Animal Behaviour 227 (2025) 123268

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Inference in wolves and dogs: the ‘cups task’, revisited

Dániel Rivas-Blanco [a, b, *, 1], Sophia D. Krause [a, c, 1], Sarah Marshall-Pescini [a], Friederike Range [a]

[a] Domestication Lab, Department of Interdisciplinary Life Sciences, Konrad Lorenz Institute of Ethology, University of Veterinary Medicine Vienna, Vienna, Austria
[b] William James Center for Research, University of Aveiro, Aveiro, Portugal
[c] Institute of Biology I, Faculty of Biology, University of Freiburg, Freiburg, Germany

ARTICLE INFO

Article history:
Received 7 August 2024
Initial acceptance 19 November 2024
Final acceptance 10 April 2025
Available online 21 July 2025
MS. number: 24-00494R

Keywords:
Canis familiaris
Canis lupus
cups task
dog
inference
wolf

Inferential reasoning, which refers to the process of arriving at a conclusion from a series of premises, has been studied in a multitude of animal species by using the ‘cups task’ paradigm. In one of the versions of this set-up, two opaque cups (one baited and one empty) are shaken in front of the animal. As only the baited cup makes a noise when shaken, the animals can locate the reward by inferring that only a baited cup would make noise that an empty cup would make no noise or both. In a previous iteration of this paradigm in wolves, *Canis lupus*, and dogs, *Canis familiaris*, wolves seemed to outperform dogs. However, assessing neither each species' inference capabilities, nor their relationship with each other was not feasible because of the lack of control conditions. In this study, several conditions in which the baited cup, the empty cup or no cups are shaken are added to tackle this issue. Results indicate that wolves and dogs made their choices based on the saliency and order of the stimuli presented and not based on inference, which is consistent with the previous study. In addition, the potential causes behind the animals' performance are discussed, and alternative paradigms that may be more apt to measure inference abilities in wolves and dogs are proposed in this study.

Animals are exposed to a variety of stimuli linked to different opportunities and challenges. In making decisions and acting appropriately, animals must process the available information in their environment (Uexküll, 1957; Völter & Call, 2017). An example of processing method is inferential reasoning, which is the ability to derive a conclusion from previously known information (Völter & Call, 2017). The effects of inferential reasoning are sometimes difficult to distinguish from something that could be explained by associative learning alone (that is, linking stimuli that covary; Premack, 1995). Although associative learning involves associating events that are simultaneously occurring in space and time (and thus requires trial and error), reasoning combines perceived events with imagined ones or events that are not necessarily co-occurring (Maier & Schneirla, 1935; discussed in Völter & Call, 2017; Premack, 1995) by retrieving and recombining information (Zeithamova et al., 2012). Accordingly, inferences can be distinguished from associative learning based on the prior knowledge of the animal and whether or not trial-and-error learning occurs before arriving at the solution. Furthermore, inferential reasoning is responsible for the ability of animals to act appropriately even when only partial information is available (Premack, 1995; Völter & Call, 2017).

Diagnostic inference (that is, reasoning from the consequent [the effect] to the antecedent [the cause]) is a type of inferential reasoning. This reasoning is based on affirming the consequent or denying the consequent. Choosing to approach a rustling bush by inferring there may be prey hidden inside would be ‘affirming the consequent’. Conversely, avoiding a still bush (because the hunter infers that no prey is hidden in it) would be ‘denying the consequent’. The latter type of inferential reasoning has been shown much more rarely in nonhuman species (Völter & Call, 2017).

In analysing inference strategies in animals, the ‘cups task’ was used as the paradigm, in which subjects can choose between two opaque cups (one containing a reward while the other remains empty) based on the information given about the contents of the cups. For example, when a baited cup is shaken, an auditory cue in the form of a rattling sound is produced, which indicates the presence of the food reward. Choosing the baited cup by using this cue to infer the presence of its cause is an example of ‘affirming the consequent’ (Völter & Call, 2017). By contrast, avoiding the cup

* Corresponding author.
E-mail address: daniel.rivas-blanco@vetmeduni.ac.at (D. Rivas-Blanco).
[1] D. Rivas-Blanco and S. D. Krause contributed equally to this work.

https://doi.org/10.1016/j.anbehav.2025.123268

that does not make a sound when shaken (and is thus supposedly empty) would be ‘denying the consequent’ (Völter & Call, 2017).

Several species have been tested in some version of the ‘cups task’, such as African grey parrots (Schloegl et al., 2012), pigs (Nawroth & von Borell, 2015) and apes (Bräuer et al., 2006; Call, 2004, 2006; Hill et al., 2011), all of which have the ability to make diagnostic inferences. On the contrary, studies on domestic dogs using the cups paradigm have shown that dogs seem to lack causal understanding (Bräuer et al., 2006; Erdőhegyi et al., 2007). Lampe et al. (2017) compared dogs with wolves, which are their closest living relatives (Vila et al., 1997). Similar to the above-mentioned studies, dogs performed at the chance level, whereas wolves were able to solve the task above the chance level (Lampe et al., 2017).

The results of the study performed by Lampe and colleagues indicate that wolves may outperform dogs in causal tasks, which would lead to an effect of domestication on dogs' causal understanding. However, Lampe and colleagues only tested one condition (‘noise’ condition), where a baited and an empty cup were shaken. They also did not include any control conditions to eliminate other potential explanations. Notably, wolves may have been following the most salient stimulus, regardless of whether they were able to infer that the reward was there. Furthermore, the cups were always shaken in the same order (with the baited cup being shaken first and the empty one being shaken second), which leaves open the possibility of wolves solving the task by staying by the side where a cup was first shaken (regardless of whether they were able to infer the presence of the food). Consequently, which strategy the wolves may have used to solve this task (be it inference by affirming the consequent, denying the consequent or both) or whether inferential reasoning was involved at all cannot be ascertained because they may have chosen the ‘full cup’ as it represented the most salient stimulus (the one that produced a noise) or by choosing the very first object being moved (primacy effect).

Thus, this study aimed to investigate the inferential reasoning strategies used by wolves and dogs by replicating the ‘noise’ condition of Lampe et al. (2017) and adding further conditions. Aside from the ‘full information’ condition (both cups, one baited and one empty, are shaken; Lampe et al., 2017), the following conditions were included: (B) ‘affirming the consequent’ (only the baited cup is shaken), (C) ‘denying the consequent’ (only the empty cup is shaken) and the control condition or (D) ‘no information’ (no cup is shaken; Table 1, Fig. 1).

According to Frank's (1980) information processing hypothesis, in general, wolves perform better than dogs on inferential reasoning tasks. This hypothesis states that selection pressures on independent problem-solving skills and causal understanding were relaxed in dogs during domestication as a result of humans providing direct (in the case of pet dogs) or indirect (e.g. refuse to feed on in case of free-ranging dogs) care (Frank, 1980; Frank & Frank, 1982, 1985; Frank & Frank, 1987; Marshall-Pescini, Cafazzo, et al., 2017; Wynne, 2021). This is further compounded by wolves possibly relying on complex mental processes such as means−ends relationships, causal understanding and strategy while hunting (Frank & Frank, 1985). Several studies have observed better problem-solving skills in wolves than in dogs, which might be

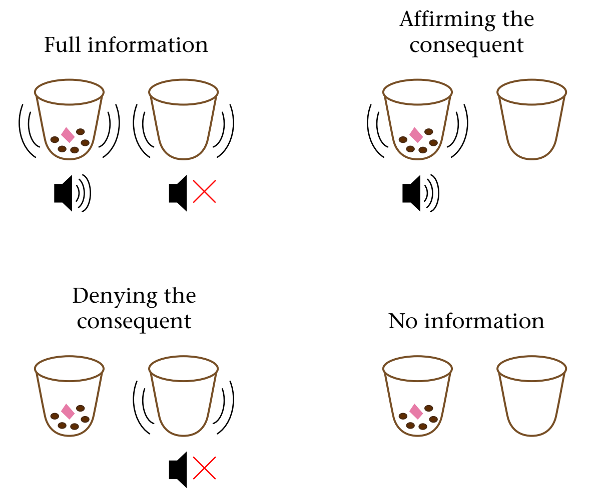

Figure 1. Schematics of all four conditions. The pictures represent the cups (light brown) with one of them containing five kibble pellets (dark brown) and one piece of the meat reward (pink). The kinetic lines indicate which cup is shaken in each condition, whereas the speaker icon represents the auditory information (rattling noise or silent) of the respective cup above it.

explained by the better causal understanding of wolves, but this difference could result from the higher persistence displayed by wolves in such tasks, leading to success by trial-and-error learning (Bensky et al., 2013; Frank & Frank, 1982, 1985; Frank, 2011; Hiestand, 2011; Marshall-Pescini, Virányi, et al., 2017; Udell, 2015).

Taken together, we expect wolves to have a better understanding of causation than dogs, performing better than chance in the ‘full information’ (both cups shaken)’ and both partial information conditions (‘affirming the consequent’ and ‘denying the consequent’). However, if the performance of wolves in the study of Lampe and colleagues was due to choosing the most salient stimulus (that is, a salience effect), then such performances were above the chance level in the ‘full information’ and ‘affirming the consequent’ conditions and below the chance level in the ‘denying the consequent’ condition. Moreover, if the subjects simply choose on the basis of the last or first stimulus shaken (that is, a recency effect or primacy effect, respectively), then they will choose the second or first shaken cup in the ‘full information’ condition and the shaken cup in the partial information conditions above the chance level. Furthermore, as no cues are given in the ‘no information’ condition, the subjects can perform at the chance level if they do not use any alternative strategies (such as choosing based on olfactory cues). A detailed overview of the expected performance for each possible strategy the animals might have used within the different conditions is given in Table 2.

