions. During the first condition, both dogs were equally rewarded, if they gave their paw to the experimenter. In the second condition, only one dog received food for giving his paw, the other was asked to work without a reward. In the third condition, the dogs did not have another dog partner: it was tested how long they are willing to work on their own if they get no food reward for giving their paw. After that, dogs were tested for social tolerance: both could eat from one bowl at the same time, and it was also tested how long the dogs sought proximity to their owners. This study found no difference between the two breed groups in regard to their inequity aversion: they worked similarly in the unfair version of the pawing task and behaved similarly with their dog partner when they were fed next to each other. The two groups, however, behaved differently with their human partner: when having to give the paw alone with a reward, cooperative breeds worked longer than independent breeds, and they also spent more time next to the owner than independent breeds (McGetrick et al., 2020).

Another study by Heberlein et al. (2017) compared to what extent dogs of independent and cooperative breeds (and dogs bred for companionship, so-called family dogs) take into account what their owner can see. At the beginning of this experiment, dogs were trained not to eat food from the floor. Afterwards, during the test setup, one food side was hidden from the owner's point of view and not for the dog, but the other side was covered by a transparent Plexiglas, leaving the food visible for both dog and owner. Now, the experimenters wanted to know if dogs took the owner's perspective into account: as dogs were still forbidden to take the food, it would be better to choose a food spot which is not visible to the owner. The results showed that independent breeds and family breeds are more sensitive to what their owners could see, because they ate first from the location where the owner could not see the food. In contrast, cooperative breeds showed no preference for feeding first from the area that was not visible for the owner (Heberlein et al., 2017). This study suggests that independent breeds may be better than cooperative breeds at predicting human behaviour in competitive situations; that is, when the task is to decide when it is safer to disobey. Another, recent, study seems to confirm this finding but raises also another possibility.

The 2021 study by Lonardo and colleagues tested whether dogs have an idea of what a person might or might not know because she was not present when an important event took place. The experiment was conducted in a small room with two containers. There were two experimenters: one person was hiding food, the other one was a so-called "communicator". Besides the two experimenters, the owner with their dog sat in the room. The dogs and the communicator could always watch how the first experimenter hid some food in one of the two containers. Then the communicator either left the room or remained inside when the first experimenter moved the food from the first to the second container. The dogs could always observe this event. Depending on her presence, the communicator could either have a true belief, if she saw the food replacement, or had a false belief, if she had been absent during the food replacement. Independent of her belief, in both conditions the communicator then always suggested the container where the food was first (now empty container) to the dog to choose. After this suggestion, the dogs were released to make a choice: they could either search the container where the first experimenter placed the food last (baited container) or they could investigate the (empty) container the communicator (either knowing or not that this container was empty) suggested them to choose. Also in this experiment dogs that belong to independent and cooperative working breeds behaved differently. As such, terriers that belong to the independent workers behaved more likely as young children and followed the communicator's suggestion more often when the communicator suggested the empty

container knowing that it was empty (that is, she had a true belief about the situation) than in the false belief condition. In contrast, Border collies behaved in the opposite way, and followed the communicator's misleading suggestion more often in the false belief condition (that is, when the informant did not know about the replacement of the food) than in the true belief condition (although the difference between conditions for Border collies was not statistically significant). Different possibilities might explain the findings. First, it could be that the terriers (independent breeds) followed the communicator in the false belief condition less than the border collies (cooperative breed), because terriers can "think more independent" of the human informant and realize easier that she was mistaken. Such a difference would be in line with the findings of Heberlein and colleagues (2017) who found that independent dogs took their own and their owner's differing perspectives better into account than cooperative dogs. Alternatively, however, it is possible that cooperative dogs can as well as independent dogs evaluate humans' past experiences but are more inclined to follow their communicative cueing even if it is (accidentally) faulty. However, cooperative breeds, like Border collies, might be more sensitive to deception than independent breeds, and may, in the true belief condition, respond stronger to the misleading informant when she pointed to the empty container despite knowing that this one was empty (Lonardo et al., 2021). This second possibility would mean that, as already suggested above, cooperative breeds do not lag behind independent breeds in assessing their human partners' perspectives and past experiences but they respond to this stronger when this assessment leads to the detection of misleading intentions.

As I have just summarized, former studies suggest diverse possibilities how cooperative and independent breeds may differ in their assessment of humans' perspectives, past experiences and intentions underlying our communication. The current study aims at further tackling this question, by using a different experimental paradigm addressing reliability assessments.

1.3. Reliability assessments

Humans as well as dogs are social animals; they can gain knowledge from conspecifics and other species. Learning and taking information from others can be very efficient and rewarding: if we are using help from others, achieving our goal can be much easier and faster (Poulin-Dubois & Brosseau-Liard, 2016). However, it can also be the other way around if our partners tend to trick others for their own benefit or provide wrong information unintentionally (Harris et al., 2018; Shafiq et al., 2012). Therefore, it would be worthwhile to be able to distinguish

unreliable information from useful information. To find out which person we can trust, we need to follow their history of trustworthiness in previous events: better to follow those that gave reliable information in our last interaction (Mills, 2013).

Adult humans often use language to make credible claims and listeners treat such statements as reliable or unreliable (Harris & Koenig, 2006). However, gaining knowledge or belief from others is not only possible by language: we can also use nonverbal communication (Harris et al., 2018). Additionally, as human adults, we have already gained a lot of experience with other people in our daily life, which can support such assessments. In contrast, children lack the same amount of experiences and mental capacity, which means that assessing someone's reliability may be for them more challenging (Mills, 2013). Having the skill to assess the reliability of others can obviously help to avoid misleading information. Especially children often have to rely on our actions to learn more about their environment, when observing others or listening to others (Harris & Koenig, 2006). In that case, they should differentiate between good and bad information, very often between reliable and unreliable informants.

Indeed, also young children can detect the trustworthiness of humans with whom they had already former experience (Harris, 2007). This has been tested in a number of paradigms that typically compare the infants' reactions to one or two informants that acted in the past reliably (gave good, correct information) or unreliably (gave bad, misleading information). Given this information, during the test, the question is whether the subjects can decide whom to trust after they got an introduction of both informants (Mills, 2013). If children are supposed to learn new words from other people, there are also age differences in when and how children come to trust a person. In a study by (Luu et al., 2013), one person always said the correct word (accurate) for an object the children already knew whereas the other person always said a wrong word (inaccurate) to a body part (e.g., call the eye arm). In addition, a third person labelled the eye with a completely unknown word (novel) for the children. Three comparisons were made in this study, accurate vs inaccurate; accurate vs novel; inaccurate vs novel, by asking children verbally which person was reliable. The results for the last condition showed that only the oldest children (5-years-old) preferred the experimenter who labelled objects with the novel word over the experimenter with whom they had bad experiences in naming of objects. This was not the case for 3- to 4-year- old children: they did not show an aversion of the inaccurate experimenter compared to the experimenter who named objects with novel words. However, 4- to 5-year-old children trusted the experimenter more from whom they had already received reliable information over the experimenter who said novel words. In contrast, the youngest group (3-year-olds) only preferred the accurate person, when they had the choice

between accurate versus inaccurate. Whenever there was the experimenter who said novel words, they did not show preference to the accurate person over a person with novel words or an aversion to the inaccurate over the person with novel words (Luu et al., 2013). The study shows that it is crucial how old the children were when they were confronted with a new situation. It looks like that children were capable to distinguish between two persons with whom they had experience, but if it comes to a novel condition, the youngest had difficulties to apply the accuracy to the known person. However, it is difficult to test whether pre-verbal children have the ability recognise the reliability of other people.

Another study has shown that, if a person had pretended that there is a toy behind a barrier, but in reality, there was nothing there, 14-month-old children followed the gaze direction of this person less often than that of a reliable person who had given a correct sign that there is a toy behind a barrier (Chow et al., 2008). In an eye-tracking study, 8-month-old children got 100 % correct cues from a reliable person, but 75 % misleading cues from another, unreliable, person. This study also found evidence that children followed the gaze of the reliable person more and searched longer in a box that the reliable person had cued, in comparison to the unreliable informant (Tummeltshammer et al., 2014). Therefore, the authors suggest that children can evaluate the reliability of informants and adjust their behaviour to that specific person. Importantly, these studies have shown that also non-verbal methods can be used to test reliability assessment and have demonstrated that pre-verbal infants can use previous cues of different informants and behave accordingly. As we cannot simply ask non-verbal infants to report the reliability status of one person, we have to create tasks where we can derive their cognitive processes from their behaviour. Importantly, such methods can then possibly be used to investigate similar abilities of non-human animals as well.

1.4. Can dogs assess the reliability of human informants?

In the current study that I will introduce here, I investigated whether dogs kept as companions can assess the reliability of human informants. As dogs are social animals that live very close to humans, it is easy to see that it would be an advantage for dogs if they could judge whom they trust and which person they should better avoid (Dwyer & Cole, 2018). Similar to non-verbal infants, dogs cannot use human language to express their thoughts; however they are capable of using our commands, speech, gaze, pointing and other human signals (Hare & Tomasello, 2005). As they often use our human-given cues, it is an interesting

question if they can detect the reliability of different persons, and respond less to unreliable informants. Some studies suggest that dogs follow human cues, even if they know that they were misled (Prato-Previde et al., 2007, Szeteci et al., 2003) whereas other studies showed that dogs can learn that there are dishonest pointing cues from a human (Petter et al., 2009). In the first experiment of this later study, there was a cooperative and a deceptive informant: the first pointed to the container with reward while the second misled the dog. The dogs approached the cooperative pointing significantly more often than the deceptive pointing. These results occurred when more trials were performed in a row by one experimenter during the first four sessions. This has however changed in the last session where the experimenters switched more frequently in a random order; in this case the dogs no more differentiated between them. Therefore, the authors speculate that it was more confusing for the dogs to maintain the human distinction, when they switched more often.

Using non-verbal paradigms similar to those applied in human infants, dogs have also been tested for their abilities to respond to the reliability of a person. One study (Takaoka et al., 2015) tested if dogs can flexibly follow or ignore human pointing after having seen where food had been hidden and whether a person pointed to this or another location. In this study, the informant was always the same person who made either correct or misleading suggestions across three different phases on the experiment: she gave in phases one and three reliable cues where food could be found, but in phase 2, her pointing cue was obviously misleading because the dog could see that the food had been hidden in the other food location. Importantly, after experiencing the misleading 2nd phase, in phase 3, dogs tended to take longer for deciding in later reliable trials with the same person than in phase 1. This suggests that they recognized the misleading behaviours of humans and also reacted to that behaviour, as they did not follow her pointing that quickly anymore (Takaoka et al., 2015).

Another study (Pelgrim et al., 2021) tested whether dogs can differentiate between two persons who provide differently reliable information. They found that dogs followed significantly more often a trustworthy person than the person who had been misleading in the past; her cue, the dogs more often ignored. Before interpreting the results of this study though, it is important to look into its methods in more details. In its most relevant third experiment (Figure 1) the dogs participated in two visits. First, they got a familiarization phase to learn and get used to getting the food under one cup. Later, two experimenters sat close to each other and gave conflicting points to the baited (reliable) and the unbaited (unreliable) clear dish. This phase gave the dog the opportunity to learn which person is pointing to the correct place and which person was misleading. After the dogs received this information, they were tested

without visible access to the reward, as they now used opaque dishes. The question was whether they would use the previous history of the two informants' reliability and chose the dish pointed to by the reliable experimenter even when not seeing the food under this dish. Dogs chose significantly above chance the accurate person in history trials and also more often during the test trials, when they could not see the food. Therefore, the authors conclude that dogs can apply past information about people's accuracy to new situations (Pelgrim et al., 2021).

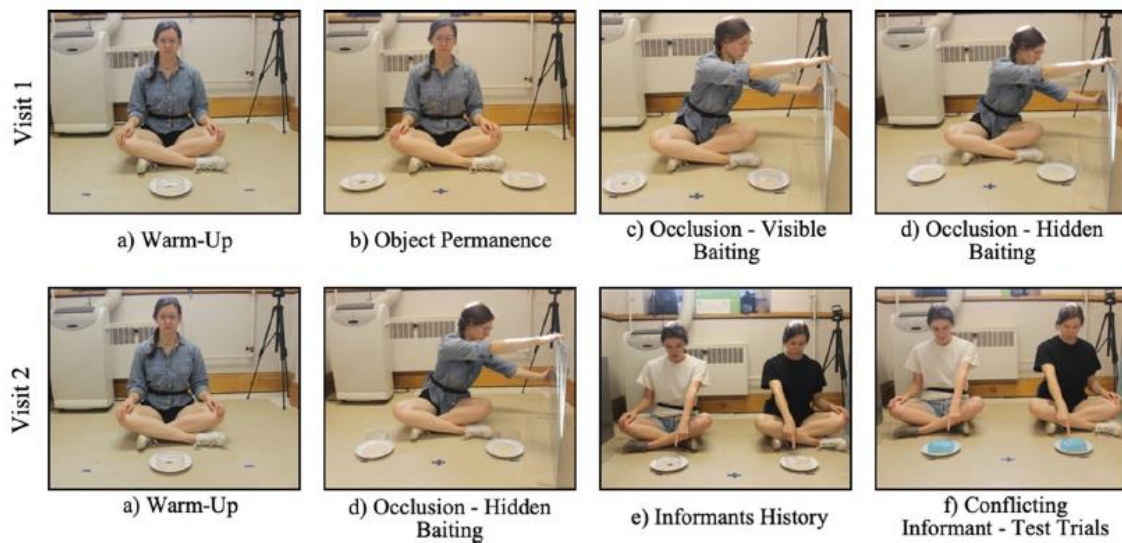


Figure 1: “Fig. 5 Procedures for the phases of Experiment 3. In Visit 1 dogs completed, in order, **a** warm-up, **b** object permanence, **c** occlusion visible baiting, and **d** occlusion-hidden baiting. In visit 2, dogs in order, **a** warm-up, **d** occlusion-hidden baiting, **e** informant history, and **f** conflicting informants’ test trials” (Pelgrim et al., 2021, p. 291).

1.5. Limitations of nonverbal paradigms so far

Note that in principle all non-verbal studies described above aim at investigating whether animals and children can make inferences about the reliability of *persons* they interact with based on the accuracy of the cues they have provided. Here I argue, however, that in these former studies the subject may have shown a preference for the accurate cue rather than for the reliable informant because they always used the same experimental setup and informant cue during training and test (e.g., pointing, gazing). For instance, in a gaze following study by Tummeltshammer et al., 2014, 8-month-old infants were tested if they preferentially follow the head movement of a person that reliably indicated where something interesting will show up, as compared to another person who sometimes looks to a wrong direction (Figure 2). They

found that infants indeed had a longer looking time for the suggested picture box of the reliable person, at the same time, this was not the case for the unreliable person. Children looked less to the boxes which was indicated by the unreliable person. However, in this study the very same photos of the two persons looking sidewise were used in the training and test conditions. The only novelty in the test condition was that these faces oriented to a new direction. Therefore, one can argue that the infants did not learn about the reliability of the people depicted on the photos but simply learnt which photo is a better cue to predict the occurrence of interesting pictures.



Figure 2: "Fig. 1 Examples of familiarization trial blocks containing four reliable face cues (left) and four unreliable face cues (right) from Experiment 1. At the start of each trial, one of two female faces appeared and looked toward one of the boxes. An animal then appeared in one of the boxes. The reliable face always looked to the correct box (i.e., where the animal would appear), but the unreliable face looked to the correct box on only one out of every four trials" (Tummeltshammer et al., 2014, p. 3).

The dog studies summarized above also suffer from the same limitations because their training session and test phase also used identical cues. For example, Takaoka et al., 2015 and Pelgrim et al., 2021 had in their studies always the same pointing cue to a baited or unbaited cup. This cue included the same finger, pointing to a container, with the same person positioned and behaving in the same way behind the pointing gesture. That is, the cue remained the same throughout the experiment, and there was no need for the subjects to think about the reliability of the persons that remains the same in whatever way they provide information. All these studies could be solved by learning about the *misleadingness of the cue*, without thinking of the *reliability of the persons*. Therefore, the question remains whether dogs

can assess the reliability to persons *across different contexts and independent of the cue* they use to provide information.

1.6. Current study

Our experiment aimed at addressing this question by providing dogs, across three subsequent phases, with information about three different cues that had no common component, and investigating whether dogs can generalize across cues what they learnt about the differential reliability of two persons. Moreover, we have also tested whether they can transfer their experiences with these people to completely novel tasks. This means that the dogs had the possibility to use their experiences in new contexts and choose to interact with one or the other person based on what they had learned about them beforehand in a different context.

The informants had either a reliable or an unreliable role. The reliability of the informants in our study was defined by how often the subjects could find food where the informant had previously done something. Each of one of three experimenters (two reliable and one unreliable) alternated giving cues to a dog where to find food reward hidden behind two screens. The reliable persons acted in every trial on the location where then food could be found. In contrast, the unreliable person cued only in 50 % of her trials the correct location of the food. In the other half of her trials, her cues were misleading because the food was not there where she had acted but on the other side. This novel setup was used to evaluate if the dogs could learn which person was reliable and which person was not trustworthy because she sometimes gave a wrong cue. Therefore, we expected that if dogs were able to detect the unreliable informant, they should follow the unreliable person less and less as the session progressed, as this person was misleading in previous trials. Furthermore, we expected that the dogs' hesitation length before making a choice would soon be higher with the unreliable person, as the dogs would lose confidence in following that person. As trust in the unreliable person should decrease, I provided the dogs with an opportunity to look through a gap before making a decision to check if the food was where the unreliable person had acted before. Therefore, over time, peeking through the gap with the unreliable person should be used more often than with the reliable person. Finally, if dogs could assess the reliability of the two informants, they should show a preference for the reliable informant also in novel contexts, e.g. ask for food or for help from the reliable, rather than the unreliable, one.

With this study, I want to conduct further research on belief revision in dogs. As discussed earlier, Lonardo and colleagues (2021) have already raised the question of whether

cooperative and independent breeds perceive and interpret human beliefs differently. They made two proposals in this regard that they could not differentiate between based on their results: Are cooperative breeds more prone to reject deception or are independent breeds better in making choices independent of faulty human cues?

I suggest that the current study has not only the potential to investigate whether dogs can assess the reliability of human informants but also to investigate how cooperative and independent breeds differ to each other in this regard. More specifically, our hypothesis is that dogs in general will show a preference for the reliable informant over the unreliable informant, but this preference may show up with a different strength in our two breed groups. We have three hypotheses in this regard:

(1) If selection for cooperative working made cooperative working dogs more sensitive to deception, dogs belonging to cooperative working breeds may better differentiate between reliable and unreliable informants than independent working dogs because they will sooner give up on following the actions of the unreliable informant than dogs of independent breeds.

(2) If selection for independent working made independent working dogs more prepared to ignore human behaviour, dogs belonging to independent working breeds may better differentiate between reliable and unreliable informants than cooperative working dogs because they will sooner be able to ignore the actions of the unreliable informant than dogs of cooperative breeds.

(3) If selection of both kinds has taken place, dogs belonging to both cooperative and independent breeds will differentiate between reliable and unreliable informants to a similar extent.

2. Material and Methods

2.1. Subjects

In total we tested 61 dogs kept as companions in private homes. Thirty-two dogs belonged to cooperative breeds (for breed distribution see: Table 1) and 29 dogs belonged to independent breeds (Table 1). Due to lack of motivation five of the independent breeds and two of the cooperative breeds were excluded during testing. The final sample consisted of 54 companion dogs, divided into 30 dogs of cooperative breeds and 24 dogs of independent breeds. The sex of the whole sample size was equally distributed, 50 % (n=27) of males and 50 % (n=27) of females. However, there were more males (n=18) than females (n=12) in the cooperative group whereas more females (n=15) than males (n=9) in the independent group. The mean age of all dogs was 6.1 years with a standard deviation (SD) of 3.3 years. The two groups separately had a mean of 6.7 years with a SD of 3.4 years for the cooperative ones and a mean of 5.5 years with a SD 3.1 years for the independent group.

Table 1: Number of participants of the breeds of the respective breed group, abbreviations and papers that have categorized these breeds as "cooperative" or "independent".

Breed group	(n=)	Breeds	Short	Papers that labelled these breeds as "cooperative" or "independent"
Cooperative (n=32)	10	Australian Shepherd	AS	(Gácsi, et al. 2009; Heberlein et al., 2017; McGetrick, et al. 2020)
	1	Australian Kelpie	AK	(Gácsi et al., 2009)
	13	Border Collie	BC	(Gácsi et al., 2009; Kovács, et al. 2016; McGetrick, et al. 2020; Lenkei, et al., 2021)
	2	Collie	C	(Heberlein et al., 2017)
	6	Shetland Sheep Dog	SSD	(Gácsi, et al. 2009; Heberlein, et al. 2017)
Independent (n=29)	2	Bedlington Terrier	BLT	(Gácsi et al., 2009)
	7	Border Terrier	BT	(Heberlein et al., 2017)
	2	Fox Terrier	FT	(Lenkei et al., 2021)
	0	Irish Terrier	IT	(Heberlein et al., 2017)
	4	Irish glen of Imaal Terrier	IGIT	
	1	Irish soft-coated Wheaten Terrier	IWT	
	3	Jack Russell Terrier	JRT	(Gácsi et al., 2009)
	8	Parson Russell Terrier	PRT	(Gácsi et al., 2009)
	1	Welsh Terrier	WT	(Gácsi et al., 2009)
	1	West Highland Terrier	WHT	(Gácsi, et al. 2009; Heberlein et al., 2017; Lenkei et al., 2021)

The breeds that I categorized as cooperative and independent are listed in (Table 1). My criteria for categorization were that cooperative breeds needed to belong:

- (1) to the same genetic clade as the Border Collie, as listed in Parker et al. (2017), and
- (2) to the subsection "sheepdog" of FCI group 1, and
- (3) mentioned as "cooperative" in at least one peer-reviewed published paper other than Lonardo et al. (2021).

Similarly, in the independent working group I included breeds that belong:

- (1) to the same genetic clade as the Parson Russell Terrier (the terrier breed most frequently represented in Lonardo et al. (2021)), as listed in Parker et al. (2017), and

- (2) to the FCI group 3, and
- (3) mentioned as “independent” in at least one peer-reviewed published paper other than Lonardo et al. (2021). (Exceptions from this criterion are the Irish Glen and the Irish Soft-Coated Terriers that have so far been identified as independent breeds only by Lonardo et al. (2021). We still included them because they are very similar to Irish Terriers, a breed that has been listed as independent also by Heberlein et al., 2017).

2.2. Experimenters

Throughout the whole study, the same three female experimenters acted as informants in three different roles. All of them were of similar age and height, and unfamiliar to the subjects. The different roles were two reliable informants (Reliable A and Reliable B) and one Unreliable informant, whereby the Reliable informants always cued the baited side, whereas the Unreliable informant gave in 50 % of the trials her cue on the unbaited side. We worked with two reliable persons for two reasons. First, we wanted to create a common rule that, similarly to everyday life, humans are in general trustworthy, and the Unreliable person’s behaviour is an exception. Second, the first reliable person introduced every novel cue. As the dogs may have needed a short familiarization with each new step and their responses to the first trials might have been influenced by this novelty rather than the reliability of this person, I did not include the Reliable A trials in the analyses. The three different roles were randomly assigned to the three experimenters prior to the experiment and counterbalanced across dogs. To make it easier for the subjects to differentiate the three informants, the three informants wore differently coloured clothes (Figure 3): one was completely white, one completely black and the third one was black and white striped. The colouring of the clothes was also randomized across the three roles and counterbalanced across subjects.



Figure 3: The three experimenters in three different colours: black, striped and white. Two of them are reliable (Reliable A, Reliable B) and one is Unreliable.

2.3. Experimental setup

The dogs were tested at the Clever Dog Lab, on the campus of the University of Veterinary Medicine, Vienna. The experimental room (Figure 5) had a size of 6m x 7m in total. Two curtains covered one part of the room (2m x 7m) where the informants could hide during the experiment (Figure 4). Further, there were two doors behind the curtain to enter the room. In front of the curtain, in the area visible for the subject, two V-shaped screens were used for hiding the food. The outer side of these screens was longer (151 cm) than their inner side (61cm). The screens had a height of 90cm, so that the subjects could not look behind them from above. However, the two screens had a 2cm gap at the tip of the screen, where the subjects could peek through and check whether food was behind the screen before making a choice. The sides of the rewards were counterbalanced for all dogs, and the order on which side (left or right) the cue was given was randomized, but food could only be found twice in a row of one side.

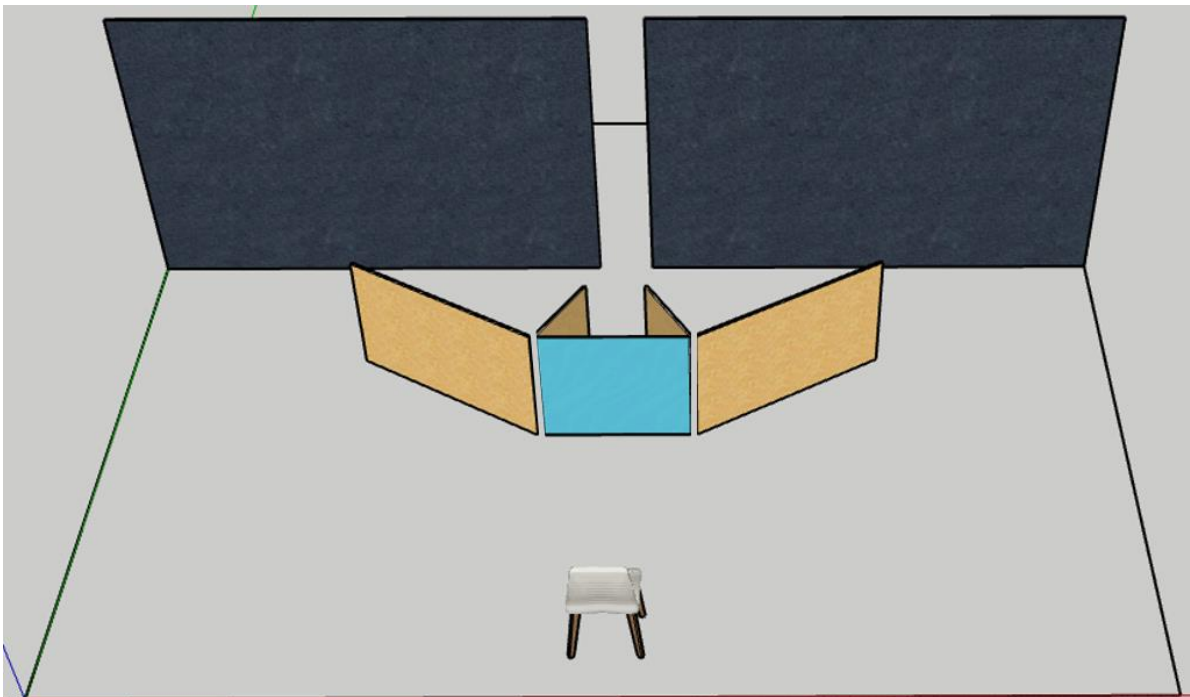


Figure 4: 3D drawing of the setup of the demonstration phase (dark grey: curtains, yellow: two screens, light blue: plexiglass panel forcing the dogs to make a choice of the screens by approaching from the side).

curtain in these trials. For the fourth transfer task, the solvable/unsolvable task, an empty jar lid was screwed to a 70m x 120m large board.

2.4. Procedure

2.4.1. Warm-up

Each experiment started with a familiarization phase in which the subjects could learn, with help of the owner, that food can be found behind the V-shaped screens indicated by the Reliable A experimenter. The owner with the dog on leash stood in the start position in front of the screens. At the same time, reliable A came out from the middle of the two curtains, held the food in front of her belly with both hands and presented it by bringing both hands to her chest while standing in the middle of the two screens and then stepped behind one of the screens. Then the owner was asked to come with the dog to the respective side and the dog could look behind the screen while standing on the side. Reliable A crouched down and put, in full view of the dog, food on the board behind the screen. She never addressed the dog during this entire process (that is, did not call its name and did not look at it). Now, the dog was allowed to take the food and the next trial could start until all four trials were performed. The sides were randomized by a program, named PsychoPy, that was running throughout the experiment on a laptop placed behind the curtain (Figure 5). After this pre-feeding, the owner was asked to sit on the chair across from the screens, with the dog in front of them.

2.4.2. Training

The aim of the training was to make sure that the dogs follow the crouching cue later in the experiment.

After the Warm-up, the owner was asked to remove the leash and hold the dog on their harness or collar. Again, Reliable A appeared in the middle of the curtain, stepped behind one of the two screens, crouched behind it whereby she placed the reward on the board, and then disappeared again between the curtain and the wall, on side of the screen that she had just cued. The owner held the dog until the releasing signal “beep” by one informant was given. At this point the owner released the dog to make a choice and was asked not to interact with the dog. Additionally, the owner could not see what the experimenters were doing, as they had to wear a cap and look to the floor in front of them. Each dog received a minimum of ten training trials, and this continued until the dog followed significantly more than they did not follow, as

determined by a binomial test or they reached maximum of 20 trials. Then a short break of approximately five minutes followed.

2.4.3. Familiarization

After the break, we repeated the warm-up phase. This was done two times on each side in an alternating and counterbalanced manner. In the next step, Reliable A introduced the gap on the tip of the screens. While the dog was held on leash by the owner in front of the screen, Reliable A moved the food up and down at the gap, allowing the dog to watch and follow the food with their nose. Then the dog was allowed to go behind the screen and eat the food. The gaps were presented once on each side.

After completing the last gap training, the owner was asked to wear a cap, sit on a chair and look down at the floor. This was done so that owners could not influence the dog's behaviour with their gaze. During the whole experiment the owner was asked not to interact with the dog as soon as a trial starts. Only when the trial ended and the second "beep" was given, the owner could call their dog back.

Next step was to introduce the three informants to the dog: all three came from one side, appearing one after the other between the wall and the curtain, and lined up in front of the curtain showing themselves simultaneously (Figure 3), which did not happen later during the course of the experiment. They remained standing in front to the screens for five seconds without looking at or talking to the dog, and finally withdrew again behind the curtain, between its other end and the wall.

2.4.4. Demonstration phase

This phase consisted of three parts with three different actions that were presented in a fixed order so that the dog could use them as cues to find food. The actions were presented in a series of trials that are summarized in Table 2. During the first action, the informant came out through the middle of the two curtains, while holding food with both hands in front of her body. The informant stepped left or right, and crouched (Figure 7) down behind one of the two screens. During crouching, she placed a piece of food on the board behind this screen, in all trials in case of the two reliable informants and in 50 % of the trials in case of the Unreliable informant. If the Unreliable person was misleading, she just crouched and took the food secretly with her behind the curtain. Meanwhile, one of the other informants inconspicuously placed another piece of reward behind the other screen from behind the curtain, with the help

of pulling and pushing the board on the floor. If the dog went to the wrong side where no food was hidden, the informants pulled back the baited board on the ground in full view of the subject and removed the food (Figure 6). The informants left the scenery between the curtain and the wall, always on the side where they had crouched. After the owner got the releasing command “beep”, the dog could make a choice. Each of the reliable informants ran four crouch trials, the Unreliable person crouched six times, and at the end of this phase Reliable B did two more crouching trials in order to show the dogs that the cues of only the Unreliable person can be misleading.

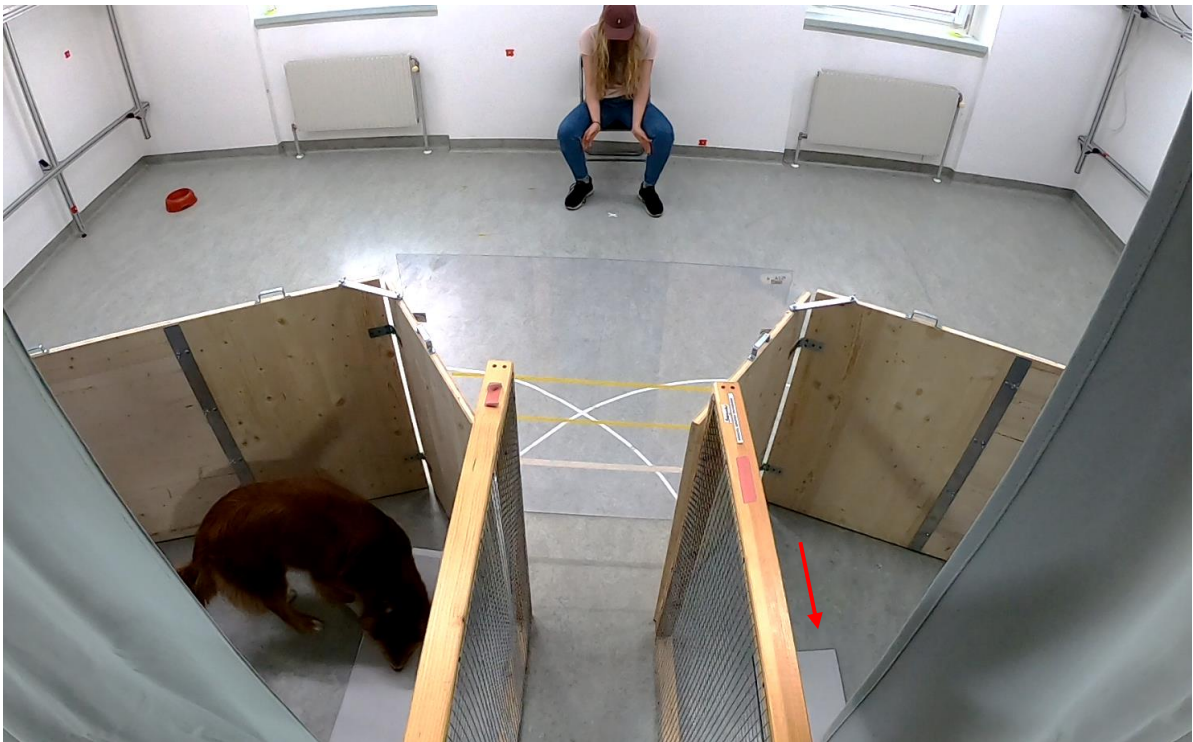


Figure 6: Wrong choice of the dog; food was removed behind the curtain on the other side.

In the second phase, after a short break, Reliable A did again two crouching trials in order to remind the subject that it is not crouching action per se that lost its informative value, but only the Unreliable informant uses it in this way. Afterwards, a new action was introduced, first by Reliable A, then Reliable B and after that the Unreliable informant showed the action. In this phase, behind each screen, a transparent bowl was placed on each board. The informant who gave the cue went out of the middle of the curtains, kneeled down between the two screens and lifted (Figure 8) one bowl above the screen to give the subject the cue where food could

be found. First, the informant presented the food in front of the body and during kneeling, both hands were moved back simultaneously behind her back and then forward to the bowls. Food was in the same hand of the cue side, while moving hands back. This was done so that the Unreliable informant could change the food from the cueing side to the other side, to give the misleading cue in 50 % of the trials. The number of trials stayed the same as in the crouching phase for all three informants.

After the second phase the second break could be claimed if the subject needed one. Then, before the third action started, Reliable A again gave a reminder of the lifting trials. During the third phase, a sound (Figure 9) cue was used, which was made with a box where dried food made a “shaking sound”. The informant who gave the cue, stepped out behind the curtain with the box in her hands, which did not make a sound during shaking, it was just used to present and make the movement of shaking. In order to connect the sound cue to one of the screens, the shaking sound was instead made by one of the other two informants behind the curtain on the corresponding side of the screen. Two sound cues of three times of shaking were used, during the first sound the informant was standing in the middle in front of the curtain. The second time the sound happened, when the informant was already behind the curtain. The informants walked backwards and disappeared in the middle of the two curtains in the lifting and sound actions. In contrast to the other phases, the order of the informants changed in the sound phase. First Reliable A conducted four trials, then it was already the turn of the Unreliable (4x), and after that Reliable B (4x) finished phase 3. This was done so that reliability is not completely confounded with the order of Reliable B and Unreliable informant showing up.

Throughout all three phases and all trials, as soon as the informant returned behind the curtain and the cueing was finished, the owner got a “beep” signal to release the dog. After that the dog could choose one of the two screens to find the food. If the dog made a wrong choice, the informants pulled the food back on the other side in full view of the subject. In this way, the animals had information in every trial whether food was on the side of cueing or on the other side. Even though the dogs could see the reward on the other side after a wrong choice, the two barriers between the two screens prevented them from walking to the other screen (Figure 6). During the three phases, two fixed breaks could be taken, one after phase 1 and the other one after phase 2 (Table 2).