METHODS

Subjects

In this study, wolves and dogs (hereafter referred to as ‘pack dogs’) that were raised under similar conditions at the Wolf

Table 1
Description of the conditions used in the experiment

Condition	Procedure	Stimulus output
Full information	Both cups were lifted and shaken one after the other, in a random order	Only one cup emitted a noise, the other did not
Affirming the consequent	Only the baited cup was lifted and shaken	Shaken cup emitted a noise
Denying the consequent	Only the empty cup was lifted and shaken	Shaken cup emitted no noise
No information	No cup was lifted or shaken	No information was given

The schematics of the conditions is shown in Fig. 1.

Table 2
Expected performance under each condition for the possible strategies the animals might use

Strategy	Full information	Partial information		No information
		Affirming the consequent	Denying the consequent	
Causal understanding	Baited cup	Baited cup	**Baited cup**[1]	Random/side-bias
Most salient stimulus	Baited cup	Baited cup	**Empty cup**	Random/side-bias
Last stimulus	**Last cup shaken**	Baited cup	Empty cup	Random/side-bias
First stimulus	**First cup shaken**	Baited cup	Empty cup	Random/side-bias
Lack of understanding	Random/side-bias	Random/side-bias	Random/side-bias	Random/side-bias

If 'baited cup' or 'empty cup' is stated in the table, then this option can be chosen above the chance level. The outcome shown in boldface indicates the result in the respective condition, which would be essential for differentiating among the possible strategies.
[1] According to Völter and Call (2017), the ability to reason by denying the consequent is less widespread among animals than the ability to reason by affirming the consequent. Therefore, random/side-bias results would be compatible with this strategy.

Science Center (WSC) in Ernstbrunn, Austria, were tested. The wolves ($N_{wolf} = 12$; 8 males; mean age = 106.5 ± 11.2 months) and pack dogs ($N_{pack\ dog} = 5$; 2 males, mean age = 98.4 ± 6.4 months) were hand raised, and they had permanent contact to humans until the age of 5 months before being introduced or separated into packs. The wolves and pack dogs were living under the same conditions in outdoor enclosures measuring 2000 to 10.000 m2.

Based on previous studies (Müller et al., 2016), dogs raised with different experiences in regard to object manipulations and opportunities to learn the physical properties of the environment should not differ in their inferential reasoning skills. Therefore, to increase our dog sample size, seven additional pet dogs were recruited ($N_{pet\ dog} = 7$; 5 males; mean age = 99.3 ± 12.5 months). These pet dogs were familiar with the WSC test environment (to minimize potential differences in testing) with four of them being former pack dogs raised at the WSC. Nevertheless, in accounting for any potential differences in understanding the contingencies of the task caused by pet dogs having more experience with human-created objects, the performance of both dog populations was compared before merging their data for further analyses (and only did so after finding that, indeed, no significant differences arose between them, as explained in Results).

Three animals (two wolves and one pet dog) were excluded because of motivational or health reasons (sight and hearing). Thus, they were not included in the statistical analyses. Therefore, for analyses, the sample size for wolves was $N_{wolf} = 10$, whereas that for pet dogs was $N_{pet\ dog} = 6$. Details about the subjects are provided in Table 3.

Table 3
Species, status, sex and age (at the time of the first training session) of animals

Subject	Species	Status	Sex	Age (months)
Chitto	Wolf	pack	M	114
Tala	*Wolf*	*pack*	*F*	*114*
Una	Wolf	pack	F	115
Nanuk	*Wolf*	*pack*	*M*	*150*
Taima	Wolf	pack	F	66
Tekoa	Wolf	pack	M	66
Geronimo	Wolf	pack	M	150
Yukon	Wolf	pack	F	150
Etu	Wolf	pack	M	71
Maikan	Wolf	pack	M	71
Kenai	Wolf	pack	M	143
Amarok	Wolf	pack	M	119
Hiari	Dog	pack	M	92
Imara	Dog	pack	F	92
Layla	Dog	pack	F	124
Panya	Dog	pack	F	92
Enzi	Dog	pack	M	92
Hakima	Dog	pet	M	136
Kilio	*Dog*	*pet*	*M*	*144*
Asali	Dog	pet	M	135
Freya	Dog	pet	F	80
Pepeo	Dog	pet	M	93
Zazu	Dog	pet	M	93
Coco	Dog	pet	F	59

Animals shown in italics were excluded from the experiment (and thus from any subsequent analyses) because of motivational or health problems. Refer to Table S1 for more details about the exact date of birth, exact date of the first training session, genetic relationships and breed information. F: female; M: male.

Experimental Conditions and Set-up

Testing environment

The experiment was performed at the WSC in Ernstbrunn, Austria. The experimental sessions were mostly conducted in two large fenced outdoor enclosures, which are used for testing purposes. However, five wolves were tested in their home enclosure because of their unwillingness to leave their enclosure. In this case, the other pack members were shifted to another enclosure during the test.

Apparatus

A table (150 cm length, 50 cm width and 113–143 cm height [adjustable]) positioned outside the enclosure behind the mesh-wire fence was used as the apparatus. The upper part of the table was surrounded by acrylic glass, through which the animals could observe the experimental procedure. In addition, a curtain (96 cm × 57 cm) was attached behind the acrylic glass in the middle of the upper part of the table, thereby concealing the experimenter and their actions to avoid providing unintended cues to the subject.

Two retractable targets, one on each side of the table, were used to allow the animals to indicate their choices in the experiment. Both targets were attached to a PVC pipe, which enables the experimenter to push them into the enclosure (within the animal's reach) and out of it again. In ensuring that the targets can move simultaneously, these two pipes were connected to each other by a third one (parallel to the long side of the table and perpendicular to the other two pipes) on the experimenter's side of the table.

For the test trials, two containers, which are plastic flower pots with a diameter of 14 cm, were used. The two cups were attached to wooden poles (45 cm in length) by two o-screws and two rubber bands, in such a way that the bands enveloped the cups while being inserted into the circular cavity of the o-screws. This contraption allowed the experimenter to shake the cups from behind the curtain with a minimum amount of movement and without having their hands within the subject's view (which could distract the subjects or allow them to receive unintentional cues from the experimenter).

To observe the subject's position and choice despite the curtain, a GoPro Hero 4 camera installed at the top of the table provided the experimenter with a live video feed on a tablet during the test and filmed the session. A second camera (Sony HDR) was placed outside the enclosure, which filmed the session diagonally sideways from the animal's point of view, to allow for double checking of the performance of the animals in each session at a later point.

A detailed illustration of the testing table is shown in Fig. 2.

Habituation

Three of the wolves that participated in this experiment (Maikan, Taima and Tekoa) have shown to be particularly neophobic in other experiments and during daily activities. Thus, a habituation phase preceded the experiment for these animals to reduce their fear towards the apparatus. In this phase, the animals were familiarized with the entire apparatus, especially the movable elements of it (moving targets, shaking cups and the noise they made) to ensure that they could freely interact with the affordances of the apparatus without fear. However, in ensuring that these animals did not have more experience on the relevant features of the task that may have given them an advantage over other subjects, they were not allowed to interact directly with the cup or see its contents. Furthermore, in preventing any potential associations from arising during the habituation phase, approaching the empty and baited cups was rewarded by the experimenter. The habituation phase lasted until the subject felt comfortable enough to interact with the table and participate in the test.

Experimental Procedure

General procedure

Prior to each training or testing day, the subjects were separated from their pack and shifted into the testing enclosure. Then, they were shortly acclimatized for 5 min, in which the subjects can roam freely in the enclosure.

Each animal was tested in approximately 2–3 sessions per day for 2–3 days a week. Each session had a time limit of 30 min. If the animal did not complete all trials within this time limit or showed no interest in the test (e.g. refused to be shifted into the test enclosure or did not approach the table), then the session was cancelled and repeated on a different day. After five consecutive cancelled sessions (training or testing), a clear lack of motivation for the test or the inability to participate in the test caused by health reasons, the animal was excluded from the experiment. Two wolves (Tala and Nanuk) and one pet dog (Kilio) were excluded in accordance with the above-mentioned criteria.

Training phase

During the training phase, the animals were trained to indicate their choice (‘left’ or ‘right’) by touching one of the targets on the side where the reward was placed. Each training session consisted of 12 trials. A training trial started by calling the animals' attention by either saying their name out loud or knocking on the centre of the table. Subsequently, a food reward (usually a small piece of beef or sausage, but chick pieces were used for some of the animals) was placed directly on one side of the table (outside of the area covered by the curtain), with the side on which the reward was placed (‘left’ or ‘right’) randomized and counterbalanced. Afterwards, the targets were pushed through the fence. If the subjects touched the target closest to the reward, then the experimenter grabbed the reward and threw it over the fence and into the enclosure. Conversely, if the subject touched the target on the side opposite to the reward, then the reward was taken back behind the curtain.