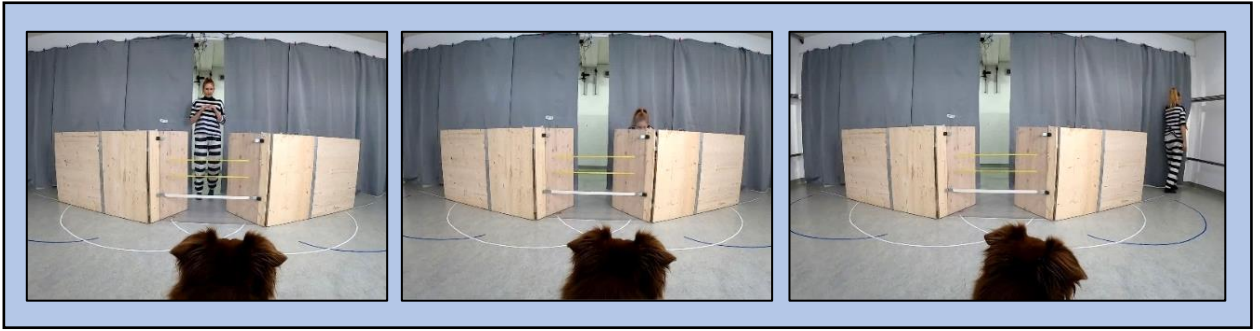


Figure 7: Action crouching

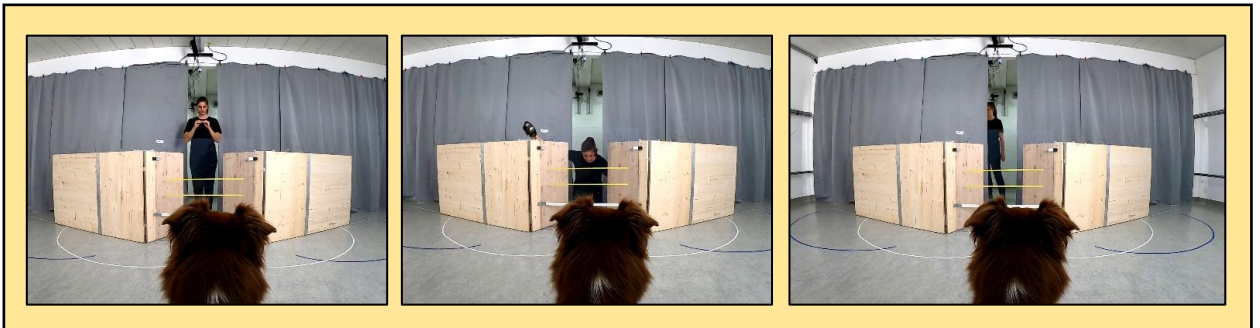


Figure 8: Action lifting

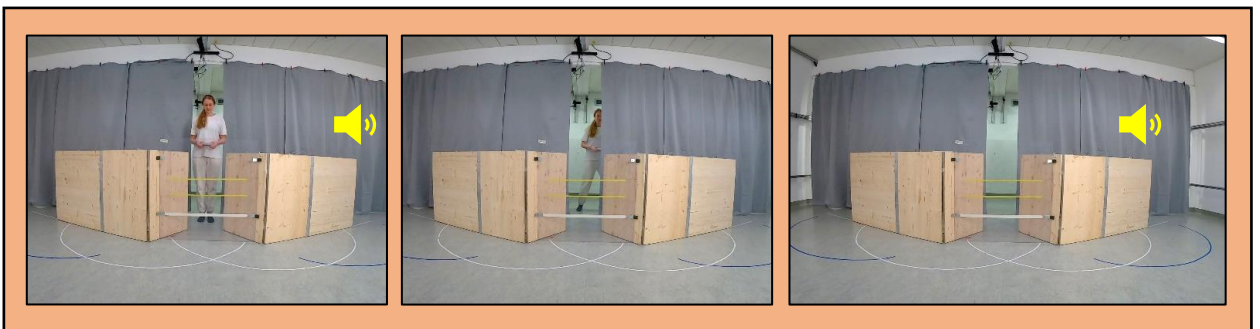


Figure 9: Action sound

Table 2: Schedule of phases with number of trials for each informant and sequence of trial.

Phase	Reliable A (not in analysis)	Reliable B	Unreliable (50 % misleading)	
<i>Warm-up</i>	4x crouch with owner help			
<i>Training</i>	10-20x crouch			
<i>Short break</i>				
<i>Familiarization</i>	4x crouch with owner help and 2x peeking introduction			
<i>Introduction of all informants</i>				
Demonstration phase 1	4x crouch			
		4x crouch		
			6x crouch	
		2x crouch		
<i>Short break</i>				
Demonstration phase 2	2x crouch reminder			
	2x lift			
		4x lift		
			6x lift	
		2x lift		
<i>Short break</i>				
Demonstration phase 3 (sequence change of informant)	2x lift reminder			
	2x sound			
			4x sound	
		4x sound		
Transfer tasks		4x screen choice		
	<i>Removal of the previous setup</i>			
	(places the 2 bowls)	2x pointing		
		2x begging		
	3x solvable	1x unsolvable		

2.4.5. Transfer tasks

After the demonstration phase, four transfer tasks were performed only by Reliable B and the Unreliable informant. We picked these four tasks in order to investigate whether the dogs formed a preference for Reliable B over Unreliable and whether this preference shows up in novel contexts as well that were increasingly different to the setup of the demonstration phase. In the first two transfer tests the informants went on providing the subjects with information where (i.e. in which of two food locations) food could be found. As Reliable B and Unreliable made simultaneous suggestions, the dogs needed to make a choice whom to follow. The last two tests were more different to the demonstration phase, as now the dogs had the chance to initiate communication with the experimenters. They could either beg for food or ask for their help to access food, and the question again was, if they showed a preference for Reliable B over the Unreliable informant.

2.4.5.1. Screen choice task:

The first transfer task was the screen choice task where the setup was the same as in the demonstration phase (Figure 5 and Figure 10). Both informants (Reliable B and Unreliable) appeared simultaneously on the opposite sides of the curtain and went to the screens to take a position in front of them while looking at the place behind the screens where food could be placed. Every dog got four screen choice trials in which the sides of the experimenters were counterbalanced. In half of the trials, both sides were baited whereas in the other half the screen indicated by Unreliable was left unbaited. The baiting event happened when all experimenters were still behind the curtain: two experimenters pushed the boards with or without reward secretly behind the screens. When both experimenters stood in position (in front of screen), Reliable A gave the releasing command for the dog from behind the curtain. As in the demonstration phase, the dog could only choose one side. After their choice, Reliable A gave the signal to the owner to call back their dog. After every trial, both experimenters disappeared behind the curtain through the middle gap and new rewards were placed before the next trials started. Before the next transfer task started, everything in front of the curtain was removed (screens, plastic panel and boards).

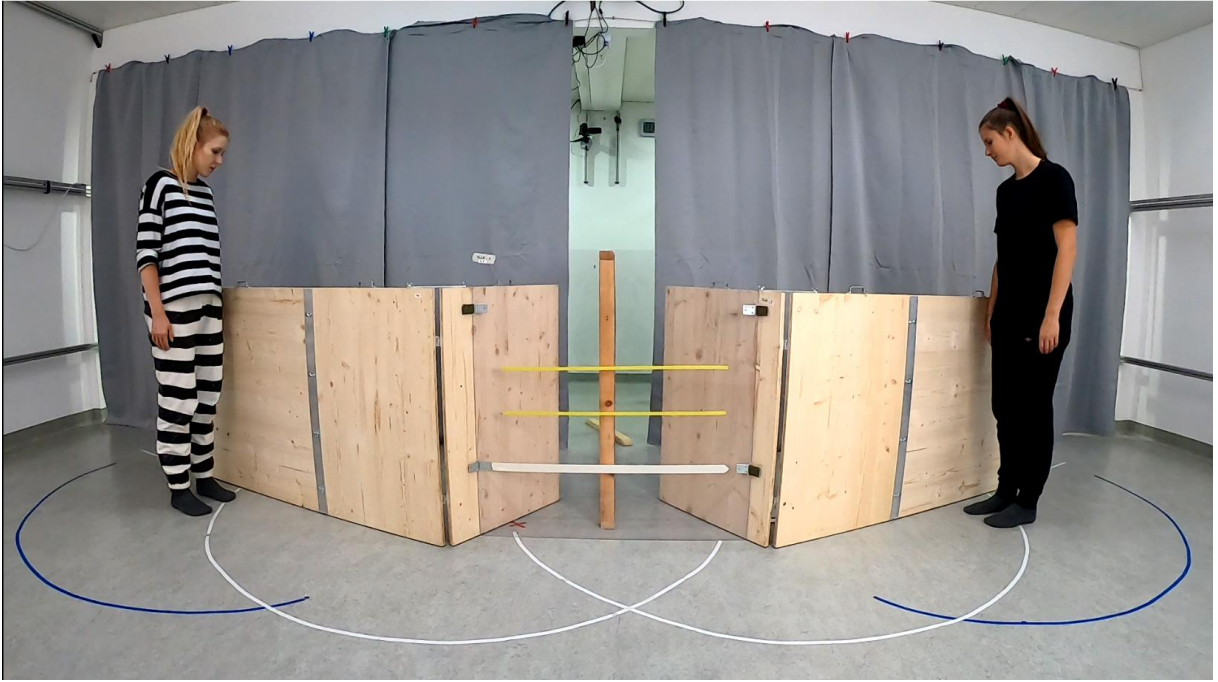


Figure 10: First transfer task, screen choice, both Informants (Reliable B and Unreliable) stand in front of the screens.

2.4.5.2. Pointing task:

The second transfer task was the pointing test (Figure 11 and Figure 12) that consisted of two trials, with starting side of the standing position (left or right) of Reliable B and Unreliable counterbalanced across dogs. First, Reliable A went out with two covered bowls, kneeled down in front of the subject and simultaneously pushed both bowls as far as possible apart. Then, without looking at the dog, she stood up and disappeared behind the curtain through the middle gap. The bowls had already been baited with food behind the curtain: in one trial both were baited whereas in the other trial only the one that was to be indicated by Reliable B. Reliable B and Unreliable appeared at the ends of the curtains, walked towards each other and took a standing position close together. Then they simultaneously started to point with the outer hand at the bowl of their side and looked at the bowls. The inner hand was on the back. The dog could choose one of the bowls, the other one was immediately removed by the informant who was pointing to it.



Figure 11: Second transfer task, pointing, both Informants (Reliable B and Unreliable) are pointing at a bowl.

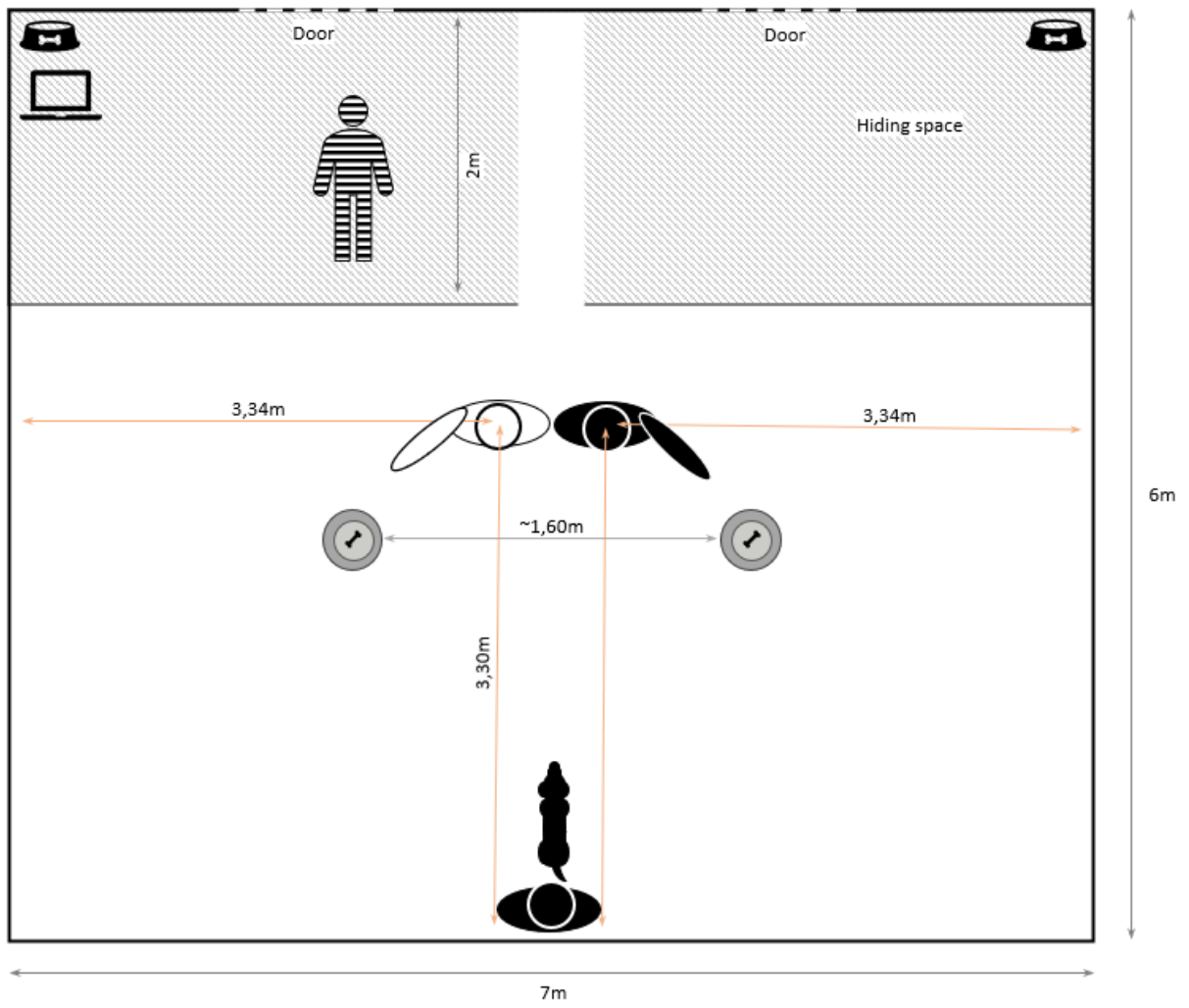


Figure 12: Setup of pointing task during the transfer phase with measurements.

2.4.5.3. *Begging task:*

In the third transfer task the dogs could beg from the informants (Figure 13 and Figure 14). Reliable B and the Unreliable informant came out again at the end of the curtains with a baited bowl (diameter 22cm) in their hands. They stood one meter apart and held bowls in front of their body without any further movements and eye contact to the dog. The subject was released for one minute, while, independent of its actions, both informants were watching the bowls in their hands in a neutral and unresponsive manner. After one minute, Reliable A signalled the owner to call the dog back and the informants disappeared behind the curtain from the sides. Then the trial was repeated after the informants changed sides behind the curtain. Starting side of the informants was counterbalanced through all dogs. The subject did not get any food in this task.

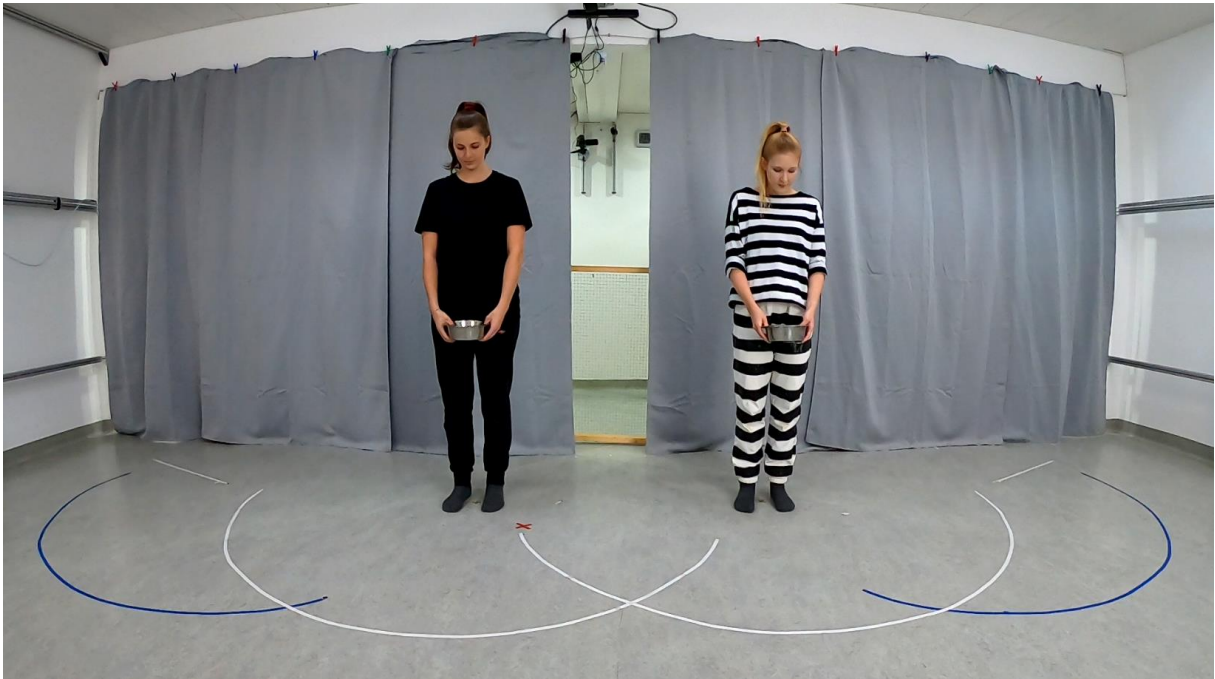


Figure 13: Third transfer task, begging, both Informants (Reliable B and Unreliable) hand a bowl with food in front of their body.