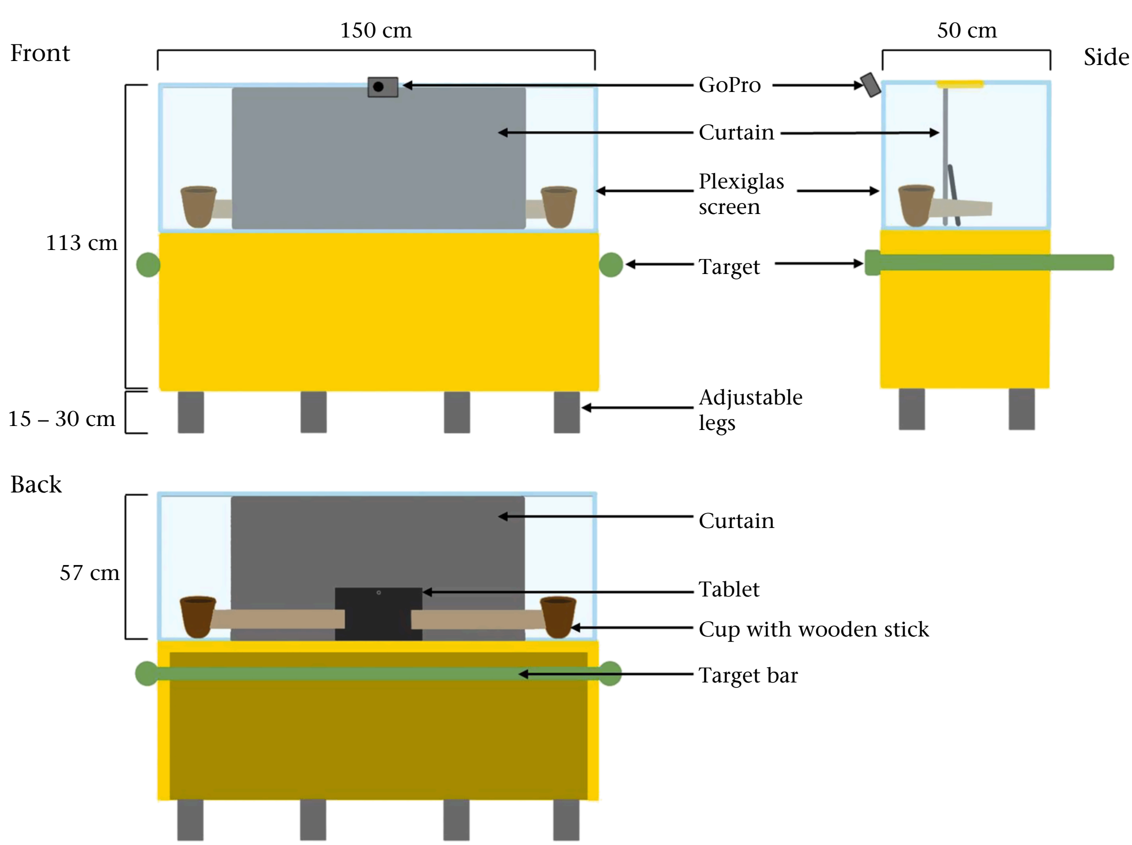

Figure 2. Schematics of the testing apparatus. The front side represents the perspective of the animal. This side was positioned against the enclosure fence from the outside. The back side was the perspective of the experimenter who was sitting or standing behind the table. The yellow parts represent the wooden parts of the table.

If the subjects chose the correct side in 10 out of 12 trials in two consecutive training sessions, then they moved forward to the testing phase. Overall, to reach this criterion, wolves needed to achieve an average of 2.9 ± 0.5 training sessions, while the pack and pet dogs needed to achieve 3.6 ± 1.2 and 2.8 ± 0.5 sessions, respectively. One of the pack dogs (Layla) needed five training sessions to move forward to the testing phase, but after failing three different pretrials in two different sessions, she received an additional three training sessions (for a total of eight). Furthermore, one of the pet dogs (Hakima) performed two additional training sessions because he was the pilot subject. Since this pilot was carried out more than 2 months before he was tested, we decided to give him an additional two training sessions. All these additional sessions were included in the above-mentioned averages. Details for the performance of each animal in the training sessions are presented in Table S1.

Testing phase

Pretest. Prior to each test session, the subject was tested for motivation, attentiveness and any side preferences in a pretest. This pre-test consisted of six trials, which are identical to the training trials described above. The subjects had to choose the correct side at least five out of those six trials to proceed with the experimental trials. If this criterion was not reached, then the animal got five more minutes of acclimatization, and the pretest was repeated. However, if the subject did not reach the criterion in two pretests on the same day, then the test was cancelled for that day and rescheduled. After three consecutive failed pretests, the animal went back to the training phase (this happened only once with one of the pack dogs named Layla).

Experimental trials

After completing the pretest, the subject proceeded with the experimental trials. Each animal was tested in eight test sessions, each session consisting of eight trials. Before a trial started, the experimenter prepared the cups behind the curtain by baiting one of them with a piece of food of the animal's liking (the same type of meat as in the training phase; baited cup) and adding five pieces of kibble (dry food) that produced a rattling sound when shaken, while the other cup remained empty (empty cup). At the beginning of each trial, the subject was lured away by throwing a piece of food far into the enclosure in line with the centre of the table to ensure that the animals approached the table centrally. This piece of food was usually a kibble pellet. In some cases, sausage or beef was used because of motivational reasons (but this piece of food was always considered of a 'lower quality' than the reward in the baited cup for each of the animals). As soon as the subject looked away, the experimenter placed the cups on the table, each one protruding out of each side of the curtain. Subsequently, the experimenter called the animal's name or knocked in the middle of the table to draw their attention back to the apparatus.

The cups were manipulated as soon as the subject looked in the direction of the table (something that the experimenter could see on the video feed of the GoPro camera). Depending on the condition, one, both (one after the other) or no cups were manipulated by lifting and shaking for several seconds until the subjects directed their attention towards them, with the experimenter placing the cups back on the table after doing so. After the manipulation, the experimenter pushed the two targets through the fence to allow them to make their choice by touching one of the targets with either their snout or their paws. After the animal touched one of the targets, both targets were pulled back to their original position behind the fence. If the subject chose the target on the side of the table where the baited cup was placed, then the choice was considered correct. In this case, the reward was taken out of the cup within the view of the subject and thrown over the fence into the enclosure. To do this, the experimenter reached into the cup with one of their hands, while the rest of their body remained concealed. Conversely, if the subject chose the side where the empty cup was placed, then the choice was considered incorrect. In this case, the cups were directly taken back behind the curtain, and no reward was given. Only the first touch to the targets was considered for the test and subsequent data analyses. That is, if the animals managed to touch a second target before the experimenter pulled the targets out of reach of the animals, then these additional touches were not taken into consideration.

The choice of the subjects was recorded by the experimenter on a prepared 'protocol sheet' and digitized after the session. This 'protocol sheet' also contained the information about each of the trials the experimenter needed to conduct, that is, the side on which the reward should be positioned and which cups (if any) should be shaken and in which sequence (in the case of the full condition).

A video of a sample trial (full information condition) can be found in Video S1.

Conditions

In examining the inferential reasoning abilities or potential alternative strategies the animals might have used in the study conducted by Lampe et al. (2017), the animals were confronted with four different conditions (three experimental [full information, affirming the consequent and denying the consequent] and one control condition [no information]).

In the 'full information' condition (similar to the one performed by Lampe and colleagues), both cups were shaken one after the other, with the baited cup emitting a rattling sound and the empty cup remaining silent. The order in which the cups were shaken in the 'full information' condition ('baited' or 'empty' first) was counterbalanced. Furthermore, in the 'affirming the consequent' condition, only the baited cup was shaken, emitting a rattling sound. On the contrary, in the 'denying the consequent' condition, only the empty cup was shaken, emitting no sound. In the 'no information' condition, no cup was shaken; thus, no information about the location of the reward was given to rule out possible alternative strategies the animals might use to solve the task.

Within each test session, two trials per condition were presented in a pseudo-randomized pattern. That is, trials of the same condition never occurred more than twice in a row, and the reward was not on the same side more than twice in a row.

A detailed overview of the conditions used in this experiment is shown in Table 1, and a graphical representation of these conditions is shown in Fig. 1.

Statistical Analyses

Analyses were performed using R 4.0.3 (R Core Team, 2020). In testing whether the performance of each species was above the chance level in each of the conditions, exact binomial tests (two-tailed) with a probability of success of 0.5 were used. To examine the effects of the condition and the species on the performance of the experiment, generalized linear mixed models (GLMM) were run using the 'glmer' function (package 'lme4'; Bates et al., 2015) with a binomial distribution and logit link function. Furthermore, we checked for model assumptions through custom functions designed by Roger Mundry and Remco Folkertsma.

Our main model aimed to investigate whether wolves and dogs performed differently across various conditions. Thus, the outcome of each trial ('success' or 'failure') was used as the response variable, while the species, condition and their interaction were used as predictor variables (fixed effects). We added the session in which each trial was performed as a control variable (fixed effect) to account for any variations in performance caused by learning or changes in motivation throughout the experiment. In addition, considering that the wolves were not fed every day (as opposed to the dogs), which might have caused a difference in motivation, the days that passed in between the date the animal was last fed and the day of the experiment (set as 0 if the subject was fed on the day before the test, 1 if one day had passed in between and so on) were added as a control variable (fixed effect). The identity of the subjects was added as a random effect, and random slopes were added whenever applicable.

In testing the effect of the interaction between the condition and the species, a χ^2 test was performed using 'ANOVA' ('car' package; Fox and Weisberg, 2019) to compare the full model described above with a null model in which the variables species and condition were removed. In investigating the effect of the species and condition, reduced models that removed the interaction between these two factors were run, which were equally compared with the null model through a χ^2 test.

Moreover, an analysis was conducted for the first two sessions only to examine the initial performance of animals in a manner comparable to the experiment performed by Lampe et al. (2017). In the initial experiment, the condition our paradigm was based upon was performed in four trials (in two different sessions) interspersed among trials testing for understanding of other causal cues and several social ones. The predictor and control variables were the same as the model described above (except for the session number, which was excluded because of the limited number of sessions analysed, which could have negatively impacted the model's stability). For the main model, a full-null comparison with a model was run in which the predictor variables (species, condition and the interaction between them) were removed. A reduced model (without the interaction between species and condition) was also run and compared with the null model.

Furthermore, to account for any potential alternative strategies that the animals might have used to solve the task (e.g. choosing the most salient or last stimulus), two more binomial tests (to compare performance against chance) and their respective models (to compare performance between the two species) were run. In checking for a potential preference towards the last cup shaken (recency effect), whether the animals chose the last cup shaken in the 'full information' condition was analysed, regardless of whether it made a rattling noise or not. This analysis also covered any potential preferences for the first cup shaken (primacy effect) because choosing the first cup above chance would be equivalent to choosing the last cup below chance. In addition, a model that investigated to which extent the animals chose the only shaken cup in the two partial information conditions (that is, the side with the most salient stimulus) was run to analyse the animals' preference towards the most salient stimulus (regardless of whether it was baited).

Plots were generated using the 'ggplot' function (package 'tidyverse'; Wickham et al., 2019). Significance brackets in the plots were added through the 'ggsignif' package (Ahlmann-Eltze & Patil, 2021), and the colours in Fig. 3 were adjusted through 'ggnewscale' (Campitelli, 2020).

Ethical Note

All procedures were approved by the 'Ethics and Animal Welfare Committee at the University of Veterinary Medicine Vienna' with the ETK number 'ETK-050/04/2021'. Participation in the experiment was voluntary for all animals. Throughout the testing sessions, the subjects had permanent access to water. With regard to the pet dogs, the owners gave their consent in advance, and they were informed beforehand about the procedure of the study. The animals could retreat into the enclosure at any time in case they were uncomfortable with the experiments.

RESULTS

Comparing Dog Populations (Pack Dogs versus Pet Dogs)

Pack and pet dogs showed similar performances in all conditions, as well as average of choices for the second cup in the full information condition and average of choices for the shaken cup in the partial information conditions (Table 4). Accordingly, no significant difference was found between pack dogs and pet dogs neither across sessions (full-null comparison: $\chi^2 = 5.054$, $P = 0.653$) nor in the first two sessions (full-null comparison: $\chi^2 = 5.010$, $P = 0.286$). Furthermore, no effect was found between pack and pet dogs regarding whether they chose the cup shaken

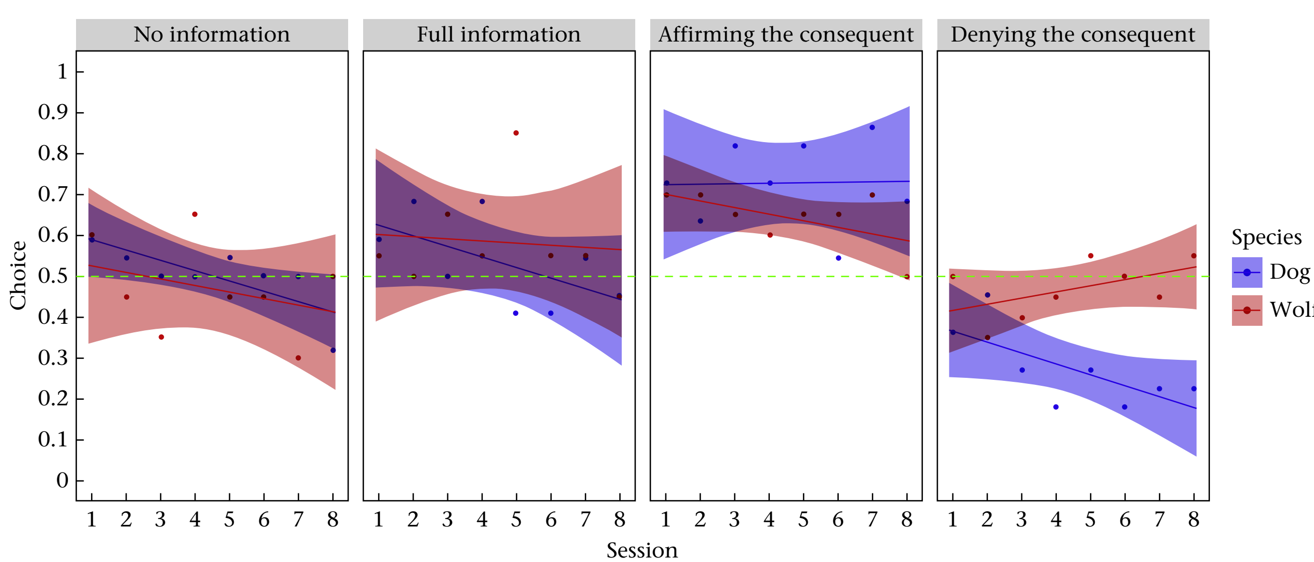

Figure 3. Performance over time of both species ($N_{wolf} = 10$; $N_{dog} = 11$) in each condition. Linear regression lines were applied by using a linear model. The green line indicates the probability of success at the chance level.

Table 4
Average correct choices for each condition and choices intended to measure potential biases (± SE) for pack and pet dogs

	Pack dogs	Pet dogs
No information	0.562 ± 0.223	0.447 ± 0.204
Full information	0.550 ± 0.224	0.521 ± 0.205
Affirming the consequent	0.700 ± 0.206	0.750 ± 0.177
Denying the consequent	0.325 ± 0.210	0.229 ± 0.172
Choosing the last cup (full information)	0.575 ± 0.222	0.625 ± 0.199
Choosing the shaken cup (partial information conditions)	0.688 ± 0.208	0.760 ± 0.174

second in the ‘full information’ condition (full-null comparison: $\chi^2 = 0.293$, $P = 0.588$) or whether they chose the shaken cup in the partial information conditions (full-null comparison: $\chi^2 = 2.347$, $P = 0.125$). Considering that pack and pet dogs did not show any significant differences in their performance, they were grouped together for further analyses, as experience seemed to have no influence on the performance of the dogs in this study. Consequently, only the two species groups (wolves and dogs) will be compared hereafter.

Comparing Species (Wolves versus Dogs) Across all Sessions (1–8)

All conditions: general performance

In all models, the ‘no information’ condition was taken into the intercept to compare the rest of the experimental conditions (‘full information’, ‘affirming the consequent’ and ‘denying the consequent’) with the control. The performance of the dogs was also taken into the intercept for all models.

Full model

Our analyses showed that the interaction between species and condition significantly influenced the performance of the animals (full-null comparison: $\chi^2 = 53.585$, $P < 0.001$). However, a significant difference was only found in the ‘denying the consequent’ condition ($GLMM_{full}$: estimate ± SE = 0.986 ± 0.320, $P = 0.002$; Table 5, Fig. 4) between the two species with dogs having worse performance (0.273 ± 0.135 on average) than wolves (0.469 ± 0.158) and the latter being at the chance level (Table S2). In all other experimental conditions, no significant difference was found between the two species (Table 5, Fig. 4).

Regarding the ‘full information’ condition, although no significant difference was found between the two species in the model (estimate ± SE = 0.318 ± 0.311, $P = 0.307$; Table 5), the performance of the wolves (0.581 ± 0.157 on average) was above chance

Table 5
Summary of the full model ($GLMM_{full}$) examining the effect of the interaction between species and condition in all sessions

Model terms	Estimate	SE	z value	*P*
(Intercept)	−0.019	0.158	(−0.123)	(0.902)
Dog versus wolf	−0.083	0.237	−0.352	0.725
No information versus full information	0.137	0.214	0.642	0.521
No information versus affirming the consequent	0.994	0.243	4.089	**<0.001**
No information versus denying the consequent	−0.985	0.227	−4.338	**<0.001**
Session no. (z-transformed)	−0.127	0.057	−2.209	**0.027**
Days passed between feeding and testing (z-transformed)	−0.035	0.073	−0.481	0.630
Dog versus wolf: full information	0.318	0.311	1.021	0.307
Dog versus wolf: affirming the consequent	−0.265	0.346	−0.767	0.443
Dog versus wolf: denying the consequent	0.986	0.320	3.084	**0.002**

The response variable (‘success’) was defined as the choice of the noisy (baited) cup. The predictors were species, condition and their interaction. The ‘no information’ condition served as our control condition; thus, it was defined as the baseline against which all experimental conditions were compared. Regarding species, dogs were defined as the baseline against which wolves were compared. The interaction of species and condition is indicated by ‘:’. Significant values are shown in boldface.

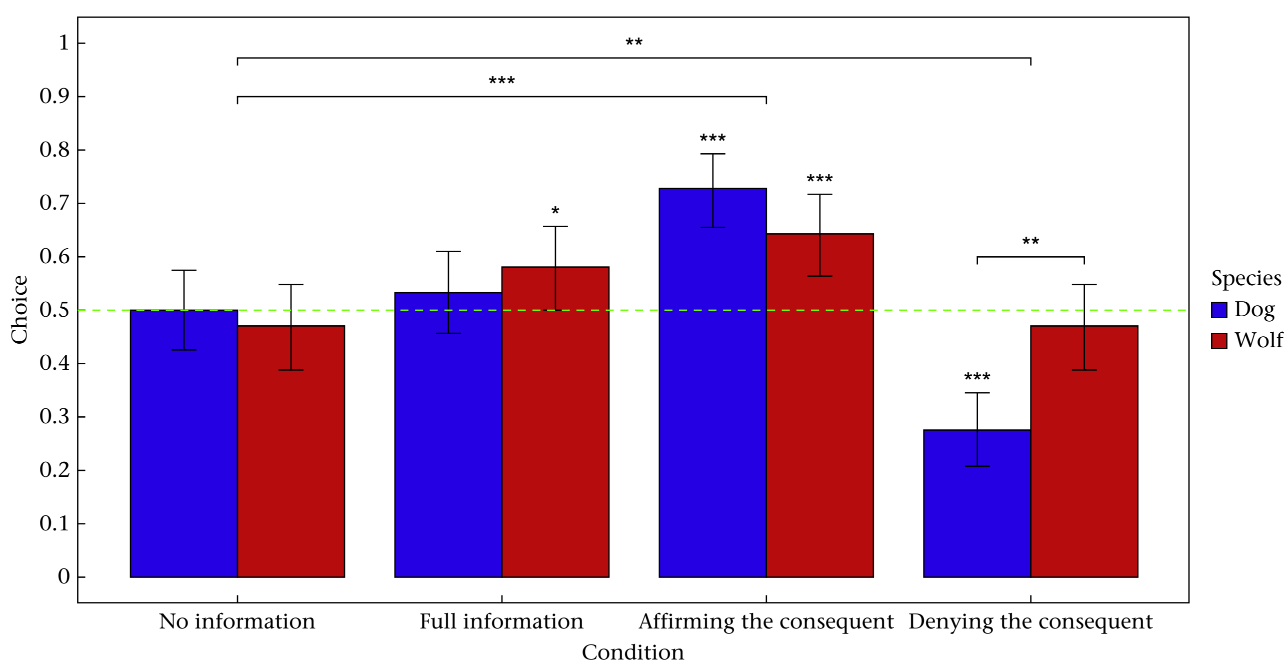

Figure 4. Performance of the two species ($N_{wolf} = 10$; $N_{dog} = 11$) in the respective conditions based on the mean success in all sessions. Error bars represent the 95% CI. The green line indicates the probability of success at the chance level. Performances above the chance level (two-tailed binomial test; probability of success = 0.5) are shown by black asterisks above the error bar, whereas performances below the chance level are shown by red asterisks above the error bar. The differences between the two species are based on the results of the full model ($GLMM_{full}$), which are shown by a black line with asterisks above them. The differences between conditions are based on the results of the reduced model ($GLMM_{reduced}$), which are shown by a black line with asterisks above them (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).