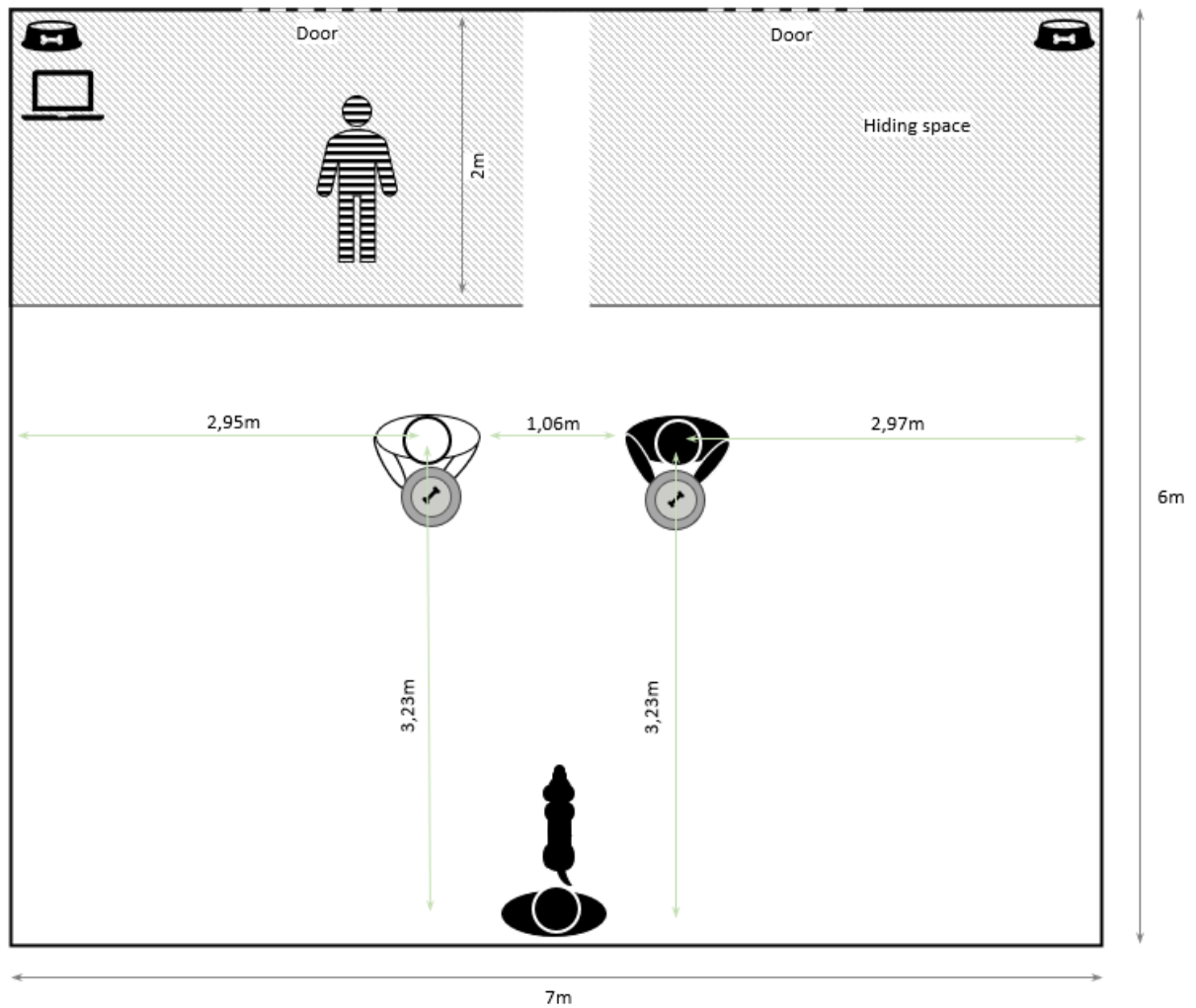


Figure 14: Setup of Begging task during the transfer phase with measurements.

2.4.5.4. Unsolvable task:

The last transfer task was a solvable/unsolvable task (Figure 15 and Figure 16). A lid (diameter 8.5 cm) of a jar (height 17.5cm) was screwed on a heavy board (70x120cm). The board was placed in the middle of the room by Reliable A (coming out from the middle). There were three solvable trials provided, where the jar stood unscrewed on the lid. Reliable A always did the baiting (kneeling down and placing food under the jar) and gave the owner the signal to release the dog when she was behind the curtain. The dog could now approach the board and make the jar fall in order to receive the reward under it. If dogs could not solve it, they got help from Reliable A and the trial was counted as “not solved” and the next trial started. In the unsolvable task, the Reliable A closed the jar with the lid on the board; the jar could no more be removed, which prevented the subject from solving this task on their own. When Reliable A disappeared

behind the curtain, Reliable B and Unreliable came appeared at the two ends of the curtains and walked near the board. During this unsolvable trial Reliable B and the Unreliable informant stood opposite to each other, on two sides of the board with the screwed jar (Figure 15). Both informants just stood there looking at the transparent jar and did not interact with the subject. The dog was released for two minutes and their behaviour was recorded until the whole test was completed. After the whole experiment was done, the dog was allowed to take the piece of food by the help of one experimenter.

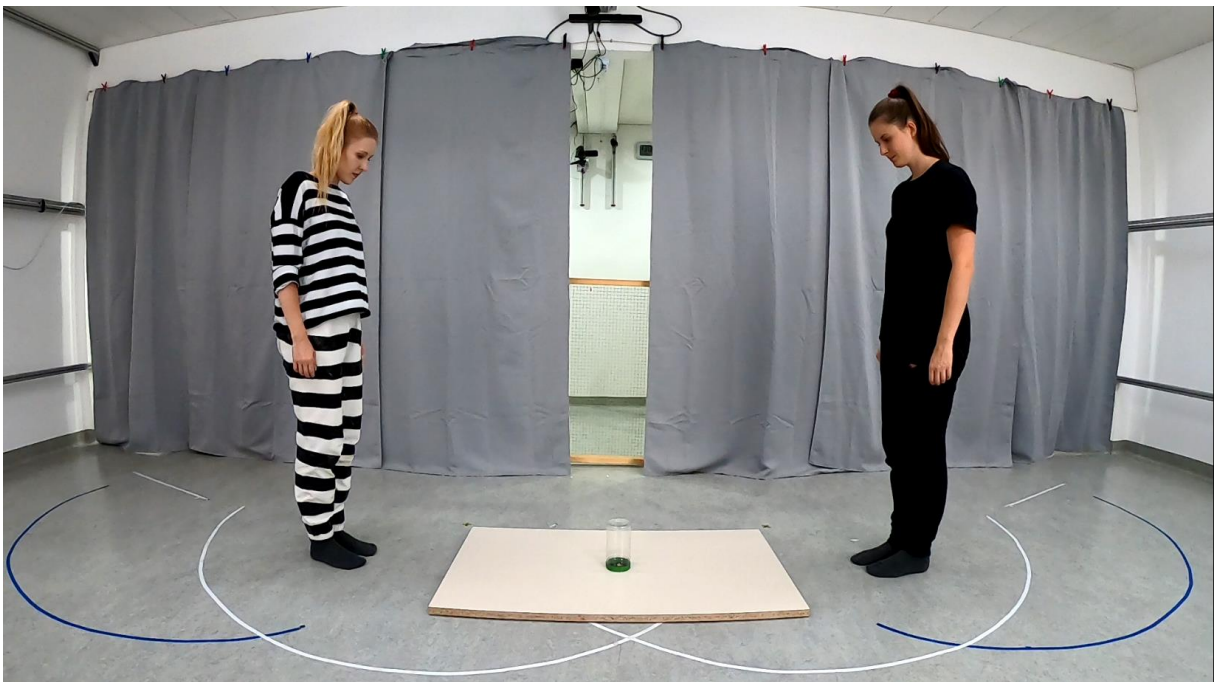


Figure 15: Fourth transfer task, unsolvable task, both Informants (Reliable B and Unreliable) stand near the unsolvable task.

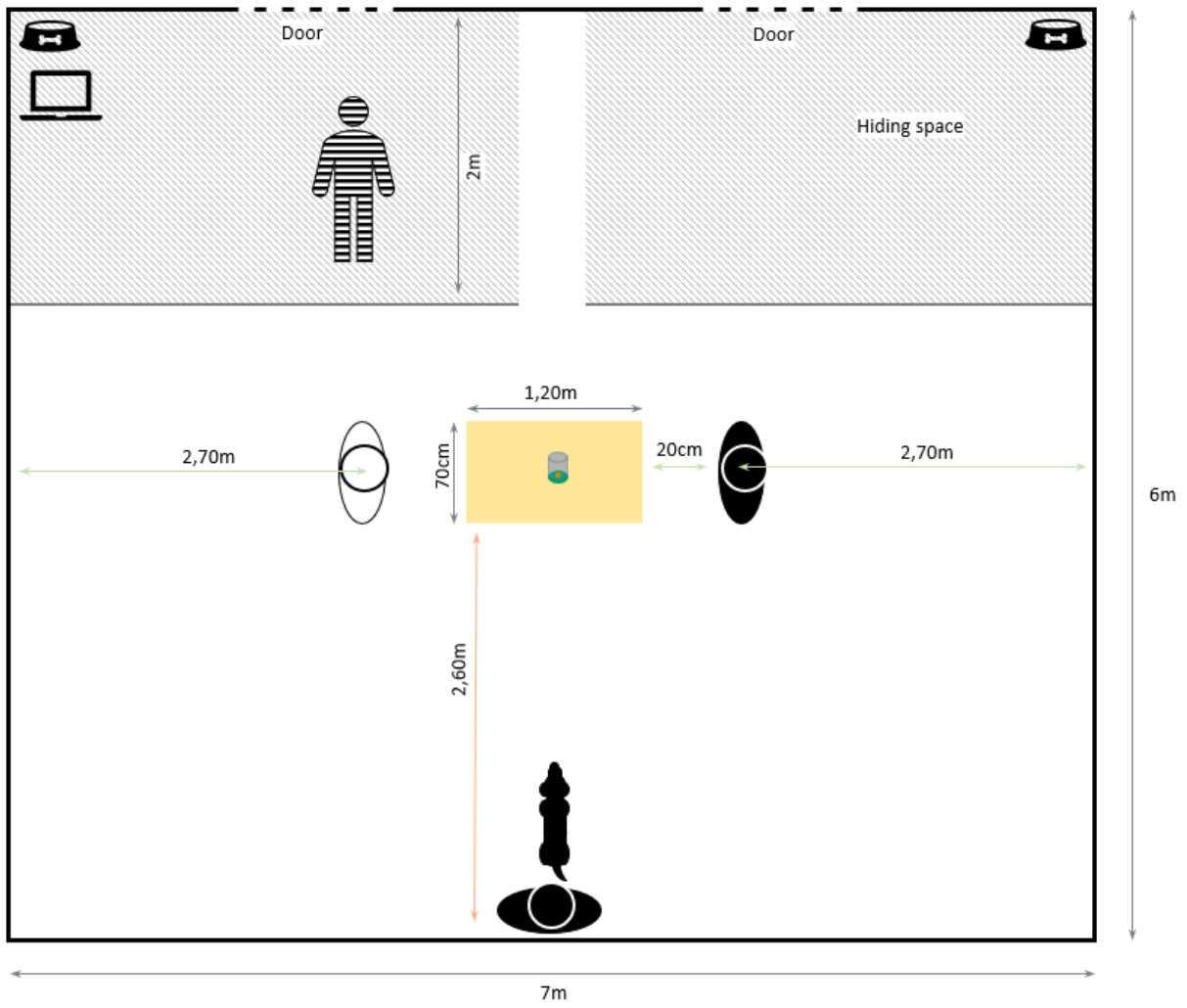


Figure 16: Setup of unsolvable task during the transfer phase with measurements.

2.5. Video scoring

All sessions were recorded with four Panasonic HD-Camcorder HC-V777 and one AXIS M3057-PLVE MkII. The videos were saved in Loopy (<http://loopbio.com>, Loopbio GmbH, Vienna, Austria), where the scoring was done as well. During the demonstration phase (Crouching, Lifting, Sound) the attentiveness of the subjects were recorded during the cue presentation (from the time the informant stepped in front of the curtain until the time she disappeared again). Attentiveness was coded when the snout of the dog was pointed to the curtain. This was coded to make sure the dogs saw the action to have an idea where the food was supposed to be. However, attentiveness was not analysed; I used it only to check if dogs were attentive during the cueing. Hesitation (latency to make a choice) was also coded in every trial, as time elapsed between the “beep” sign that signalled the owner to release the dog and the subject making a choice. The hesitation length was measured in order to check whether dogs had a longer latency to make a choice in the Unreliable than in the Reliable trials, presumably because in the Unreliable trials they were more unsure where to go. A choice was defined as the subject crossing the extended line of outer side of a screen with their nose. Before a choice was made, peeking could be coded as an event. Peeking was coded when dogs' snout was within 10cm to the gap of the screen or above the edge of the screen for at least one second.

For the four different transfer tasks, we have coded in each trial whether the dog firstly chose Reliable B or Unreliable. During the screen choice task, a choice was coded whenever the subject crossed the extended line of the screens. In the pointing trials, a choice was counted when a movement of a paper covering one of the bowls was visible. If a second “beep” was given and the dog still did not approach any of the bowls, “no choice” was coded. In regard to the last two transfers, begging and unsolvable test, we had a ground marking on the floor (see figures of transfer tasks). In the begging task, whenever the dog entered with their snout one of the two white circles marked on the ground around the informants, their choice was coded. In the unsolvable trial, a choice was coded when the dog was in one of the white circles and either sat, stood or lay there, and their nose was orientated to one experimenter for at least one second.

In addition to choice, proximity to the two informants was also coded in the screen choice, begging and unsolvable tasks. Proximity was coded when the dog's nose was within the marked circle on the ground to the respective informant and they either touched the

experimenter or vocalized in their direction or stood, sat, lay for at least one second with their nose orientated to the experimenter. The blue circles marked on the floor showed the distance in the screen choice task (Figure 15) while the white circles on the ground were used in the begging and unsolvable tasks.

2.5.1. Inter-Observer Reliability-Test

To test the consistency of our results among two internal video coders and an external video scoring person, an Inter-Observer Reliability-Test (IOR) was conducted using R (version 4.1.2, R Core Team, 2021). The *icc* function (Intraclass Correlation Coefficient) (Wolak et al., 2012) with tolerance of 1 was employed to measure hesitation length, resulting in 93,7% agreement. Transfer contact on the left side yielded an agreement outcome of 97,9%, while in the right side it was 97,6%. Peeking showed an agreement of 99,1%, and choice exhibited 98,9% agreement. Overall, the two internal video coders and one external video scorer achieved over 90% agreement in all variables.

2.6. Statistical analysis

All analyses were conducted in R (version 4.1.2, R Core Team, 2021), *p*-values smaller than 0.05 were used for significance testing.

2.6.1. Cue following model (Demonstration phase)

This model tested whether the reliability of the informant, the informants' actions and the dogs' breed group affected the dogs' probability to follow the informants' cues. The dependent variable was the dogs' choice (whether they followed informant's cue or not, binary variable) during the demonstration phase. Following the cue was scored as 1 and not following was scored as 0. To analyse this variable, we fitted a binomial Generalized Linear Mixed Effect Model (GLMM, Baayen, 2008) with logit link function, using the R function *glmer* of the package *lme4* (Bates et al., 2015).

As key test predictor we included the three-way interaction between informant role (Reliable B or Unreliable), action (crouch, lift and sound) and breed group (independent or cooperative). We included the interaction among these predictors, because our hypothesis was that the effect of dogs' breed, informant role, and action type was not the same across the levels of the

predictors (for example, that the difference in choice between breed groups would depend on the informant role and action shown). As a control predictor we included the side on which the action was performed (left or right). We included all theoretical identifiable random slopes (Barr et al., 2013; Schielzeth & Forstmeier, 2009), namely the random slope of the interaction between action and informant role within the random intercept of subject ID; the random slope of the interaction between informant role, action and breed group within experimental dyad (identity of the Reliable and Unreliable experimenters); finally, the random slope of the interaction between action, informant role and breed group within colour dyad (the colour of the clothes of the Reliable and Unreliable experimenters). We additionally included the random slope of the action side within all three random intercepts. We included all correlations between random intercepts and slopes (Matuschek et al., 2017). Before fitting the model, we dummy coded and centered the random effects to ensure that the results were not dependent on reference categories. We excluded the trials in the analysis in which the dogs did not make a choice. In total the model was based on 1728 observations from 54 different dogs, each completing 32 demonstration trials.

We tested the overall effect of the interaction by comparing the full model with a null model that lacked the three-way interaction, but retained the control fixed effect (action side) and the same random effect structure as the full model (Forstmeier & Schielzeth, 2011). The full-null model comparison was conducted with a likelihood ratio test (Dobson, 2002). The significance of the individual fixed effects was evaluated using likelihood ratio tests using the function `drop1` with argument `test` set to `Chisq`.

As for all the models described in this thesis, the stability was evaluated by comparing the estimates of models excluding each level of the random effects (Nieuwenhuis et al., 2012). See at the columns “min and max” in all model tables in the result section.

We checked the absence of collinearity by looking at variance inflation factors (VIF) (Field, 2005) using the results of a linear model fitted using the function `lm` of the package `car`. Collinearity was not an issue as all the VIFs were close to 1. We checked that the BLUPs (best linear unbiased predictors) were normally distributed. Confidence intervals were calculated by conducting a parametric bootstrap with 1000 repetitions and confidence level of 95 %.

2.6.2. Cue following model - first exposure (Demonstration phase)

As in the choice model above, the dependent variable was the dogs' choice, but only the trials of the first exposure to each experimenter and action were included in this analysis (that is, 1 crouching, 1 lifting and 1 sound trial for each of Reliable B and Unreliable). In total, the model was based on 324 observations (54 dogs, each completing six demonstration trials). We initially tried to fit to this reduced dataset the same binomial model described above for the choices during all demonstration trials. However, we incurred in complete separation probably due to the high rates in cue following and consequent low variation in the response. For this reason, we fitted a binomial model comprising as test predictors three 2-way interactions: between informant role and action, between action and breed group and between informant role and breed group. We controlled for the role of the side of the room on which the action was shown by adding this factor as a fixed effect. We included all theoretically identifiable random slopes (Barr et al., 2013; Schielzeth & Forstmeier, 2009): the random slope of informant role, action and action side within subject; the random slopes of all three interactions and of action side within experimenter dyad and colour dyad. All factors were centred prior to entering the random slope component of the model, to avoid the estimates being dependent on their reference categories. Correlation between random slopes and intercepts were not included due to convergence issues.

To avoid multiple testing, as an overall test of the test predictors, we compared the fit of the full model to that of a null model lacking the fixed effects of the test predictors and of all three 2-way interactions between them but being otherwise identical to the full model. The significance of the individual predictors was then assessed using likelihood ratio tests using the R function `drop1`.

Again collinearity was not an issue, as evidenced by the VIF that we obtained fitting a linear model, lacking the interactions, with the function `lm` of the package `car`. The BLUPs were normally distributed. The model stability and confidence intervals were calculated as described above.

Additionally, we explored whether overall the dogs' choices during the demonstration phase were significantly above chance level (0.5). We conducted two separate two-tailed, one-sample t-tests, one for the dogs' performance during first exposures and one for the overall performance during the demonstration, based on each subject's aggregated data, and adjusted the *p*-values using the Bonferroni correction.

2.6.3. Hesitation model (Demonstration phase)

The hesitation data was analysed to test if the latency to make a choice depends on the informant's role, the action type and the breed group. Our prediction was that the latency should have been longer for the Unreliable experimenter. The dependent variable was the hesitation length of the dog (measured in seconds from releasing the dog until the dog made a choice). To analyse this model we fitted a GLMM with gamma error distribution and log link function, using the R function `glmer` of the package `lme4`.

The key test predictor was the same as above (the 3-way interaction between informant role, action and breed group). As control predictors we included again the side of the action and additionally the peeking variable (peeking: yes=1, no=0). The peeking variable was added to control for the increase in latency potentially produced by the peeking behaviour. We included the random slopes of the interaction between action and informant role as well the random slope of action side within subject. Both within the random intercept of colour dyad and experimenter dyad, we included the random slope of the 3-way interaction between action, informant role and breed group as well as the random slope of action side. Finally, we included the random slope of peeking within all three random intercepts. Before entering the model, the random slope factors were manually dummy coded and centred. The correlations between random intercepts and slopes were removed due to convergence issues.

Trials with no choice were excluded from the latency analysis. Hence, the observations entering the model were 1715 (three dogs having 30 observations, seven dogs having 31 observations, the remaining 44 dogs having all 32 observations).

After fitting the model, we checked its assumptions by inspecting a qq-plot of the residuals and a plot of residuals against fitted values. These suggested that the residuals did not raise from a gamma distribution and were not homogeneous. However, because they showed less heteroscedasticity than those of a linear mixed model fitted to the same data, we continued the analysis on the gamma model. A test of overdispersion revealed that the model was underdispersed (dispersion parameter: 0.25), hence its estimates might be more conservative.

We assessed collinearity using the function `VIF` of the package `car` (version 3.1.1; Fox & Weisberg, 2011) for a model fitted without the interaction but retaining the fixed effects and the

same random effects. The maximum VIF was 1.00; hence, collinearity among predictors was not present.

The model stability was assessed removing one level of the grouping factors at a time and re-fitting the model. This revealed the model to be stable.

As an overall test of the test fixed effects and of their interactions, we fitted a null model lacking all test predictors (the main effects of informant role, breed group and action, their 3-way interaction and all 2-way interactions between them), but that was otherwise identical to the full model (Forstmeier & Schielzeth, 2011), using a likelihood ratio test (Dobson, 2002). As an overall test of the significance of the 3-way interaction (which reflected our initial hypothesis) we fitted a reduced model lacking only the 3-way interaction but not the main effects and all 2-way interactions, again using a likelihood ratio test. We used the function `drop1` to test for the main effect of the control predictors.

2.6.4. Choice of informant (Transfer phase)

The transfer task provides the possibility to choose between the two informants at the same time. Therefore, dogs' preference over one experimenter (Reliable B or Unreliable) was measured. The dependent variable was preferred informant role (Reliable B, coded as 1 or Unreliable, coded as 0)

To analyse the effect of breed group (cooperative or independent) on the dogs' choices during the transfer phases, we fitted a binomial model with logit link function. We included tasks (screen choice, pointing, begging or unsolvable), trial number within the task and side of the reliable informant as control predictors. We included the random slopes of trial number, task and reliable informant's side within subject ID, experimenter dyad and colour dyad. We additionally included the random slope of breed group within the experimenter dyad and colour dyad. The trial number was z transformed to a mean of 0 and a standard deviation of 1 and the random slope of task, breed group, reliable informant side were manually dummy coded and centred before entering the model. We checked the absence of collinearity among predictors by using the function `VIF` of the package `car` and then we assessed the range and distribution of the BLUPs, which were no issue. *P*-values for the fixed effects were obtained based on likelihood ratio test by using the function `drop1` with the argument `test set "chisq"`,

which compass the full with respective reduced models each lacking one of the predictors (Barr et al., 2013).

2.6.5. *Proximity to informant (Transfer phase)*

For each trial of the screen choice, begging and unsolvable tasks of the transfer phase, we calculated the proportion of time dogs spend close to the reliable experimenter over the total time they spent close to both experimenters. We removed from the analysis all trials (209 out of 378 trials) in which the dogs were not close to any of the experimenters. We squeezed the distribution of the response to exclude the extreme values of 0 and 1 (Smithson & Verkuilen, 2006). We fitted the proportion of time dogs spent close to the reliable experimenter a GLMM with beta error structure and logit link function. We included breed group as only test predictor and transfer tasks, trial number and side of the reliable experimenter as control predictors. The random effect structure was identical to the one described for the binomial above. We checked for the absence of overdispersion, by using the package DHARMA (Hartig, 2022), which revealed a dispersion parameter of 1.025 (p -value=0.656), which indicates the model was not overdispersed. P -values were again obtained with the function `drop1` as described above. Collinearity among predictors, evaluated as described above, was not an issue as all VIFs were close to 1. Confidence intervals (confidence level: 95 %) were calculated using the function “`confint`” of the R package `stats` (R Core Team, 2022).

3. Results

3.1. Demonstration phase

3.1.1. Follow cue all Demonstration trials

First, the full-null comparison did not show any significant effect ($\chi^2=5.945$, $df=11$, $p=0.877$) (Table 3). The three-way interaction of informant role, action and breed group did not show any significant effect on cue following during the demonstration phase of all trials ($\chi^2= 0.728$, $df=2$, $p=0.695$). However, we could find a significant effect of the side where Reliable B showed the action. If the experimenter (Reliable B) gave the action (cue) on the left side from the dog's perspective, dogs followed her cue more likely ($\chi^2=4.326$, $df=1$, $p=0.038$).

Overall, on average, the dogs followed the experimenters' cue on 80% of the trials (confidence interval: 76% - 84%), which is significantly above chance level (two-tailed, one-sample t-test: $t = 15.03$, $df = 53$, $p<0.001$). The cooperative breeds followed the Reliable B on average in 79% of all trials and the Unreliable 74% of all trials. For the independent breeds, the average of following the cue was 83% of all trials for the Reliable B and 84% of all trials for the Unreliable (Figure 17).

Table 3: Results of the full model - follow cue Demonstration phase (dependent variable: follow cue for all trials)

Fixed Factor	Term	Estimate	Std. Error	χ^2	Df	CI min	CI max	p-value	Min	Max
	intercept	2.623	0.606			1.749	3.923		2.085	3.231
Informant role	Informant role ¹	-0.287	0.681			-1.410	0.867		-0.747	0.364
Action	Action lift ²	0.068	0.718			-1.110	1.409		-0.538	0.580
	Action sound ²	-0.425	0.785			-1.672	0.788		-1.103	0.105
Breed group	Group ³	0.278	0.733			-0.824	1.721		-0.379	0.711
Action side	Action side ⁴	-0.909	0.485	4.33	1	-2.078	0.040	0.038	-1.185	0.645
Informant role*action*breed group	Informant role*action lift*group	0.063	0.928			-1.880	1.842		-0.966	0.864
	Informant role*action sound*group ⁵	1.021	1.225	0.728	2	-0.732	3.180	0.695	0.147	1.886

Note:

1 reference category Reliable B

2 reference category crouch

3 reference category cooperative

4 reference category left

5 the reported test refers to the overall effect of the interaction. The significance of the individual test predictors is not reported, because they were tested in an interaction.

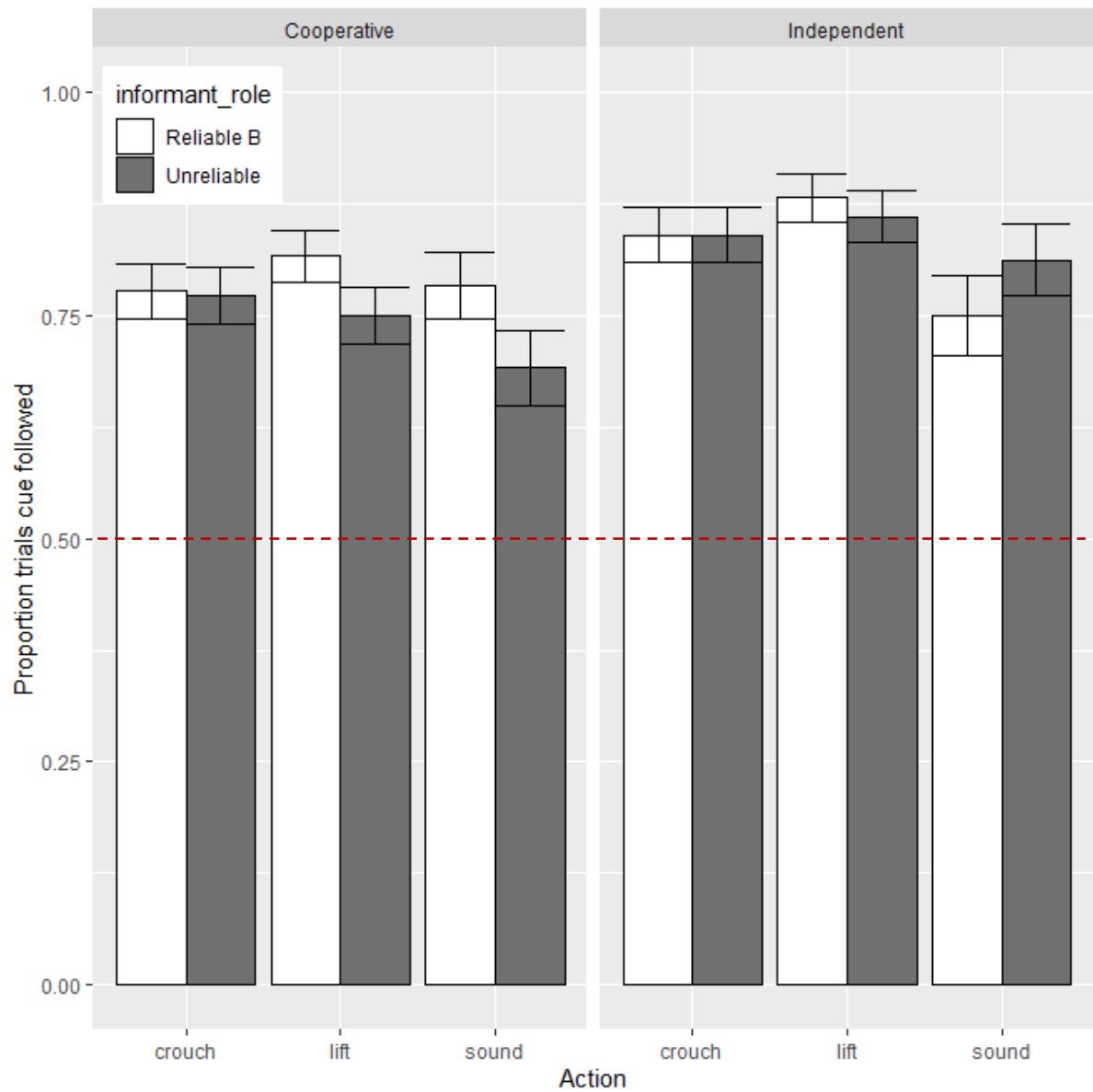


Figure 17: Cue following proportion of both breed groups (cooperative and independent) in all three actions (crouch, lift and sound) with both informants (Reliable and Unreliable) during all demonstration trials.

3.1.2. *Follow cue first exposure*

After investigating the cue following in all demonstration trials of Reliable B and Unreliable, we run an analysis for every first trial per action and informant. The full-null model comparison only approached significance ($\chi^2=16.92$, $df=9$, $p=0.050$), suggesting that the three two-way interactions (between informant role and action, between action and breed group and between informant role and breed group) did not affect significantly dogs' choices during the first trials (Figure 18). None of the three interactions taken individually was significant either (Table 4). The control predictor of action side did not contribute significantly to explaining the response either. Overall, the dogs followed the experimenters' cues in 78% of the trials (confidence interval: 73% - 84%), which is significantly above chance level (two-tailed, one sample t-test: $t=10.95$, $df=53$, $p<0.001$). On average, the cooperative breeds followed the reliable informant's cues in 76% of the first trials and the unreliable informant's cues in 68%. The independent breeds followed the reliable experimenter's cues on 86% of the first trials and the unreliable informant's on 89% of the trials.

Table 4: Results of the full model - follow cue Demonstration phase (dependent variable follow choice during first exposure).

Fixed factor	Term	Estimate	Std. Error	χ^2	Df	CI min	CI max	p-value	Min	Max
	intercept	1.378	0.599			0.482	24.663		0.996	14.116
Informant role	Informant role ¹	-0.194	0.598			-10.058	7.978		-0.911	22.467
Action	Action lift ²	1.257	0.807			-1.822	18.083		-3.77	18.293
	Action sound ²	-0.120	0.653			-11.694	8.686		-0.995	21.616
Breed group	Breed group ³	0.599	0.716			-6.085	6.585		-0.135	12.710
Action side	Action side ⁴	-0.529	0.428	1.586	1	-5.963	1.252	0.208	-15.895	-0.336
informant role*breed group ⁵	informant role*breed group	0.738	0.736	1.033	1	-4.174	15.486	0.309	-6.839	6.785
action*breed group	action lift*breed group ⁵	0.358	1.175	0.418	2	-7.879	15.703	0.811	-3.192	28.923
	action sound*breed group	0.557	0.889			-5.072	13.804		-17.264	6.599
Informant role*action	Informant role*action lift ⁵	-1.324	0.869	4.452	2	-23.51	3.335	0.108	-30.686	0.103
	Informant role*action sound	0.434	0.777			-8.5	15.232		-25.404	1.606

Note:

1 reference category Reliable B

2 reference category crouch

3 reference category cooperative

4 reference category left

5 the reported test refers to the overall effect of the interaction. The significance of the individual test predictors is not reported, because they were tested in an interaction.

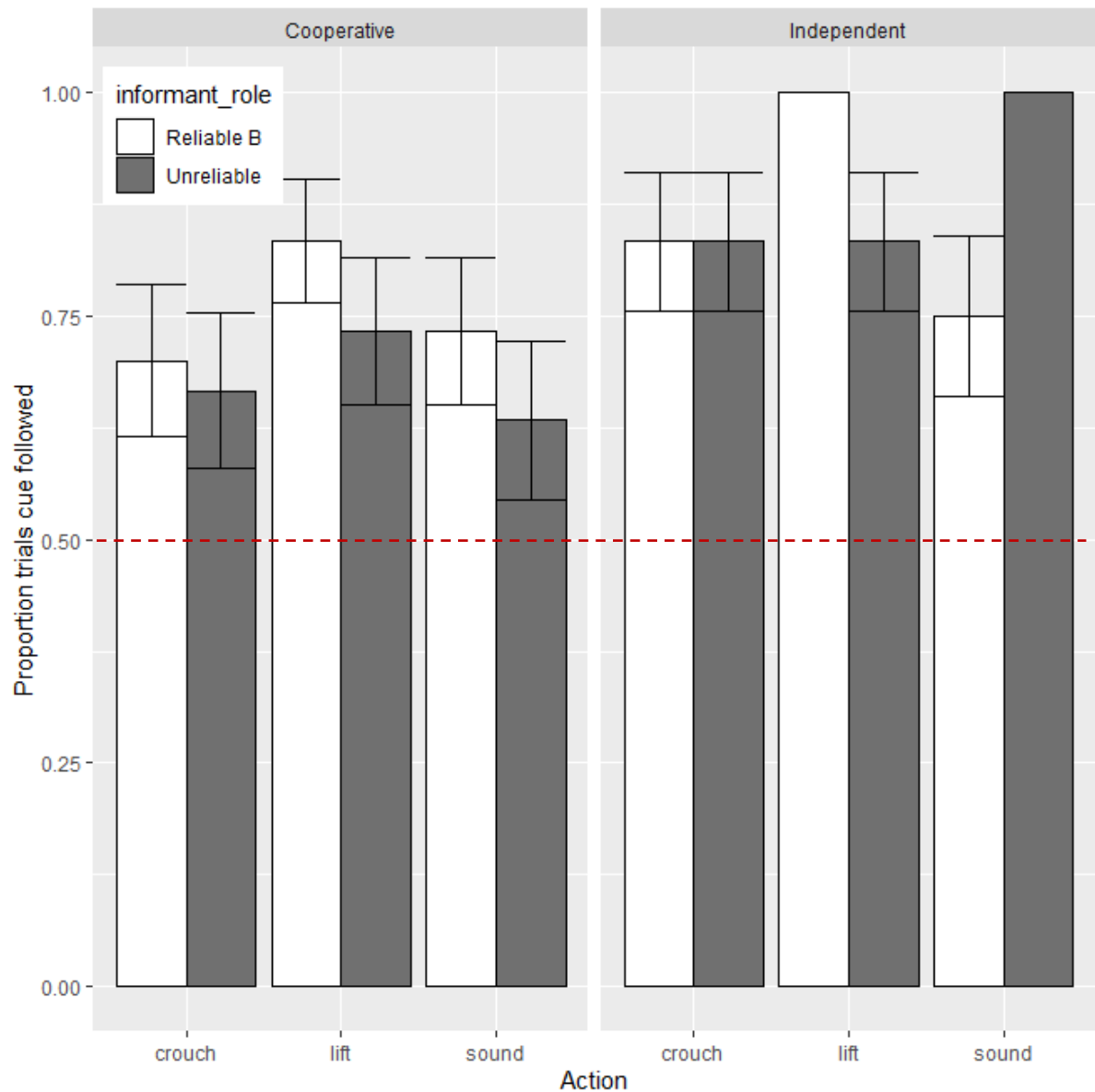


Figure 18: Cue following of both breed groups (cooperative and independent) in all three actions (crouch, lift and sound) with both informants (Reliable and Unreliable) during the demonstration phase in their first trial per action and informant.

3.1.3. Hesitation model

Same as for cue following, the full-null model comparison did not reach significance ($\chi^2= 6.81$, $df=11$, $p=0.814$), indicating that none of the test predictors (neither the main effects, nor the three way interaction nor the two-way interactions) significantly predicted the dogs' latency to make a choice.

There was no significant effect of the three-way interaction of informant role, action and breed group on the hesitation length of the dogs (full-reduced model comparison: $\chi^2=0$, $df=2$, $p=1$). The variable action side had a significant effect in dogs' hesitation length (Figure 19), with dogs taking more time to make a choice when the action was performed on the right than on the left side from the dogs perspective ($\chi^2=4.085$, $df=1$, $p=0.043$). The variable peeked did not affect the dogs' hesitation significantly ($\chi^2=2.997$, $df=1$, $p=0.083$) (Table 5).

Table 5: Results of the full model - hesitation during Demonstration phase (dependent variable: hesitation length to make a choice).