Table 6
Summary of the reduced model ($GLMM_{reduced}$) examining the effect of species and condition without the interaction of both in all sessions

Model terms	Estimate	SE	z value	*P*
(Intercept)	−0.143	0.132	(−1.087)	(0.277)
Dog versus wolf	0.179	0.142	1.260	0.207
No information versus full information	0.289	0.155	1.857	0.063
No information versus affirming the consequent	0.872	0.189	4.610	**<0.001**
No information versus denying the consequent	−0.494	0.158	−3.127	**0.002**
Session no. (z-transformed)	−0.125	0.057	−2.194	**0.028**
Days passed between feeding and testing (z-transformed)	−0.037	0.077	−0.477	0.634

The response variable ('success') was defined as the choice of the noisy (baited) cup. The predictors were species and condition, whereas the interaction between them was removed. The 'no information' condition served as our control condition; thus, it was defined as the baseline against which all experimental conditions were compared. Regarding species, dogs were defined as the baseline against which wolves were compared. Significant values are shown in boldface.

level, whereas that of the dogs (0.534 ± 0.151 on average) did not differ from the chance level (Table S2). However, the lower boundary of the 95% CI for the 'full information' condition of the wolves was almost at the 50% 'chance' level (0.501; Table S2). Thus, whether the wolves performed above the chance level remains unknown. In the 'affirming the consequent' condition, the wolves (0.644 ± 0.152 on average) and dogs (0.727 ± 0.135 on average) performed above the chance level (Table 6) with no significant difference between them (GLMMfull: estimate ± SE = -0.265 ± 0.346, $P = 0.443$; Table 5).

Reduced model

In testing whether there were differences between the conditions in both species and whether wolves and dogs had any general differences in performance, a reduced model was run by removing the interaction between condition and species. Analysis showed a significant difference between the reduced and the null model (reduced-null comparison: $\chi^2 = 38.934$, $P < 0.001$), with no effect of species (estimate ± SE = 0.179 ± 0.142, $P = 0.207$; Table 6). In addition, the session number showed a negative effect (estimate ± SE = -0.125 ± 0.057, $P = 0.028$; Table 6, Fig. 3), indicating that performance gradually decreased as the experiment progressed.

With regard to the effect of condition on performance, both partial information conditions ('affirming the consequent' and 'denying the consequent') differed significantly from the control condition ('no information', Table 6 and Fig. 4). Subjects chose the correct side more often in the 'affirming the consequent' condition than in the 'no information' condition (estimate ± SE = 0.872 ± 0.189, $P < 0.001$; Table 6). By contrast, subjects performed worse in the 'denying the consequent' condition than in the 'no information' condition (estimate ± SE = -0.494 ± 0.158, $P = 0.002$; Table 6).

Comparing Species (Wolves versus Dogs) and Conditions in Sessions 1 and 2

In the study conducted by Lampe et al. (2017), the animals were tested in only four trials (two trials per session; two sessions). Accordingly, the version of the cups task performed by Lampe et al. (2017) is equivalent to the first four trials of the 'full information' condition in this study. Consequently, the performance of the first two sessions was analysed separately to investigate the initial performance of the animals and compare the results from the 'full information' condition with the study of Lampe et al. (2017).

In this model, the interaction between species and condition had no effect on the performance of the animals (full-null comparison: $\chi^2 = 12.094$, $P = 0.097$). A second model was also fit in which the interaction was removed to test its effect on the performance of species and condition separately ($GLMM_{reduced}$). A significant difference was found between this model and the null (reduced-null comparison: $\chi^2 = 11.157$, $P = 0.024$); however, none of the experimental conditions differed significantly from the control (Table 7, Fig. 5).

Furthermore, species had no effect on the performance in the first two sessions, as the performance of the dogs did not differ significantly from that of the wolves (estimate ± SE = -0.154 ± 0.302, $P = 0.611$; Table 7). Both species performed at the chance level in the 'no information', 'full information' and 'denying the consequent' conditions (Table S3). In the 'affirming the consequent' condition, wolves (0.700 ± 0.147) and dogs (0.681 ± 0.142) performed above the chance level (Table S3, Fig. 5).

Bias Towards the Second Cup Shaken in the 'Full Information Condition'

Based on the performance in the full information condition, whether the animals used inferential reasoning, followed by the most salient stimulus, or chose the last stimulus presented remains unknown. Therefore, the choice of the last cup shaken in the 'full information' condition was examined for further analysis.

The results indicate that wolves and dogs chose the cup shaken last in the 'full information' condition above the chance level (binomial test: wolves: probability of success: 0.656, C.I. 95%: 0.577−0.729, $P < 0.001$; dogs: probability of success: 0.602, C.I. 95%: 0.526−0.675, $P = 0.017$). Therefore, the choices of both species were at least partially informed by the position of the last cup shaken (recency effect).

In addition, a χ^2 test was performed to test whether the two species showed differences in this preference towards the second cup shaken, thereby comparing the full model that included species as a predictor variable with a null model without this factor. No significant differences were found between the two models (full-null comparison: $\chi^2 = 1.198$, $P = 0.274$), indicating that both species showed no differences in their biases towards the second/last cup.

However, considering the individual performance of the animals, the results indicate large variation among subjects in both species regarding the choice of the cup shaken second when it was baited or not (Fig. S1). For example, three wolves (Taima, Amarok and Etu) and three dogs (Imara, Asali and Zazu) chose the cup

Table 7
Summary of the reduced model ($GLMM_{reduced}$) examining the effect of species and condition without the interaction of both in the first two sessions

Model terms	Estimate	SE	*z value*	*P*
(Intercept)	0.265	0.263	(1.006)	(0.314)
Dog versus wolf	−0.154	0.302	−0.509	0.611
No information versus full information	0.146	0.312	0.469	0.639
No information versus affirming the consequent	0.613	0.323	1.901	0.057
No information versus denying the consequent	−0.535	0.325	−1.647	0.100
Days passed between feeding and testing (z-transformed)	0.019	0.151	0.128	0.898

The response variable ('success') was defined as the choice of the noisy (baited) cup. The predictors were species and condition, whereas the interaction between them was removed. The 'no information' condition served as our control condition; thus, it was defined as the baseline against which all experimental conditions were compared. Regarding species, dogs were defined as the baseline against which wolves were compared.

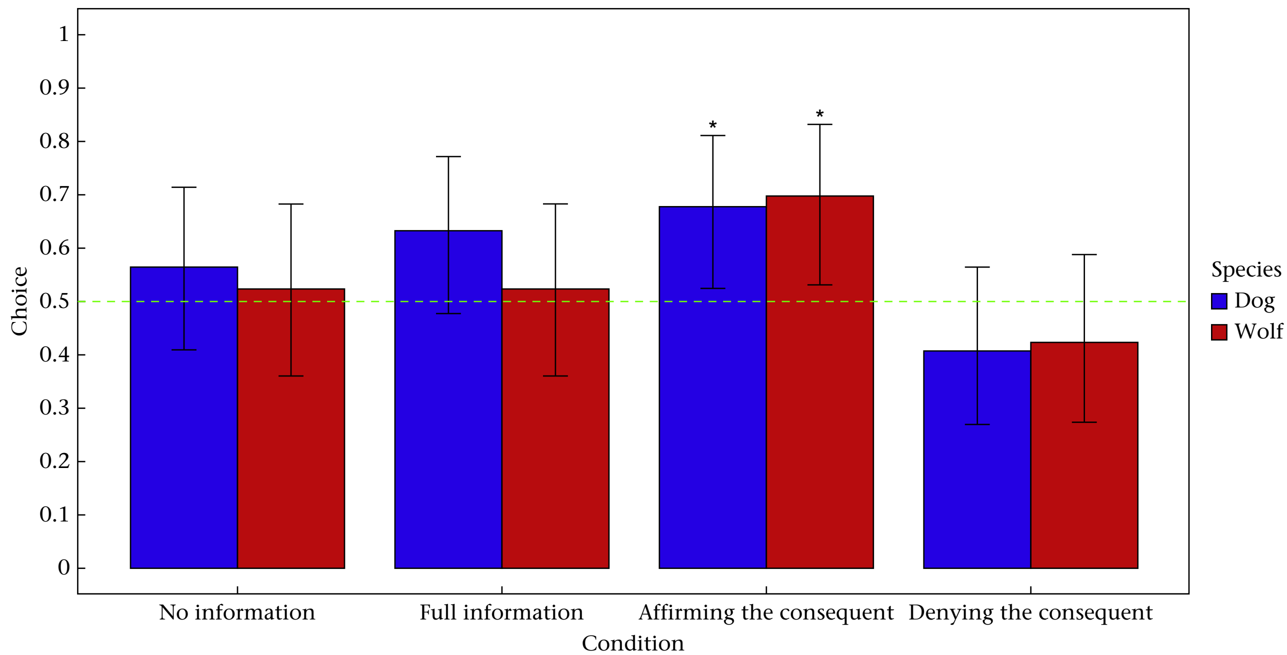

Figure 5. Performance of the two species ($N_{wolf} = 10$; $N_{dog} = 11$) in their respective conditions based on the mean success in the first two sessions. Error bars represent the 95% CI. The red line indicates the probability of success at the chance level. Performances above the chance level (two-tailed binomial test; probability of success = 0.5) are shown by black asterisks above the error bar (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$).

shaken second in 12 or more out of 16 trials (Fig. S2). Only one wolf (Chitto) chose the baited cup in the full information condition in a total of 12 trials (six trials baited cup shaken first; six trials baited cup shaken second) and only one dog (Hakima) in a total of 11 trials (four trials baited cup shaken first; seven trials baited cup shaken second), whereas no other subject achieved these performances in the 'full information' condition (Fig. S1).