Fixed factor	Term	Estimate	Std. Error	χ^2	Df	CI min	CI max	p-value	Min	Max
	intercept	1.255	0			0.863	1.639		1.193	1.319
Informant role	Informant role ¹	-0.072	0			-0.448	0.337		-0.176	0.002
Action	Action lift ²	-0.267	0.001			-0.698	0.116		-0.334	-0.243
	Action sound ²	-0.025	0			-0.62	0.419		-0.156	-0.016
Breed group	Group ³	0.392	0			-0.148	0.973		0.280	0.616
Action side	Action side ⁴	0.015	0	4.085	1	-0.161	0.194	0.043	-0.003	0.068
Peeked	Peeked ⁵	0.736	0	2.997	1	0.187	1.281	0.083	0.613	0.837
Informant role*action*breed group ⁶	Informant role*action lift*group	0.037	0			-1.024	0.976		-0.029	0.361
	Informant role*action sound*group	-0.013	0	0	2	-1.105	1.107	1	-0.196	0.303

Note:

1 reference category Reliable B

2 reference category crouch

3 reference category cooperative

4 reference category left

5 reference category peeked not

6 the test reported refers to the comparison between full and reduced model lacking the three-way interaction but retaining all main effects and two-way interactions

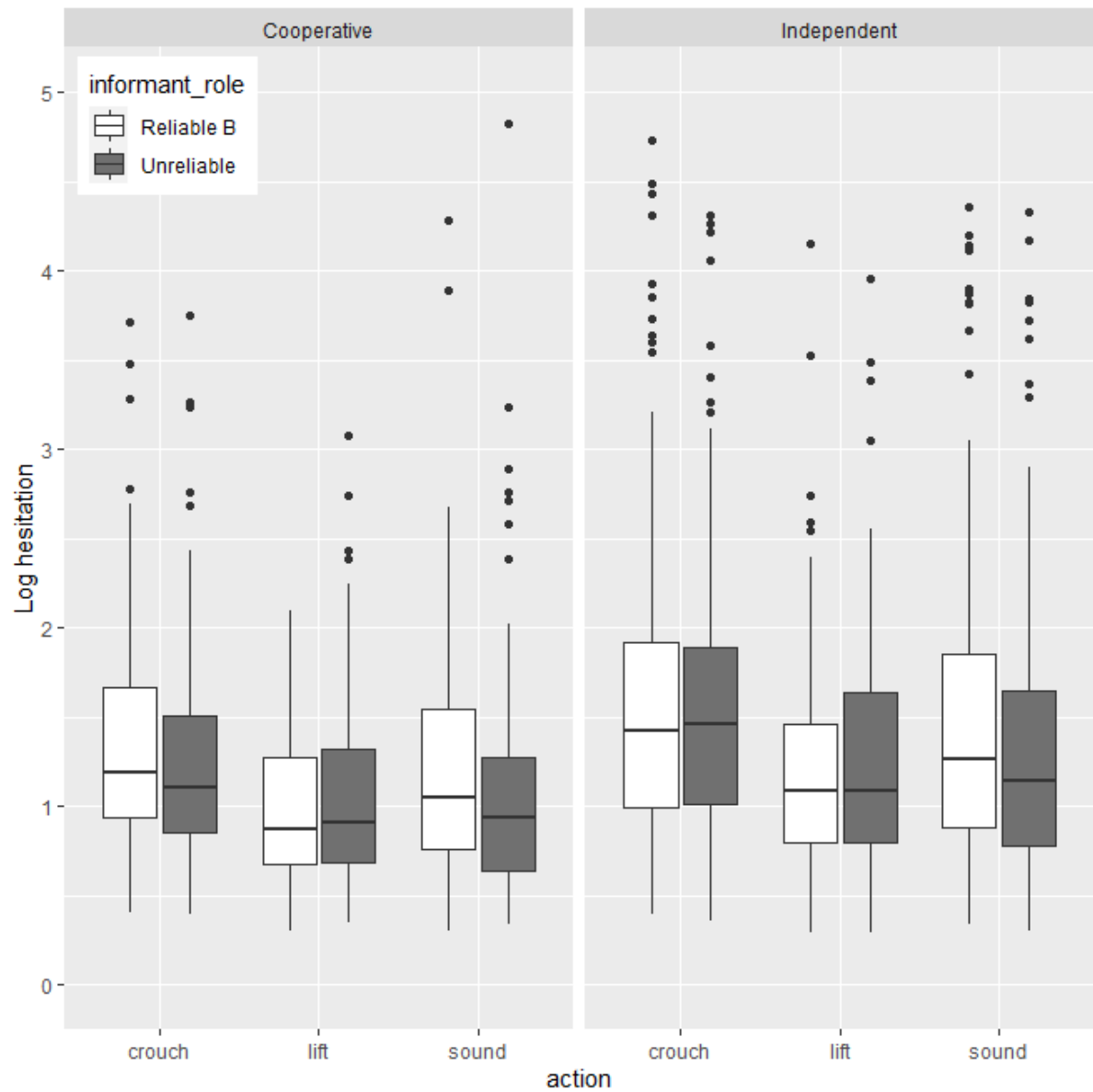


Figure 19: Hesitation of both breed groups (cooperative and independent) for all three actions (crouch, lift and sound) with both informants (Reliable and Unreliable) during the demonstration phase.

3.2. Transfer Tasks

3.2.1. Transfer-Choice

The dogs' breed group, the task and trial number did not have an influence on the dogs' choices for the transfer phase (Figure 20). At the group level dogs chose equally often the Reliable and Unreliable experimenter (the side on which the Reliable experimenter stood was chosen in 50% of the trials, 227 trials and the side of the Unreliable experimenter on 231 trials). The dogs were significantly more likely to choose the Reliable experimenter when this was on the right compared to the left side of the room ($\chi^2= 4.588$, $df=1$, $p=0.032$) (Table 6).

Table 6: Results of the full model - choice during Transfer Tasks (dependent variable: preferred informant role).

Fixed factor	Term	Estimate	Std. Error	χ^2	Df	CI min	CI max	p-value	Min	Max
	intercept	-0.217	0.28			-0.795	0.352		-0.531	0.057
Breed group¹	Breed group independent	-0.247	0.202	1.504	1	-0.704	0.165	0.220	-0.384	-0.099
Transfer Task²	Transfer pointing	0.069	0.387			-0.675	0.935		-0.216	0.547
	Transfer Screen choice	0.039	0.31			-0.573	0.663		-0.277	0.253
	Transfer unsolvable⁴	0.021	0.466	0.036	3	-0.980	1.098	0.998	-0.444	0.397
	Trial number	-0.003	0.121	0.001	1	-0.254	0.244	0.981	-0.065	0.075
	Reliable side³	0.622	0.256	4.588	1	0.114	1.207	0.032	0.451	0.750

Note:

1 reference category cooperative

2 reference category begging

3 reference category left

4 the reported test refers to the overall effect of the interaction. The significance of the individual test predictors is not reported, because they were tested in an interaction.

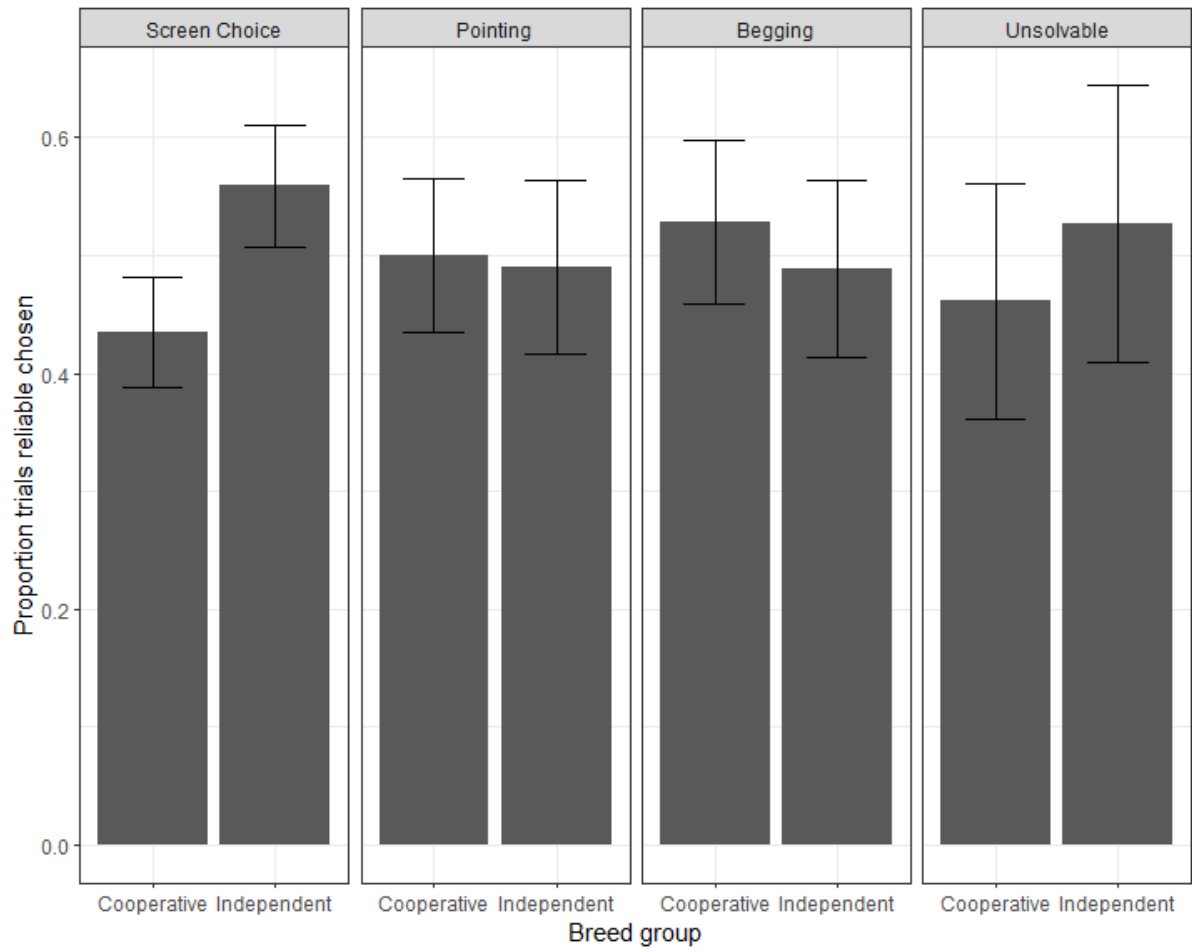


Figure 20: Proportion to choose Reliable over Unreliable in all four transfer tasks (Screen choice, Pointing, Begging and Unsolvable) for both breed groups (cooperative and independent).

3.2.2. Transfer-Proximity

On average the dogs spent equal proportions of time in proximity to the two experimenters (mean proportion of time spent close to Reliable B=0.5; Figure 21). Dogs' proportion of time spent close to Reliable B did not differ significantly across the three transfer tasks (screen choice, begging and unsolvable; $\chi^2= 0.608$, $df=2$, $p=0.738$; Table 7).

Table 7: Results of the full model - proximity during Transfer Tasks (dependent variable: proportion of time spent close to Reliable B experimenter).

Fixed factor	Term	Estimate	Std. Error	χ^2	Df	CI min	CI max	p-value	Min	Max
	intercept	0.266	0.208			-0.142	0.674		0.147	0.449
Breed group¹	Breed group independent	0.03	0.211	0.020	1	-0.385	0.445	0.888	-0.139	0.153
Transfer Task²	Transfer Screen choice	0.114	0.406			-0.681	0.910		-0.180	0.351
	Transfer unsolvable	-0.186	0.271	0.608	2	-0.717	0.345	0.738	-0.338	-0.052
	Trial number	-0.073	0.126	0.351	1	-0.321	0.175	0.554	-0.208	0.015
	Reliable side³	-0.425	0.229	2.697	1	-0.874	0.024	0.101	-0.634	-0.171

Note:

1 reference category cooperative

2 reference category begging. The reported test refers to the overall effect of the predictor.

3 reference category left

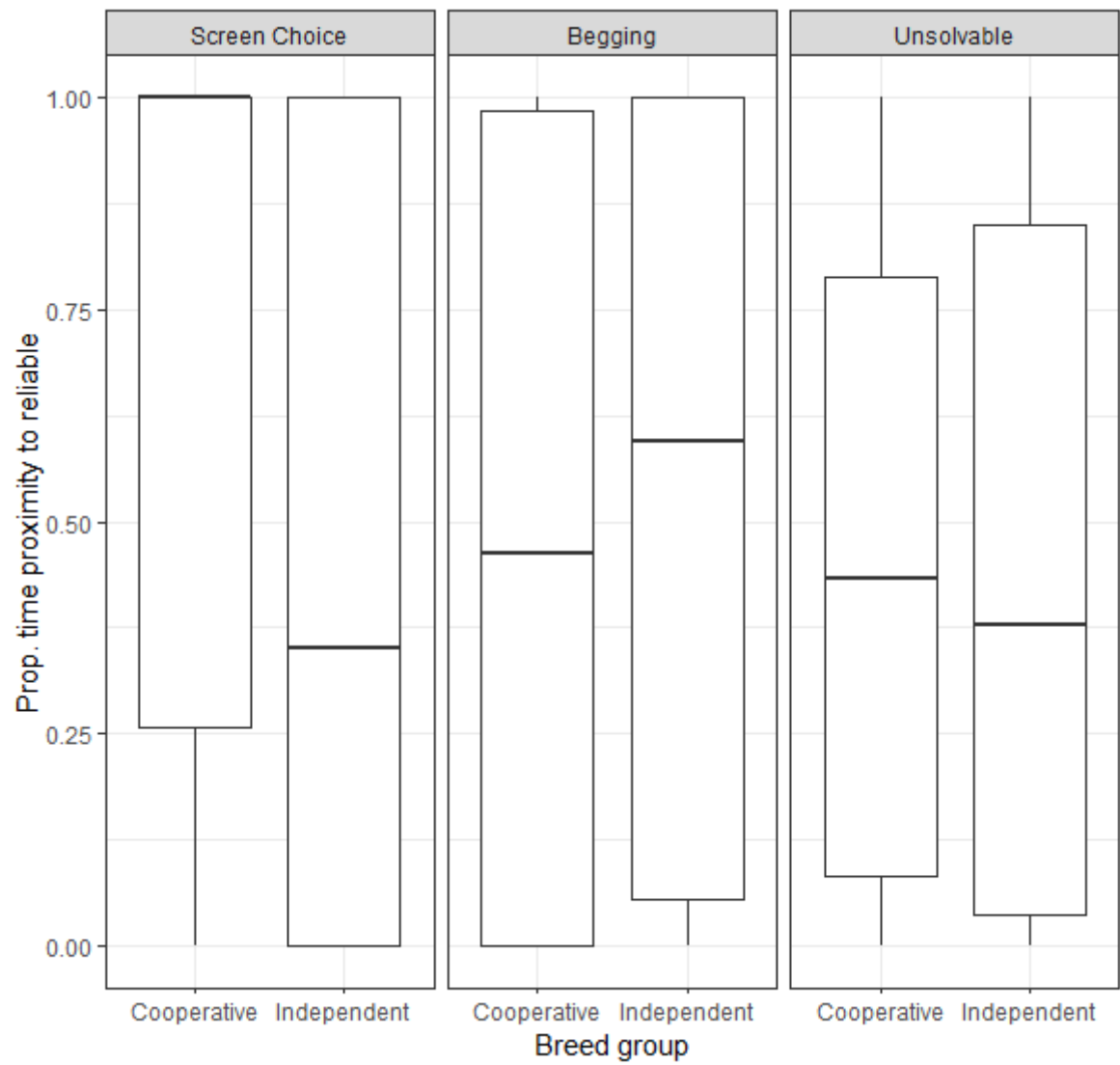


Figure 21: Proportion of time proximity to Reliable B over the total time they spent time near both experimenters.

4. Discussion

The purpose of the study was to find out whether dogs can assess the reliability of human informants with novel cues in different contexts and whether dogs belonging to cooperative vs. independent working breeds differ in this regard.

4.1. Reliability assessment

Despite the fact that dogs could follow novel communicative cues from human informants, they did not react to the Unreliable person. In addition, an exploratory analysis revealed significant results that the dogs followed the informants' cue above chance during the first exposure (78%) and in all trials (80%). However, our predictions during the demonstration phase were not confirmed. There was no significant evidence that dogs followed the Unreliable informant less than the Reliable informant, as we originally predicted. Furthermore, there was no longer hesitation to make a choice with the Unreliable informant than with the Reliable informant. In general, peeking could only be observed very rarely. We had predicted that it should have occurred more often when dogs were previously misled by a person, as the dogs might have tried to look for themselves where the food was hidden, when confronted with the Unreliable experimenter. Even during the transfer test, dogs did not show a preference for the Reliable informant over the Unreliable, their choices and proximity proportion was equal for both informants. Thus, dogs did not behave differently in our experiment if a person was unreliable to them in previous trials. There was only a tendency for the full-null model comparison for the first exposure, however there was no significant effect in the three two-way-interactions (between informant role and action, between action and breed group and between informant role and breed group), which could explain this effect. The only significant results we got do not answer our main questions, they were rather included in the analyses as control predictors. During the demonstration phase (all trials), dogs chose the left side more often, but we had some left biased dogs which may have had an influence on these results. Another significant side preference was shown during the transfer phase; in this case the dogs chose more often the right side when Reliable B was standing there. One simple explanation could be that dogs preferred this side more, as the exit door was on this side of the room (behind the curtain). This however does not explain why they preferentially chose the other side during the demonstration phase. Probably this preference explains why the dogs hesitated less when the action was performed on the left, their preferred, side of the room.

Our negative finding seems to be in contrast with the Pelgrim et al. (2021) and Takaoka et al. (2015) studies where dogs responded differently to misleading suggestions from the experimenters and avoided their pointing more often than those of a reliable pointer. However, our task was more complicated than former studies, as we asked the dogs to generalize across actions and learn about the reliability of the informants themselves instead of just following reliable cues and ignoring unreliable ones. Therefore, our findings may indicate limited capacities of dogs to assess the reliability of human informants by generalizing across their diverse cues. Furthermore, our study was more demanding than former ones by not providing direct information about the location of food prior to the informant's cue. In both former dog studies (Pelgrim et al., 2021; Takaoka et al., 2015) the dogs could first observe the baiting process and then the informant made a misleading suggestion for the unbaited spot. This direct information about the actual location of food might be important for dogs to recognize when a cue is misleading. In contrast, in our study, the dogs received information about the food location only after having watched the cue and having made a choice. To gain this information, when they did not find the food, they had to pay attention to the other side and notice that food was being removed from there. This means that they could learn about the misleadingness of the cue based on a posteriori information, and they had only a few (6 or 4) trials with the Unreliable informant and only in half of those they had the chance to recognize this contingency. Therefore, our negative findings may reflect limited inferential capacities of dogs, at least under these hard conditions.

Alternatively, it is possible that dogs are capable of making such inferences but this capacity does not manifest in contexts where the dogs need to respond to human-given cues. Former studies have suggested that during domestication dogs were selected to follow human communicative cues, and for this reason they have difficulties to avoid or ignore such cues, even if they are misleading (Petter et al., 2009; Szetei et al., 2003). Dwyer & Cole (2018), for instance, made it very obvious for the dogs where the food was hidden by using transparent containers. Still, dogs kept following the human cues that were directed at the clearly empty container. Whether this readiness to listen to people even though they have made wrong suggestions in previous attempts is due to innate predispositions (Barnard et al., 2019) or individual learning is not clear (Petter et al., 2009). It may well be that the life-long experiences of dogs contribute to this "positivity bias". All dogs engage in similar interactions on a daily basis at least with their owner, or possibly with multiple people: our behaviour often indicates where food or other reward can be found. Especially our companion dogs usually do not have bad experiences with people, because most of the time dogs are surrounded by people who

do not want to harm them. Thus, these interactions are mostly positive, which may make it more difficult for dogs to recognize an unreliable person because, like children, dogs do not have the life experience we adults have (Mills, 2013). Additionally, the general context of the experiment and the Clever Dog Lab itself may have made the dogs identify this experiment as a hiding-search game of food. If so, this may have made it even more difficult to expect behaviours from a person that would be unrelated to the food location.

This tendency of dogs to follow human cues even if they are misleading can explain why dogs in our experiment did not respond differently to the cues of the unreliable informant during the demonstration phase, but cannot explain why dogs did not show a preference for the Reliable informant during the transfer tasks. In this phase the task was either to select of identical cues of two informants who were present simultaneously or there were no cues at all but rather the subjects could initiate an interaction. The finding that the dogs failed to differentiate between the Reliable and Unreliable informants also in these contexts seems to suggest that the dogs could not infer the reliability of human informants in different contexts. It is likely that dogs experienced every new setting with humans just as a new task with new chances to get food from the informants (novel task – novel informants' role). A recent study by Völter et al. (2023) also found no evidence that dogs transfer their previous experiences in a new context in an unwilling/unable paradigm. Although dogs reacted differently to the experimenters' intentions during the demonstration phase, they did not generalize their intentions to a new context where dogs could show their preference for one person in a pointing task (Völter et al., 2023). Similarly, another study (Carballo et al., 2020) investigated if dogs can differentiate between attitudes (generous and selfish) of people. Only dogs who chose first the generous person were tested in an unsolvable task as well. Even dogs who were selected to show a capability to differentiate between people's attitudes, did not show a preference (first gaze) for the generous person in an unsolvable task. They only found longer looking behaviours towards the generous person (Carballo et al., 2020). The study shows us once again that it is difficult for dogs to apply their experiences to new contexts. This could be the reason why we could not detect a preference for Reliable B in our transfer tasks, as we did not even exclude dogs who did follow both informants equally.

Before making this conclusion, we must consider whether there are simpler, rather procedural reasons why dogs failed to differentiate between the two informants in this task. An experimental explanation for this could be that the Unreliable person gave the dogs enough rewards even if this was only the case in 50% of her trials. During training or even in everyday

situations, dogs are used to not constantly receiving rewards for every single correct behaviour (Petter et al., 2009). Therefore, they must not be surprised or even forgive if they were not rewarded. Adding to this, for companion dogs it might not be that costly to lose some rewards, therefore it does not matter for them to get less food, as our dogs normally receive enough food and the ability to differentiate between reliable and unreliable humans might not be advantageous considered pet dogs' social ecology.

Making learning even more difficult, the total number of misleading trials was actually only around 13% (maximum of all 86 trials with ten misleading trials during Demonstration phase and Screen Choice Transfer). Trials of the Reliable informants can be counted to a maximum of 76 trials and ten reliable trials with the Unreliable (as she had only 50% misleading trials). It could be worthwhile to minimize the number of rewarded trials for the Unreliable person, so that dogs can recognize the misleading person easier. Other studies have in total 50% misleading trials (of 200 trials in 5 sessions) (e.g. Petter et al., 2009). For instance, only in 25% the Unreliable experimenter gives the correct cue on the rewarded side, this might be easier for dogs to distinguish the Unreliable person to the Reliable one. Or one could at least minimize the number of trials with the Reliable informants.

Another reason could be that dogs did not distinguish between the different informants. If dogs did not pay attention to the identity of the person, they cannot react differently in next trials with an Unreliable person. In our study, the three informants alternated quite frequently, which could also have led to confusion in dogs, similar to that seen in Petter et al. (2009) study in which they found no significant effect of dogs following the cooperators significantly more than the deceptor when they alternated more frequently. Anyway, there was also an event during the demonstration phase which might disturb dogs' memory after every trial, barriers between the two screens were pushed forward, which made a sound and movement. This event could be distracting for the dogs.

Additionally, the total length of the experiment was too long; it lasted between one and two hours, depending on the dog's individual speed. This means that it is unlikely that they could concentrate during the entire experiment on the different roles of the informants, even though we provided three breaks. Maybe the approximately five to ten minutes breaks were too short for the dogs or they could not relax properly as they were in an unfamiliar setting. Another overwhelming fact for dogs could be that there were too many different tasks and this distracted them from the different roles of the informants. As we had three different cues during our demonstration phase and four different transfer tasks, it might happen that dogs concentrated more on the tasks and not on the informants' roles.

4.2. Comparison of breed groups

In addition, we did not establish direct communication with the dogs, but the dogs could only observe us giving them new signals without any common component. Further interest of this study was the comparison of two working breed groups, where three hypotheses were proposed at the beginning. One was suggesting that if cooperative breeds are better in detecting deception, they would differentiate between the two informants sooner and/or stronger than dogs of independent breeds. The other hypothesis was that if independent breeds are better in ignoring human behaviour, they react sooner and/or stronger with less following of the Unreliable person. Third, we suggested that if both of these differences between breed groups are at place, they would both follow the Reliable informant more than the Unreliable one. In contrast with our predictions, none of the two breed groups followed the Reliable informant more than the Unreliable one, either in the demonstration phase or in the transfer tasks. As this study could not find any significant effect that dogs can detect the reliability of a human informant and react to it, it is not particularly surprising that no breed difference was found either.

As our findings seem to be in contrast with Lonardo et al. (2021), where their findings support breed differences in a True-False-Belief study and Heberlein et al. (2017), revealed in a breed difference in the understanding of humans' perception. I would first like to discuss the differences between the studies as this might explain the divergent outcomes. In comparison to the True-False-Belief study, we had many more trials: in our study dogs could have a maximum of 86 trials per dog. Already the training trials could reach a maximum of 20 trials (on average, dogs needed around 13 trials). Whereas the True-False-Belief study included only a maximum of twelve trials (one final test trial) in total. This big difference could already show the possibility that dogs got too many training trials in our study and after that had more difficulties to ignore the misleading suggestions. Therefore, it is quite possible that dogs simply followed the rule of going where a human made the action. Another big difference can be mentioned in the communicative intention. In Lonardo's and colleagues (2021) study, the experimenter addressed the dogs before cueing: they had eye contact to the dog and gaze movements between dog and bucket while they also talked to the dog. Also, in Heberlein et al. (2017) study, the owners who had previously forbidden the dog to take the food, always had direct eye contact with the dogs. In various contexts it has been shown that dogs respond differently to the same information if it was provided in a communicative manner instead of only observing the same demonstration (Byosiere et al., 2022; Kaminski et al., 2012; Pongrácz

et al., 2004; Topál et al., 2010). In our study the dogs could only observe the experimenters' suggestions during the demonstration but were not directly addressed; thus, it may be that the dogs had difficulties in understanding the misleading intention. We never communicated directly to the dog that there must be food behind the screen, we only raised a small piece of food in front of our body, with neither eye-contact nor verbal interaction with the dogs. Therefore, dogs might not react with less following to the Unreliable as they did not perceive it as an intention to be deceptive. This could be especially crucial for cooperative breeds that the Unreliable person gave ostensive cues that there is food hidden for the dogs to be misled. For a cooperative breed it would make sense to be more sensitive if humans communicate ostensively towards them, as they were bred to work with the human and follow commands (Gácsi, McGreevy, et al., 2009). Thus, even if cooperative breeds recognized the Unreliable informant, they might have difficulties to ignore human-given cues and still follow her misleading trials as previous trials were reliable. Perhaps, cooperative breeds accepted the unreliable behaviour as an accidental mistake. For the cooperative breeds, following a human could be a sufficient reward, or they interpret the situation as a game. Finally, as already mentioned above, they simply tend to cooperate with humans. For the dogs, this could also mean that we have not actually communicated directly that there is something to be found here, just because we have done something behind one screen, does not mean that we claim that food is there.

Another big difference between our study and Lonardo et al. (2021) and Heberlein et al. (2017) was the possibility to see the baiting process. The study from Scheider et al. (2013) has shown that it is important for dogs to know about the food location to follow their suggestions or not. In comparison if dogs could watch the baiting or not, dogs more likely followed the pointing if they could not watch the actual baiting. The study shows that dogs who know where food is hidden ignore the human gesture more likely and do not see it as a strong command. In Lonardo et al. (2021) study, the dogs had the possibility to observe beforehand where food was placed, which, as discussed earlier, may have made the interpretation of the humans' behaviour easier. And in Heberlein et al. (2017) study, the dogs also had the possibility to rely on their own knowledge. Dogs understood that they were not allowed to take food and they showed that they understood the perception of their caregiver: at least family and independent breeds reacted differently in taking the food from the hidden place. As argued above, the lack of attention calling may explain why we found no evidence for cooperative breeds detecting sooner the Unreliable experimenter. We find no explanation however why we found no evidence of terriers being better at ignoring human cues. Therefore,

it may be interesting to consider another hypothesis that Lonardo's et al. (2021) have raised. They suggested that terriers may have followed the human in the true belief trial more than Border collies because terriers were more prepared to perceive her suggestion as a new goal. A similar argument may explain in our study as well why terriers did not stop following the unreliable informant. Assuming that they realized her unreliability, they might have continued to follow her if they had assumed that the informant's intention was to show something new. It should also be mentioned here that we had different actions, this could encourage the belief that the informant is doing something new.

Although only one breed group was labelled as cooperative according to Gácsi, McGreevy, et al., 2009, it could still be that both were selected for cooperation with humans, so it could be that one group just doesn't need as much eye contact, which we never had in our study. As already discussed above, dogs in general may feel obliged to follow human cues independent of their working style.

My third hypothesis was that both breed groups have gone through relevant changes (i.e. the cooperative breeds became more prepared to detect deception and the independent breeds to ignoring misleading cues), which predicted that both breed groups should be similarly skilled in assessing humans' reliability. Given our negative results, we obviously found not support for this hypothesis.

4.3. Novel cue following without ostensive communication

Our study did not demonstrate that dogs can assess the reliability of human informants, or rather, they did not respond differently to informants based on their reliability. However, our study shows that dogs can follow different humans and diverse novel cues of different humans. During the demonstration phase, dogs followed on average in around 80% of the trials the human given cues (crouch, lift and sound actions) to find food in a two choice object task. All these cues were novel for the dogs as they did not contain any pointing, leaning or gazing to one side. Here it should be highlighted, that the informants made sure they never made eye contact with the dogs or used any other direct communication (e.g. calling dogs name, tongue snap) to the dogs. Our study was an observational study with no ostensive cues, as we did not address dogs' attention. However, studies (e.g. Kaminski et al., 2012; Tausin et al., 2015) have shown that dogs follow humans better when they have addressed the dogs, e.g. through eye contact, naming or eye movements. This kind of communication is called ostensive communication, which stands in contrast with our study, as we did not give ostensive cues to

the dogs. In our study there were only observational cues through all actions and phases. Interestingly, many researchers found out that dogs will follow better with ostensive cues than with non-ostensive cues (e.g. Duranton et al., 2017). However, as we had more training trials, this could have made a big difference in the ability to follow non-ostensive cues, as they had many opportunities to learn from previous trials where food can be found.

Conclusion

The aim of this work was to further explore whether dogs can differentiate an Unreliable person from Reliable ones and therefore respond differently to her in new contexts. Another big interest was, if dogs' working style and therefore breed has an influence on their differentiation skill of detecting Reliable and Unreliable informants. As three hypotheses were suggested at the beginning: First, if cooperative breeds are better in detecting deception, they should react to an Unreliable informant than independent breeds. Second, if independent breeds are better at ignoring misleading suggestions, they should ignore more often the Unreliable person than cooperative breeds. Finally, if both breed groups were selected as suggested above, both breed groups should react similar to the Unreliable informant. However, the predictions of this study were not confirmed, both breed groups followed the Reliable and Unreliable to a similar extent. Further, they did not hesitate more to make a choice with the Unreliable as predicted at the beginning. Even, during the transfer tasks, no preference for Reliable B over Unreliable could be detected. Therefore, both breed groups did not react differently to the Unreliable informant over the Reliable at least in our complex designed study where several methodological points may have influenced the negative outcome. Dogs were asked to detect an Unreliable person by three novel cues with no common component and they should generalize their prior experiences over different contexts. Also, the informants' communication that the cue will now be misleading was not given in our experiment because the dogs could not observe the baiting process. Further, for dogs it was only an observational context as we did not address dogs in advance. However, it might be difficult for dogs to avoid human made cues, also if the cues were misleading. Possibly the extended training phase at the beginning made dogs even more inclined to follow in upcoming trials. In addition, the misleading trials did not happen often which made it even more difficult for dogs to react differently next time. Furthermore, dogs did not even show a preference for the Reliable person, when they could choose between the two informants, which is explained by the challenge to generalize to new contexts. Due to methodological differences to former studies, it could not be determined that dogs can differentiate between Reliable and Unreliable informants and further no breed differences could be found as in previous studies. Although dogs did not differentiate between the two informants, they were really good in following three different novel non-ostensive cues, irrespective of the informants' role or breed group category. To gain better insight into whether dogs are capable of assessing human reliability, more research is needed to develop new methods for suitable reliability assessments.

Acknowledgment

In the last year, I get to know many different people and words will never express how thankful I am to have you all on my side. You gave me a valuable insight into research and working in the team has been a truly pleasant experience.

First, I would like to express my deepest appreciation to my supervisors, Zsófia Virányi and Lucrezia Lonardo. Your expertise, patience, and continuous guidance have been instrumental in shaping this research and pushing me to achieve my best. I am very grateful for your insightful feedback, valuable suggestions, and endless support.

I would like to express my gratitude to the head of the department, Ludwig Huber, for providing me with the necessary resources and opportunities to conduct this research.

I'm very glad to have such a big help for the statistics, without your assistance, such a complex and accurate statistical analysis would not have been possible. Thank you very much Lucrezia Lonardo, Remco Folkertsma and Kirsten Blakey for your invaluable help.

And the rest of the ARED team (Kirsten Blakey, Giacomo Melis, Ariane Veit and Eva Rafetseder), who were always ready to answer questions and provide help whenever needed. I was really happy to get the opportunity to be part of this project and I am happy to see you all again!

Last but not least, I want to thank these two women, Kea Amelung and Kinga Kovács, even in stressful situations we managed to work as a great supportive team and I can say from the bottom of my heart, testing with you guys has been the most fun of this long journey!

For organizational help, I want to thank Karin Bayer (lab manager) for her endless supportive advices and for technical help I want to thank Peter Füreder and Wolfgang Aust-Berger.

Most of all, I want to thank all owners who voluntarily participated with their dogs, which is not self-evident, but of great importance for this research, without you, this study would not exist! Thank you for your essential contribution and time.

Finally, I want to thank my family, friends and especially my boyfriend Paul Kelterer who gave me support in many different ways. Thank you for always being there for me!

References

- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511801686>
- Barnard, S., Passalacqua, C., Pelosi, A., Valsecchi, P., & Prato-Previde, E. (2019). Effects of breed group and development on dogs' willingness to follow a human misleading advice. *Animal Cognition*, *22*(5), 757–768. <https://doi.org/10.1007/s10071-019-01272-3>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). <https://doi.org/10.18637/jss.v067.i01>
- Byosiere, S., Mundry, R., Range, F., & Virányi, Z. (2022). Selective responding to human ostensive communication is an early developing capacity of domestic dogs. *Developmental Science*. <https://doi.org/10.1111/desc.13361>
- Carballo, F., Cavalli, C., Martínez, M., Dzik, V., & Bentosela, M. (2020). Asking for help: Do dogs take into account prior experiences with people? *Learning & Behavior*, *48*(4), 411–419. <https://doi.org/10.3758/s13420-020-00425-6>
- Chow, V., Poulin-Dubois, D., & Lewis, J. (2008). To see or not to see: Infants prefer to follow the gaze of a reliable looker. *Developmental Science*, *11*(5), 761–770. <https://doi.org/10.1111/j.1467-7687.2008.00726.x>
- Dobson, A. J. (2002). *An introduction to generalized linear models* (2nd ed). Chapman & Hall/CRC.

- Duranton, C., Range, F., & Virányi, Z. (2017). Do pet dogs (*Canis familiaris*) follow ostensive and non-ostensive human gaze to distant space and to objects? *Royal Society Open Science*, 4(7), 170349. <https://doi.org/10.1098/rsos.170349>
- Dutrow, E. V., Serpell, J. A., & Ostrander, E. A. (2022). Domestic dog lineages reveal genetic drivers of behavioral diversification. *Cell*, 185(25), 4737-4755.e18. <https://doi.org/10.1016/j.cell.2022.11.003>
- Dwyer, C., & Cole, M. R. (2018). Domesticated dogs (*Canis familiaris*) tend to follow repeated deceptive human cues even when food is visible. *Learning & Behavior*, 46(4), 442–448. <https://doi.org/10.3758/s13420-018-0356-8>
- Field, A. (2005). *Discovering statistics using SPSS*. (1. ed., reprinted). SAGE Publ.
- Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, 65, 47–55.
- Fox, J., & Weisberg, S. (2011). *An R companion to applied regression*. Sage publications.
- Frantz, L. A. F., Bradley, D. G., Larson, G., & Orlando, L. (2020). Animal domestication in the era of ancient genomics. *Nature Reviews Genetics*, 21(8), 449–460. <https://doi.org/10.1038/s41576-020-0225-0>
- Gácsi, M., Gyoöri, B., Virányi, Z., Kubinyi, E., Range, F., Belényi, B., & Miklósi, Á. (2009). Explaining Dog Wolf Differences in Utilizing Human Pointing Gestures: Selection for Synergistic Shifts in the Development of Some Social Skills. *PLoS ONE*, 4(8), e6584. <https://doi.org/10.1371/journal.pone.0006584>
- Gácsi, M., McGreevy, P., Kara, E., & Miklósi, Á. (2009). Effects of selection for cooperation and attention in dogs. *Behavioral and Brain Functions*, 5(1), 31. <https://doi.org/10.1186/1744-9081-5-31>

- Hare, B., & Tomasello, M. (2005). Human-like social skills in dogs? *Trends in Cognitive Sciences*, 9(9), 439–444. <https://doi.org/10.1016/j.tics.2005.07.003>
- Harris, P. L. (2007). Trust. *Developmental Science*, 10(1), 135–138. <https://doi.org/10.1111/j.1467-7687.2007.00575.x>
- Harris, P. L., & Koenig, M. A. (2006). Trust in Testimony: How Children Learn About Science and Religion. *Child Development*, 77(3), 505–524. <https://doi.org/10.1111/j.1467-8624.2006.00886.x>
- Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive Foundations of Learning from Testimony. *Annual Review of Psychology*, 69(1), 251–273. <https://doi.org/10.1146/annurev-psych-122216-011710>
- Hartig, F. (2022). *DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed)*. <https://CRAN.R-project.org/package=DHARMA>
- Heberlein, M. T. E., Turner, D. C., & Manser, M. B. (2017). Dogs' (*Canis familiaris*) attention to human perception: Influence of breed groups and life experiences. *Journal of Comparative Psychology*, 131(1), 19–29. <https://doi.org/10.1037/com0000050>
- Junttila, S., Valros, A., Mäki, K., Väättäjä, H., Reunanen, E., & Tiira, K. (2022). Breed differences in social cognition, inhibitory control, and spatial problem-solving ability in the domestic dog (*Canis familiaris*). *Scientific Reports*, 12(1), 22529. <https://doi.org/10.1038/s41598-022-26991-5>
- Kaminski, J., Schulz, L., & Tomasello, M. (2012). How dogs know when communication is intended for them. *Developmental Science*, 15(2), 222–232. <https://doi.org/10.1111/j.1467-7687.2011.01120.x>
- Lenkei, R., Carreiro, C., Gácsi, M., & Pongrácz, P. (2021). The relationship between functional breed selection and attachment pattern in family dogs (*canis familiaris*). *Applied Animal Behaviour Science*, 235, 105231. <https://doi.org/10.1016/j.applanim.2021.105231>