Bias Towards the Most Salient Stimulus in the Partial Information Conditions

In investigating which strategies the animals used, whether the animals had a preference towards the most salient stimulus (that is, the cup that was shaken) in the partial information conditions ('affirming the consequent' and 'denying the consequent') was analysed.

When taking the data from the two partial information conditions, both species chose the cup that was shaken (salience effect), regardless of whether it was the baited or empty cup, above the chance level (binomial test: wolves: probability of success: 0.588, C.I. 95%: 0.531–0.642, $P = 0.002$; dogs: probability of success: 0.727, 95% CI: 0.678–0.773, $P < 0.001$). There was an effect of species (full-null comparison: $\chi^2 = 7.401$, $P = 0.007$) with a significant difference between wolves and dogs ($GLMM_{full}$: estimate ± SE = −0.566 ± 0.206, $P = 0.006$), particularly dogs choosing the cup that was shaken (most salient stimulus) significantly more often than wolves (Table S4).

DISCUSSION

In this study, whether wolves and dogs (pack and pet dogs) use inferential reasoning to infer the location of a hidden reward was investigated by using the presence or absence of sound to inform their choices. Our results did not reveal any significant differences between pet dogs and pack dogs; thus, the data of both populations were pooled. Wolves and dogs solved the 'affirming the consequent' (only baited cup shaken) condition but not the 'full information' condition (both cups shaken). Furthermore, dogs performed below the chance level in the 'denying the consequent' condition (only empty cup shaken). Both species had a preference towards the last cup shaken in the full information condition and towards the only cup shaken in the partial information conditions.

Considering our results, the strategy that was most likely used by both species does not necessarily involve the use of inferential reasoning but rather relies on the animals responding to other simple cues available. The choices made by the wolves and dogs corresponded to a recency effect (bias towards the last stimulus presented) and a salience effect (bias towards the most conspicuous stimulus presented), respectively, with the latter being more pronounced in the case of the dogs.

The salience effect is well known for influencing associative learning processes, with the stimuli that are proportionally more intense (in comparison with other stimuli) favouring faster associations (Frieman & Reilly, 2015). This effect has been suspected to interfere with a multitude of experiments in the field of animal behaviour, with one stimulus or feature of the paradigm attracting the attention of the subjects regardless of whether it was conducive to solve the task (see Eckert et al., 2022; Lazzaroni et al., 2020; Marshall-Pescini et al., 2012 for a few examples). With regard to studies examining inference in animals, the possibility that a different stimulus might overshadow their potential inferential reasoning abilities has been considered several times in the past (see Mikolasch et al., 2012; O'Hara et al., 2015; O'Hara et al., 2016; Shaw et al., 2013 for a few examples). In our experiment, the subjects only focused on the movement of the cups, which is consistent with the results of previous studies on inference in dogs (Bräuer et al., 2006; Erdőhegyi et al., 2007). This observation could be attributed to the prey drive of species (Bastos et al., 2024), their history of training and participation with other tasks that imply movement or both.

Recency effects are equally well known, and they have been studied in humans and nonhuman animals across a range of cognitive tasks (Castro & Larsen, 1992; Murphy et al., 2006; Rubio et al., 2023; Thomas et al., 1984). For example, dogs place a higher value upon the most recently-acquired information when it comes to foraging (Devenport & Devenport, 1993). Accordingly, the movement of the first cup could have been 'de-valued' by the time the second cup was presented. This issue could theoretically be tackled in future versions of this paradigm by presenting both stimuli simultaneously. However, presenting both cups simultaneously could risk the animals failing to pay attention to both of them at the same time because of physical limitations of our apparatus (chief among which, the fact that the targets needed to be far enough from each other to avoid cross-associations with the wrong target).

Collectively, wolves and dogs responded to the saliency of the stimuli and the order in which they were presented rather than performing any inferences about the position of the reward. Using these strategies is neither optimal nor disadvantageous. From the way our sessions were presented, choosing the most salient stimulus (which, judging by our results, would be the last cup that was moved) would yield a reward half of the time (as there was an equal number of trials for all four conditions: $P_{\text{full information}} = 0.5$, $P_{\text{affirming}} = 1$, $P_{\text{denying}} = 0$, $P_{\text{no info}} = 0.5$; $P_{\text{total}} = 0.5$), which may have been good enough for the animals to lack the need towards developing a better strategy. By contrast, our subjects were able to solve the training trials above the chance level and the functionally equivalent pretest trials at the beginning of each session. However, the animals were trained to solve those trials, which was reinforced on each testing day. Furthermore, those trials were visually distinct (because of the absence of the cups) to allow the animals to develop different strategies to solve testing and training/pretesting trials.

Another strategy that we encountered is side-bias, as several individuals participating in this study (five out of 10 wolves, four out of five pack dogs and two out of six pet dogs) showed a tendency towards choosing one side of the apparatus (Table S5). towards one side when two options are presented to the animals is common in experiments involving non-human animals especially those involving complex paradigms (see Hare & Tomasello, 1999; Laschober et al., 2021; Szabo et al., 2019 for a few examples; see Bolló et al., 2023 for a study investigating the potential mechanisms behind side bias in dogs). Although research on the proximate causes of side biases remains scant, one of the potential reasons behind it is that the animals engage with the tasks in which this bias presents itself as if they were operant conditioning tasks with a variable ratio schedule of reinforcement (that is, a reinforcement scheme in which rewards are handed out with a certain probability but with no way of knowing which of the responses will be reinforced; Miltenberger, 2016). Thus, if the animal ignores the stimuli presented and engages solely (or most often) with one of the options, then this behaviour would be strongly reinforced, as rewards would be provided half of the time but at random intervals from the animal's perspective (a process that has been described as the reason for the addictiveness of gambling; Lozano Bleda & Pérez Nieto, 2012).

Overall, the results of this study are consistent with the available literature for the 'cups task' in dogs (Bräuer et al., 2006; Lampe et al., 2017) but not so for the one in wolves (Lampe et al., 2017). In our study, wolves did not show a better performance in the full information condition than in the control, although they did solve the task above the chance level in the study conducted by Lampe and colleagues. Although the performance of the animals was only analysed in the first four trials of our 'full information' condition to make our study more comparable with that of Lampe et al. wolves did not perform above the chance level in our experiment, and no differences were found between the two species. These differences could be due to two main procedural differences: First, Lampe and colleagues tested the animals in two different causal conditions ('shape': two wooden shapes lying on the table, one hiding food underneath; 'noise': two shaken containers, one of which was baited and emitted a noise when shaken) and in several social conditions that involved witnessing human actions. Thus, these social conditions can influence the animals' performance by increasing the animals' attention towards the experiment. By contrast, our study did not include any social conditions, which might have reduced the animals' motivation. A decrease in performance was also observed throughout our experiment, which could be explained by losing motivation towards carrying out the experiment. Second, the two cups were always shaken in the same order in the 'noise' condition in Lampe's study: first the baited cup and then the empty cup. Considering that the most salient stimulus was always presented first in the study by Lampe and colleagues, saliency may have played an even larger role in that study than in ours, as going towards the most salient stimulus and ignoring any other, less salient one presented afterwards could explain the results observed and would have been the optimal strategy in this scenario. The recency effect in our study could be due to the increased complexity as a result of (1) counterbalancing, as the animals had to choose the first object and ignore the second in half of the trials while the opposite held true in the other half, or (2) presenting too many trials that may have looked too similar to each other in practice despite their different functions, which may have confused the animals. In particular, in the study by Lampe and colleagues, the wolves may have had an easier time acquiring an association towards the baited cup, as it was the most salient stimulus, and its order of presentation remained consistent. However, in our study, the presence of other conditions, as well as the counterbalancing of the order in which the cups were presented, may have introduced additional cues that had an impact on the animals' choice, such as increasing their disposition towards choosing the cup that was moved last (which was now the correct choice in half of the trials).

At first glance, the performance of the dogs in the previous study was at chance level, but dogs seemed to have tended to select the most recent stimulus (percentage of trials in which the dogs chose the last stimulus: pack dogs: 0.61 ± 0.14; pet dogs: 0.59 ± 0.16). This statistical nonsignificance could be explained by dogs having received on average less trials than wolves (wolves: four trials for all animals; pack dogs 3.17 ± 0.30 trials per animal; pet dogs 3.20 ± 0.33 trials per animal), allowing for the possibility of dogs showing a significant recency effect in Lampe and colleagues' study as well if they were tested in four trials. Alternatively, the performance of the dogs in Lampe et al.' s study could also be explained by saliency alone, assuming that the most salient feature for them was the movement in and of itself rather than the combination of movement and noise (as it seemed to be the case in our experiment).

Collectively, our results and those presented by Lampe and colleagues can be explained through the effect of saliency (and to a lesser extent, recency) alone, albeit these effects did not appear in quite the same manner between species and studies.

Within the wider context of the application of the auditory version of the cups task across species, the performance of wolves and dogs is related to that of non-ape animals (and even some ape species, with orangutans showing only partial preference towards the container that made a sound or no preference at all; Call, 2004; De Petrillo & Rosati, 2020; Hill et al., 2011). For example, squirrel monkeys (Marsh et al., 2015) solved the task above the chance level when only the baited cup was shaken, did so as well when the baited and empty cups were shaken (although one of the three subjects failed to do so), and chose below the chance level when only the empty cup was shaken. This performance is slightly similar to what we observed in our study (and equally telling of a potential salience effect).