- Lonardo, L., Völter, C. J., Lamm, C., & Huber, L. (2021). Dogs follow human misleading suggestions more often when the informant has a false belief. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1955), 20210906. <https://doi.org/10.1098/rspb.2021.0906>
- Luu, B., Rosnay, M. de, & Harris, P. L. (2013). Five-year-olds are willing, but 4-year-olds refuse, to trust informants who offer new and unfamiliar labels for parts of the body. *Journal of Experimental Child Psychology*, *116*(2), 234–246. <https://doi.org/10.1016/j.jecp.2013.06.003>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315.
- McGetrick, J., Brucks, D., Marshall-Pescini, S., & Range, F. (2020). No evidence for a relationship between breed cooperativeness and inequity aversion in dogs. *PLOS ONE*, *15*(6), e0233067. <https://doi.org/10.1371/journal.pone.0233067>
- Miklósi, Á., Faragó, T., Fugazza, C., & Gácsi, M. (Eds.). (2018). *Der Hund: Geschichte, Biologie, Rassen* (J. Wissmann, C. Wink, & M. Niehaus, Trans.). Haupt Verlag.
- Miklósi, Á., Kubinyi, E., Topál, J., Gácsi, M., Virányi, Z., & Csányi, V. (2003). A Simple Reason for a Big Difference. *Current Biology*, *13*(9), 763–766. [https://doi.org/10.1016/S0960-9822\(03\)00263-X](https://doi.org/10.1016/S0960-9822(03)00263-X)
- Miklósi, Á., & Topál, J. (2013). What does it take to become ‘best friends’? Evolutionary changes in canine social competence. *Trends in Cognitive Sciences*, *17*(6), 287–294. <https://doi.org/10.1016/j.tics.2013.04.005>
- Mills, C. M. (2013). Knowing when to doubt: Developing a critical stance when learning from others. *Developmental Psychology*, *49*(3), 404–418. <https://doi.org/10.1037/a0029500>
- Nieuwenhuis, R., Te Grotenhuis, H. F., & Pelzer, B. J. (2012). *Influence. ME: tools for detecting influential data in mixed effects models*.

- Parker, H. G., Dreger, D. L., Rimbault, M., Davis, B. W., Mullen, A. B., Carpintero-Ramirez, G., & Ostrander, E. A. (2017). Genomic Analyses Reveal the Influence of Geographic Origin, Migration, and Hybridization on Modern Dog Breed Development. *Cell Reports*, *19*(4), 697–708. <https://doi.org/10.1016/j.celrep.2017.03.079>
- Parker, H. G., Shearin, A. L., & Ostrander, E. A. (2010). Man's Best Friend Becomes Biology's Best in Show: Genome Analyses in the Domestic Dog. *Annual Review of Genetics*, *44*(1), 309–336. <https://doi.org/10.1146/annurev-genet-102808-115200>
- Pelgrim, M. H., Espinosa, J., Tecwyn, E. C., Marton, S. M., Johnston, A., & Buchsbaum, D. (2021). What's the point? Domestic dogs' sensitivity to the accuracy of human informants. *Animal Cognition*, *24*(2), 281–297. <https://doi.org/10.1007/s10071-021-01493-5>
- Petter, M., Musolino, E., Roberts, W. A., & Cole, M. (2009). Can dogs (*Canis familiaris*) detect human deception? *Behavioural Processes*, *82*(2), 109–118. <https://doi.org/10.1016/j.beproc.2009.07.002>
- Pongrácz, P., Gómez, S. A., & Lenkei, R. (2020). Separation-related behaviour indicates the effect of functional breed selection in dogs (*Canis familiaris*). *Applied Animal Behaviour Science*, *222*, 104884. <https://doi.org/10.1016/j.applanim.2019.104884>
- Pongrácz, P., Miklósi, Á., Timár-Geng, K., & Csányi, V. (2004). Verbal Attention Getting as a Key Factor in Social Learning Between Dog (*Canis familiaris*) and Human. *Journal of Comparative Psychology*, *118*(4), 375–383. <https://doi.org/10.1037/0735-7036.118.4.375>
- Pongrácz, P., Miklósi, Á., Vida, V., & Csányi, V. (2005). The pet dogs ability for learning from a human demonstrator in a detour task is independent from the breed and age. *Applied Animal Behaviour Science*, *90*(3–4), 309–323. <https://doi.org/10.1016/j.applanim.2004.08.004>

- Poulin-Dubois, D., & Brosseau-Liard, P. (2016). The Developmental Origins of Selective Social Learning. *Current Directions in Psychological Science*, 25(1), 60–64. <https://doi.org/10.1177/0963721415613962>
- Prato-Previde, E., Marshall-Pescini, S., & Valsecchi, P. (2007). Is your choice my choice? The owners' effect on pet dogs' (*Canis lupus familiaris*) performance in a food choice task. *Animal Cognition*, 11(1), 167–174. <https://doi.org/10.1007/s10071-007-0102-7>
- R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/>
- Scheider, L., Kaminski, J., Call, J., & Tomasello, M. (2013). Do domestic dogs interpret pointing as a command? *Animal Cognition*, 16(3), 361–372. <https://doi.org/10.1007/s10071-012-0577-8>
- Schielezeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, 20(2), 416–420.
- Serpell, J. (Ed.). (2017). *The domestic dog: Its evolution, behavior and interactions with people* (Second edition). Cambridge University Press.
- Shafto, P., Eaves, B., Navarro, D. J., & Perfors, A. (2012). Epistemic trust: Modeling children's reasoning about others' knowledge and intent. *Developmental Science*, 15(3), 436–447. <https://doi.org/10.1111/j.1467-7687.2012.01135.x>
- Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11(1), 54.
- Szetei, V., Miklósi, Á., Topál, J., & Csányi, V. (2003). When dogs seem to lose their nose: An investigation on the use of visual and olfactory cues in communicative context between dog and owner. *Applied Animal Behaviour Science*, 83(2), 141–152. [https://doi.org/10.1016/S0168-1591\(03\)00114-X](https://doi.org/10.1016/S0168-1591(03)00114-X)

- Takaoka, A., Maeda, T., Hori, Y., & Fujita, K. (2015). Do dogs follow behavioral cues from an unreliable human? *Animal Cognition*, *18*(2), 475–483. <https://doi.org/10.1007/s10071-014-0816-2>
- Tauzin, T., Csík, A., Kis, A., Kovács, K., & Topál, J. (2015). The order of ostensive and referential signals affects dogs' responsiveness when interacting with a human. *Animal Cognition*, *18*(4), 975–979. <https://doi.org/10.1007/s10071-015-0857-1>
- Topál, J., Miklósi, Á., Sümegei, Z., & Kis, A. (2010). Response to Comments on “Differential Sensitivity to Human Communication in Dogs, Wolves, and Human Infants”. *Science*, *329*(5988), 142–142. <https://doi.org/10.1126/science.1184152>
- Tummeltshammer, K. S., Wu, R., Sobel, D. M., & Kirkham, N. Z. (2014). Infants Track the Reliability of Potential Informants. *Psychological Science*, *25*(9), 1730–1738. <https://doi.org/10.1177/0956797614540178>
- Völter, C. J., Lonardo, L., Steinmann, M. G. G. M., Ramos, C. F., Gerwisch, K., Schranz, M.-T., Dobernig, I., & Huber, L. (2023). Unwilling or unable? Using three-dimensional tracking to evaluate dogs' reactions to differing human intentions. *Proceedings of the Royal Society B: Biological Sciences*, *290*(1991), 20221621. <https://doi.org/10.1098/rspb.2022.1621>
- Wobber, V., Hare, B., Koler-Matznick, J., Wrangham, R., & Tomasello, M. (2009). Breed differences in domestic dogs' (*Canis familiaris*) comprehension of human communicative signals. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *10*(2), 206–224. <https://doi.org/10.1075/is.10.2.06wob>
- Wolak, M. E., Fairbairn, D. J., & Paulsen, Y. R. (2012). *Guidelines for Estimating Repeatability*. (3(1); *Methods in Ecology and Evolution*, pp. 129–137).

Appendix

List of figures

Figure 1: “Fig. 5 Procedures for the phases of Experiment 3. In Visit 1 dogs completed, in order, a warm-up, b object permanence, c occlusion visible baiting, and d occlusion-hidden baiting. In visit 2, dogs in order, a warm-up, d occlusion-hidden baiting, e informant history, and f conflicting informants’ test trials” (Pelgrim et al., 2021, p. 291).	17
Figure 2: "Fig. 1 Examples of familiarization trial blocks containing four reliable face cues (left) and four unreliable face cues (right) from Experiment 1. At the start of each trial, one of two female faces appeared and looked toward one of the boxes. An animal then appeared in one of the boxes. The reliable face always looked to the correct box (i.e., where the animal would appear), but the unreliable face looked to the correct box on only one out of every four trials” (Tummeltshammer et al., 2014, p. 3).	18
Figure 3: The three experimenters in three different colours: black, striped and white. Two of them are reliable (Reliable A, Reliable B) and one is Unreliable.	23
Figure 4: 3D drawing of the setup of the demonstration phase (dark grey: curtains, yellow: two screens, light blue: plexiglass panel forcing the dogs to make a choice of the screens by approaching from the side).	24
Figure 5: Experimental setup during the demonstration phase.	25
Figure 6: Wrong choice of the dog; food was removed behind the curtain on the other side.	28
Figure 7: Action crouching	30
Figure 8: Action lifting	30
Figure 9: Action sound	30
Figure 10: First transfer task, screen choice, both Informants (Reliable B and Unreliable) stand in front of the screens.	33
Figure 11: Second transfer task, pointing, both Informants (Reliable B and Unreliable) are pointing at a bowl.	34
Figure 12: Setup of pointing task during the transfer phase with measurements.	35
Figure 13: Third transfer task, begging, both Informants (Reliable B and Unreliable) hand a bowl with food in front of their body.	36
Figure 14: Setup of Begging task during the transfer phase with measurements.	37
Figure 15: Fourth transfer task, unsolvable task, both Informants (Reliable B and Unreliable) stand near the unsolvable task.	38
Figure 16: Setup of unsolvable task during the transfer phase with measurements.	39

Figure 17: Cue following proportion of both breed groups (cooperative and independent) in all three actions (crouch, lift and sound) with both informants (Reliable and Unreliable) during all demonstration trials.	49
Figure 18: Cue following of both breed groups (cooperative and independent) in all three actions (crouch, lift and sound) with both informants (Reliable and Unreliable) during the demonstration phase in their first trial per action and informant.	52
Figure 19: Hesitation of both breed groups (cooperative and independent) for all three actions (crouch, lift and sound) with both informants (Reliable and Unreliable) during the demonstration phase.	54
Figure 20: Proportion to choose Reliable over Unreliable in all four transfer tasks (Screen choice, Pointing, Begging and Unsolvable) for both breed groups (cooperative and independent).....	56
Figure 21: Proportion of time proximity to Reliable B over the total time they spent time near both experimenters.	58

List of Tables

Table 1: Number of participants of the breeds of the respective breed group, abbreviations and papers that have categorized these breeds as "cooperative" or "independent".	22
Table 2: Schedule of phases with number of trials for each informant and sequence of trial.	31
Table 3: Results of the full model - follow cue Demonstration phase (dependent variable: follow cue for all trials)	48
Table 4: Results of the full model - follow cue Demonstration phase (dependent variable follow choice during first exposure).	51
Table 5: Results of the full model - hesitation during Demonstration phase (dependent variable: hesitation length to make a choice).	53
Table 6: Results of the full model - choice during Transfer Tasks (dependent variable: preferred informant role).	55
Table 7: Results of the full model - proximity during Transfer Tasks (dependent variable: proportion of time spent close to Reliable B experimenter).	57
Table 8: Dog data (breed: see Table 1 for full breed name) and experimental data about the reliability and colour combination (e.g., striped shirt=Kinga, Reliable=Kinga, therefore Reliable A was striped and the person was Kinga.....	79

Table 8: Dog data (breed: see Table 1 for full breed name) and experimental data about the reliability and colour combination (e.g., striped shirt=Kinga, Reliable=Kinga, therefore Reliable A was striped and the person was Kinga).

Run ID	Breed	Sex	Age	Greeter	Striped shirt	Black shirt	White shirt	Reliable A	Reliable B	Unreliable	Training Criterion met
dog1	AS	F	8.7	Reliable A	Kinga	Franzi	Kea	striped	black	white	1
dog2	AK	F	12.3	Reliable A	Franzi	Kea	Kinga	striped	white	black	0
dog3	WT	F	5.8	Reliable A	Franzi	Kinga	Kea	white	striped	black	0
dog4	IGIT	M	7.2	Reliable A	Franzi	Kinga	Kea	striped	black	white	1
dog5	BT	F	3.3	Reliable A	Franzi	Kinga	Kea	white	black	striped	1
dog6	BLT	M	4.4	Reliable A	Franzi	Kinga	Kea	white	striped	black	1
dog7	BLT	M	13.4	Reliable A	Kea	Franzi	Kinga	black	striped	white	1
dog8	SSD	M	2.6	Reliable A	Franzi	Kea	Kinga	white	black	striped	1
dog9	SSD	F	11.1	Reliable A	Kea	Franzi	Kinga	black	white	striped	1
dog10	AS	F	6.3	Reliable A	Kea	Kinga	Franzi	striped	black	white	0
dog11	AS	M	4.3	Reliable A	Kinga	Franzi	Kea	striped	black	white	0
dog12	BC	M	5.8	Reliable A	Franzi	Kinga	Kea	white	black	striped	1
dog13	AS	F	3.1	Reliable A	Kea	Kinga	Franzi	black	striped	white	1
dog14	BC	M	6.8	Reliable A	Kinga	Franzi	Kea	white	striped	black	1
dog15	SSD	M	12.1	Reliable A	Kea	Kinga	Franzi	white	striped	black	0
dog16	SSD	M	6.2	Reliable A	Kea	Kinga	Franzi	striped	white	black	1
dog17	SSD	M	4.3	Reliable A	Kea	Franzi	Kinga	striped	black	white	0
dog18	BC	M	11.4	Reliable A	Kinga	Franzi	Kea	striped	black	white	0
dog19	PRT	F	11.0	Reliable A	Kea	Kinga	Franzi	black	striped	white	1
dog20	JRT	F	1.9	Reliable A	Franzi	Kea	Kinga	black	white	striped	1
dog21	JRT	F	2.1	Reliable A	Kinga	Kea	Franzi	striped	black	white	0
dog22	C	F	6.3	Reliable A	Franzi	Kinga	Kea	white	striped	black	1
dog26	PRT	M	5.0	Reliable A	Franzi	Kea	Kinga	striped	black	white	1
dog27	IGIT	F	5.6	Reliable A	Kea	Franzi	Kinga	black	striped	white	1
dog28	BC	M	7.2	Reliable A	Kea	Kinga	Franzi	black	striped	white	1
dog29	BC	M	2.7	Reliable A	Franzi	Kea	Kinga	white	striped	black	0
dog30	BC	M	10.8	Reliable A	Franzi	Kea	Kinga	black	white	striped	0
dog31	IGIT	F	9.3	Reliable A	Kinga	Franzi	Kea	black	white	striped	1
dog32	PRT	M	2.5	Reliable A	Franzi	Kea	Kinga	black	striped	white	1
dog33	BC	M	12.3	Reliable A	Kinga	Kea	Franzi	white	black	striped	0
dog34	PRT	F	8.4	Reliable A	Kea	Kinga	Franzi	striped	black	white	1
dog35	BC	M	5.8	Reliable A	Kinga	Franzi	Kea	black	white	striped	1
dog36	IGIT	F	2.6	Reliable A	Franzi	Kea	Kinga	white	black	striped	1
dog37	SSD	F	3.0	Reliable A	Kinga	Franzi	Kea	black	white	striped	1
dog38	WHT	F	2.7	Reliable A	Franzi	Kea	Kinga	striped	white	black	1
dog39	BC	F	1.4	Reliable A	Kea	Kinga	Franzi	black	white	striped	0
dog42	AS	F	9.3	Reliable A	Kea	Kinga	Franzi	striped	white	black	1

dog43	PRT	F	8.0	Reliable A	Kea	Kinga	Franzi	white	striped	black	1
dog44	BT	M	4.9	Reliable A	Kea	Franzi	Kinga	black	white	striped	0
dog45	AS	F	5.4	Reliable A	Kea	Franzi	Kinga	striped	black	white	1
dog46	AS	F	10.1	Reliable A	Kinga	Kea	Franzi	white	striped	black	0
dog47	AS	M	5.2	Reliable A	Kinga	Franzi	Kea	striped	black	white	1
dog49	BC	F	1.1	Reliable A	Kea	Kinga	Franzi	striped	black	white	1
dog50	BC	M	8.8	Reliable A	Kinga	Franzi	Kea	striped	black	white	1
dog51	BC	M	9.2	Reliable A	Kinga	Franzi	Kea	white	striped	black	1
dog52	AS	M	3.3	Reliable A	Franzi	Kea	Kinga	striped	black	white	1
dog53	FT	M	1.3	Reliable A	Kea	Franzi	Kinga	white	black	striped	1
dog54	PRT	M	1.2	Reliable A	Kinga	Franzi	Kea	black	striped	white	1
dog55	BT	F	6.6	Reliable A	Kea	Kinga	Franzi	striped	white	black	0
dog56	PRT	F	5.5	Reliable A	Franzi	Kinga	Kea	striped	white	black	1
dog57	PRT	F	4.3	Reliable A	Franzi	Kinga	Kea	striped	black	white	1
dog59	BT	F	5.9	Reliable A	Kea	Kinga	Franzi	striped	white	black	1
dog60	BT	M	8.3	Reliable A	Kinga	Franzi	Kea	white	striped	black	1
dog61	BC	M	3.2	Reliable A	Kea	Franzi	Kinga	black	white	striped	1