However, comparisons among studies following this paradigm are difficult to make in general, as different versions of it may change the testing regiment in small but crucial ways. For example, in an experiment conducted in African grey parrots (Schloegl et al., 2012), the subjects that were not successful in the equivalents to our full information and affirming the consequent conditions underwent a training phase before being tested again. In the training phase, they were shown the inside of the empty cup and the fact that it made no noise when shaken and that dropping a piece of food inside the cup and shaking it afterwards would make a noise. Similarly, in another experiment conducted in capuchin monkeys (Sabbatini & Visalberghi, 2008), the animals were allowed to freely interact with the containers after finding that only a few animals were able to solve the task. They were retested after this exposure to a higher rate of success (one out of eight animals solving the task in the first testing phase versus three out of eight in the one after the exposure). Compared with the aforementioned studies, these training methods were not used in our study, as our subjects never saw the contents of the cups throughout the experiment, and no other demonstration that could convey that a baited cup would make noise and empty cup would make none was provided either.

The question arises as to what prior knowledge of the contingencies should be expected of the animals attempting to solve this task. From a modern human's umwelt (sensu Uexküll & Mackinnon, 1926J. von. Uexküll & Mackinnon, 1926), a noise coming from a moving opaque container would signify the presence of something inside it. However, most animals lack any prior experience with this particular scenario (thus, they are not equipped to interact with the task in ways beyond the saliency of the stimulus at hand). All of our subjects had prior experience with the sound of kibble pellets against plastic because of their daily feeding and other experiments (some of them had been previously tested in Lampe et al.' s version of the experiment), but none of them had experience with the specific features of this version of the experiment. Therefore, although a capacity to make inferences based on auditory stimuli, which may be used in the wild by predators to locate their prey, could be utilized to solve a task, whether such an ability would require prior experience with the stimulus used to make inferences about it remains unclear.

To tackle the issues intrinsic to this paradigm (and in particular, to the current study's implementation of it), we propose future endeavours regarding the study of inferential reasoning in wolves and dogs to focus on the use of touchscreens and similar automatized apparatuses. Touchscreen-based apparatuses have been successfully used in dogs and wolves in the past (Aust et al., 2008; Dale et al., 2019; Laude et al., 2016; Range et al., 2008; Rivas-Blanco et al., 2020), and they have been successfully used to investigate inferential reasoning capabilities in other species (Aust et al., 2008; Mikolasch et al., 2013; O'Hara et al., 2015, 2016). The stimulus presented through such means would be simpler and easier to control, thereby reducing the possibility of animals making associations towards irrelevant features of the task and tackling the question of how much previous knowledge on the contingencies of the task would be required for it to be solved (as all the animals would need to be trained on the stimuli presented). Furthermore, by removing the need for an experimenter, any potential distractions and cueing derived from their presence would be removed as well. In addition, an apparatus such as this could either use motion sensors to start each trial only when the subject is positioned at a certain distance of the apparatus or start the trials by the animal's contact with the apparatus, thereby eliminating the need to use food to reposition them between trials (and, thus, the potentially confounding factor of being rewarded even when the incorrect choice was made). Moreover, insights gained through experiments like these could be used as a basis to design future studies with physical stimuli, which would further pose the question of not only whether dogs and wolves are capable of inferring, but if they are, in which way they utilize such abilities in their natural environment.

Notably, new physical versions of the cups task have been recently developed and used to tackle potential alternative explanations such as the ones found in the current experiment. These new paradigms increase the complexity of the task by adding extra cups and grouping them by the amount of information given to the animals. Consequently, choosing one side results in the subject always getting food, but choosing the other gives them a 50% chance (Leahy & Carey, 2020; Mody & Carey, 2016). These new versions have yielded mixed results thus far in nonhuman animals, with several primate species being successful in a version with three cups (Call, 2022) and in one African grey parrot (Pepperberg et al., 2019) and four baboons (Ferrigno et al., 2021) being successful in a version with four cups. However, according to Call (2022), this latter case should be viewed with caution, as the baboons may have learned the discrimination over the course of the task. Similarly, chimpanzees also struggled when presented with three or four cups, although the study's last experiment showed that the last case may have been at least partially a result of the working memory demands posed by the task (Engelmann et al., 2023). Nevertheless, these more complex versions (at least under their current formulation) might not be appropriate for our target species because of the performance shown by wolves and dogs with the current paradigm. Indeed, the presence of additional cups, as well as the extended presentation time, may give wolves and dogs additional confounding factors to latch on to (and also increased working memory demands; Engelmann et al., 2023).

Conclusions

In this study, neither sufficient evidence that wolves use causal reasoning nor any differences in performance were found between wolves and dogs. Our results indicate that wolves and dogs favoured the most salient stimulus and the most recent one presented potentially because of the confounding factors that arose from the manner in which the paradigm was implemented. Thus, whether the animals lack the ability for inferential reasoning or rather it was the experimental design of the current study remains unclear, which caused unintended distractions that biased the performance of the animals. Future studies should consider the use of automatized apparatuses to reduce these potential distractions.

Author Contributions

Sophia D. Krause: Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Data curation. **Sarah Marshall-Pescini:** Writing – review & editing, Supervision, Conceptualization. **Friederike Range:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Dániel Rivas-Blanco:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation.

Data Availability

All data generated or analysed during this study are included as supplementary material.

Declaration of Interest

The authors declare no competing interests.

Acknowledgments

We thank Kathryn Roddie and the trainers at the Wolf Science Center Core facility for their contribution to data collection. We also thank Erik Hollerweger and Kathryn Roddie for crafting the apparatus used in the experiment. This study was funded by the Austrian Science Fund (Fonds zur Förderung der wissenschaftlichen Forschung, grant number P33928–B). DRB was funded by a Marietta Blau Grant from the Austrian Agency for International Cooperation in Education, Science and Research (grant number MMC-2023-07030).

Supplementary Material

Supplementary material associated with this article is available at https://doi.org/10.1016/j.anbehav.2025.123268.

References

Ahlmann-Eltze, C., & Patil, I. (2021). Ggsignif: R package for displaying significance brackets for 'ggplot2. https://doi.org/10.21105/joss.03138

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

Bastos, A. P., Warren, E., & Krupenye, C. (2024). What evidence can validate a dog training method? *Learning & Behavior*, 1–2. https://doi.org/10.3758/s13420-024-00658-9

Bensky, M. K., Gosling, S. D., & Sinn, D. L. (2013). The world from a dog's point of view: A review and synthesis of dog cognition research. *Advances in the Study of Behavior, 45*, 209–406. https://doi.org/10.1016/B978-0-12-407186-5.00005-7

Bolló, H., File, B., Topál, J., & Kis, A. (2023). Side bias behaviour in dogs shows parallels to the hemispatial neglect syndrome. *Applied Animal Behaviour Science, 263*(9), Article 105921. https://doi.org/10.1016/j.applanim.2023.105921

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

Call, J. (2004). Inferences about the location of food in the great apes (*Pan paniscus, Pan troglodytes, Gorilla gorilla*, and *Pongo pygmaeus*). *Journal of Comparative Psychology, 118*(2), 232–241. https://doi.org/10.1037/0735-7036.118.2.232

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

Call, J. (2022). The "avoid the empty cup" hypothesis does not explain great apes'(*Gorilla gorilla, Pan paniscus, Pan troglodytes, Pongo abelii*) responses in two three-cup one-item inference by exclusion tasks. *Journal of Comparative Psychology, 136*(3), 172. https://doi.org/10.1037/com0000321

Campitelli, E. (2020). Ggnewscale: Multiple fill and colour scales in 'ggplot2'. *R package version, 0.4*, 5.

Castro, C. A., & Larsen, T. (1992). Primacy and recency effects in nonhuman primates. *Journal of Experimental Psychology: Animal Behavior Processes, 18*(4), 335–340. https://doi.org/10.1037/0097-7403.18.4.335

De Petrillo, F., & Rosati, A. G. (2020). Logical inferences from visual and auditory information in ruffed lemurs and sifakas. *Animal Behaviour, 164*(1), 193–204. https://doi.org/10.1016/j.anbehav.2020.03.010

Devenport, J. A., & Devenport, L. D. (1993). Time-dependent decisions in dogs (*Canis familiaris*). *Journal of Comparative Psychology, 107*(2), 169–173. https://doi.org/10.1037/0735-7036.107.2.169

Eckert, J., Bohn, M., & Spaethe, J. (2022). Does quantity matter to a stingless bee? *Animal Cognition, 25*(3), 617–629. https://doi.org/10.1007/s10071-021-01581-6

Engelmann, J. M., Haux, L. M., Völter, C., Schleihauf, H., Call, J., Rakoczy, H., & Herrmann, E. (2023). Do chimpanzees reason logically? *Child Development, 94*(5), 1102–1116. https://doi.org/10.1111/cdev.13861

Erdőhegyi, Á., Topál, J., Virányi, Z., & Miklósi, Á. (2007). Dog-logic: Inferential reasoning in a two-way choice task and its restricted use. *Animal Behaviour, 74*(4), 725–737. https://doi.org/10.1016/j.anbehav.2007.03.004

Ferrigno, S., Huang, Y., & Cantlon, J. F. (2021). Reasoning through the disjunctive syllogism in monkeys. *Psychological Science, 32*(2), 292–300. https://doi.org/10.1177/0956797620971653

Frank, H. (1980). Evolution of canine information processing under conditions of natural and artificial selection. *Zeitschrift für Tierpsychologie, 53*(4), 389–399. https://doi.org/10.1111/j.1439-0310.1980.tb01059.x

Frank, H. (2011). Wolves, dogs, rearing and reinforcement: Complex interactions underlying species differences in training and problem-solving performance. *Behavior Genetics, 41*(6), 830–839. https://doi.org/10.1007/s10519-011-9454-5

Frank, H., & Frank, M. G. (1982). Comparison of problem-solving performance in six-week-old wolves and dogs. *Animal Behaviour, 30*(1), 95–98. https://doi.org/10.1016/S0003-3472(82)80241-8

Frank, H., & Frank, M. G. (1985). Comparative manipulation-test performance in ten-week-old wolves (*Canis lupus*) and Alaskan malamutes (*Canis familiaris*): A Piagetian interpretation. *Journal of Comparative Psychology, 99*(3), 266–274. https://doi.org/10.1037/0735-7036.99.3.266

Frank, H., & Frank, M. G. (1987). The University of Michigan canine information-processing project (1979-1981). In H. Frank (Ed.), *Man and wolf: Advances, issues, and problems in captive wolf research* (Ed., pp. 143–167). Dr W Junk Publishers.

Frieman, J., & Reilly, S. (2015). *Learning: A behavioral, cognitive, and evolutionary synthesis*. Sage Publications.

Hare, B., & Tomasello, M. (1999). Domestic dogs (*Canis familiaris*) use human and conspecific social cues to locate hidden food. *Journal of Comparative Psychology, 113*(2), 173–177.

Hiestand, L. (2011). A comparison of problem-solving and spatial orientation in the wolf (*Canis lupus*) and dog (*Canis familiaris*). *Behavior Genetics, 41*(6), 840–857. https://doi.org/10.1007/s10519-011-9455-4

Hill, A., Collier-Baker, E., & Suddendorf, T. (2011). Inferential reasoning by exclusion in great apes, lesser apes, and spider monkeys. *Journal of Comparative Psychology, 125*(1), 91–103. https://doi.org/10.1037/a0020867

Lampe, M., Bräuer, J., Kaminski, J., & Virányi, Z. (2017). The effects of domestication and ontogeny on cognition in dogs and wolves. *Scientific Reports, 7*(1), Article 11690. https://www.nature.com/articles/s41598-017-12055-6

Laschober, M., Mundry, R., Huber, L., & Schwing, R. (2021). Kea (*Nestor notabilis*) show flexibility and individuality in within-session reversal learning tasks. *Animal Cognition, 24*(6), 1339–1351. https://doi.org/10.1007/s10071-021-01524-1

Laude, J. R., Pattison, K. F., Rayburn-Reeves, R. M., Michler, D. M., & Zentall, T. R. (2016). Who are the real bird brains? Qualitative differences in behavioral flexibility between dogs (*Canis familiaris*) and pigeons (*Columba livia*). *Animal Cognition, 19*(1), 163–169. https://doi.org/10.1007/s10071-015-0923-8

Lazzaroni, M., Marshall-Pescini, S., Manzenreiter, H., Gosch, S., Přibilová, L., Darc, L., McGetrick, J., & Range, F. (2020). Why do dogs look back at the human in an impossible task? Looking back behaviour may be over-interpreted. *Animal Cognition, 23*(3), 427–441. https://link.springer.com/article/10.1007/s10071-020-01345-8.

Leahy, B. P., & Carey, S. E. (2020). The acquisition of modal concepts. *Trends in Cognitive Sciences, 24*(1), 65–78. https://doi.org/10.1016/j.tics.2019.11.004

Lozano Bleda, J. H., & Pérez Nieto, M.Á. (2012). Impulsivity, intelligence, and discriminating reinforcement contingencies in a fixed-ratio 3 schedule. *Spanish Journal of Psychology, 15*(3), 922–929. https://doi.org/10.5209/rev_sjop.2012.v15.n3.39384

Müller, C. A., Riemer, S., Virányi, Z., Huber, L., & Range, F. (2016). Inhibitory control, but not prolonged object-related experience appears to affect physical problem-solving performance of pet dogs. *PLoS One, 11*(2), Article e0147753. https://doi.org/10.1371/journal.pone.0147753

Marsh, H. L., Vining, A. Q., Levendoski, E. K., & Judge, P. G. (2015). Inference by exclusion in lion-tailed macaques (*Macaca silenus*), a hamadryas baboon (*Papio hamadryas*), capuchins (*Sapajus apella*), and squirrel monkeys (*Saimiri sciureus*). *Journal of Comparative Psychology, 129*(3), 256–267. https://psycnet.apa.org/doi/10.1037/a0039316.

Marshall-Pescini, S., Cafazzo, S., Virányi, Z., & Range, F. (2017a). Integrating social ecology in explanations of wolf–dog behavioral differences. *Current Opinion in Behavioral Sciences, 16*(2), 80–86. https://doi.org/10.1016/j.cobeha.2017.05.002

Marshall-Pescini, S., Passalacqua, C., Miletto Petrazzini, M. E., Valsecchi, P., & Prato-Previde, E. (2012). Do dogs (*Canis lupus familiaris*) make counterproductive choices because they are sensitive to human ostensive cues? *PLoS One, 7*(4), Article e35437. https://doi.org/10.1371/journal.pone.0035437

Marshall-Pescini, S., Virányi, Z., Kubinyi, E., & Range, F. (2017b). Motivational factors underlying problem solving: Comparing wolf and dog puppies' explorative and neophobic behaviors at 5, 6, and 8 weeks of age. *Frontiers in Psychology, 8*, 180. https://doi.org/10.3389/fpsyg.2017.00180

Mikolasch, S., Kotrschal, K., & Schloegl, C. (2012). Is caching the key to exclusion in corvids? The case of carrion crows (*Corvus corone corone*). *Animal Cognition, 15*(1), 73–82. https://doi.org/10.1007/s10071-011-0434-1

Mikolasch, S., Kotrschal, K., & Schloegl, C. (2013). Transitive inference in jackdaws (*Corvus monedula*). *Behavioural Processes, 92*, 113–117. https://doi.org/10.1016/j.beproc.2012.10.017

Miltenberger, R. G. (2016). *Behavior modification: Principles and procedures*. Cengage Learning.

Mody, S., & Carey, S. (2016). The emergence of reasoning by the disjunctive syllogism in early childhood. *Cognition, 154*, 40–48. https://doi.org/10.1016/j.cognition.2016.05.012

Murphy, J., Hofacker, C., & Mizerski, R. (2006). Primacy and recency effects on clicking behavior. *Journal of Computer-Mediated Communication, 11*(2), 522–535. https://psycnet.apa.org/doi/10.1111/j.1083-6101.2006.00025.x

Nawroth, C., & von Borell, E. (2015). Domestic pigs' (*Sus scrofa domestica*) use of direct and indirect visual and auditory cues in an object choice task. *Animal Cognition, 18*(3), 757–766. https://doi.org/10.1007/s10071-015-0842-8

O'Hara, M., Auersperg, A. M., Bugnyar, T., & Huber, L. (2015). Inference by exclusion in Goffin cockatoos (*Cacatua goffini*). *PLoS One, 10*(8), Article e0134894. https://doi.org/10.1371/journal.pone.0134894

O'Hara, M., Schwing, R., Federspiel, I., Gajdon, G. K., & Huber, L. (2016). Reasoning by exclusion in the kea (*Nestor notabilis*). *Animal Cognition, 19*(5), 965–975. https://link.springer.com/article/10.1007/s10071-016-0998-x

Pepperberg, I. M., Gray, S. L., Mody, S., Cornero, F. M., & Carey, S. (2019). Logical reasoning by a Grey parrot? A case study of the disjunctive syllogism. *Behaviour, 156*(5-8), 409–445. https://doi.org/10.1163/1568539X-00003528

Premack, D. (1995). Cause/induced motion: Intention/spontaneous motion. In J.-P. Changeux, & J. Chavaillon (Eds.), *Origins of the human brain* (pp. 286–309). Clarendon Press/Oxford University Press.

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

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

Rivas-Blanco, D., Pohl, I. M., Dale, R., Heberlein, M. T. E., & Range, F. (2020). Wolves and dogs may rely on non-numerical cues in quantity discrimination tasks when given the choice. *Frontiers in Psychology, 11*, 2412. https://doi.org/10.3389/fpsyg.2020.573317

Rubio, P., Muñiz-Moreno, J., & Loy, I. (2023). Primacy and recency in snails (*Cornu aspersum*). *Journal of Experimental Psychology: Animal Learning and Cognition, 49*(4), 226–236. https://doi.org/10.1037/xan0000365

Sabbatini, G., & Visalberghi, E. (2008). Inferences about the location of food in capuchin monkeys (*Cebus apella*) in two sensory modalities. *Journal of Comparative Psychology, 122*(2), 156–166. https://doi.org/10.1037/0735-7036.122.2.156

Schloegl, C., Schmidt, J., Boeckle, M., Weiß, B. M., & Kotrschal, K. (2012). Grey parrots use inferential reasoning based on acoustic cues alone. *Proceedings of the Royal Society B: Biological Sciences, 279*(1745), 4135–4142. https://doi.org/10.1098/rspb.2012.1292

Shaw, R. C., Plotnik, J. M., & Clayton, N. S. (2013). Exclusion in corvids: The performance of food-caching Eurasian jays (*Garrulus glandarius*). *Journal of Comparative Psychology, 127*(4), 428–435. https://psycnet.apa.org/doi/10.1037/a0032010

Szabo, B., Noble, D. W., & Whiting, M. J. (2019). Context-specific response inhibition and differential impact of a learning bias in a lizard. *Animal Cognition, 22*(3), 317–329. https://doi.org/10.1007/s10071-019-01245-6

Thomas, D. R., Moye, T. B., & Kimose, E. (1984). The recency effect in pigeons' long-term memory. *Animal Learning & Behavior, 12*(1), 21–28. https://psycnet.apa.org/doi/10.3758/BF03199809

Udell, M. A. R. (2015). When dogs look back: Inhibition of independent problem-solving behaviour in domestic dogs (*Canis lupus familiaris*) compared with wolves (*Canis lupus*). *Biology Letters, 11*(9), Article 20150489. https://doi.org/10.1098/rsbl.2015.0489

Uexküll, L. von (1957). A stroll through the worlds of animals and men. In C. H. Schiller (Ed.), *Instinctive behavior, the development of modern a concept*. New York, NY: International University Press.

Uexküll, J. von, & Mackinnon, D. L. (1926). In J. von Uexküll, & K. Paul (Eds.), *Theoretical biology*. Trench, Trubner & Co., Ltd.

Völter, C. J., & Call, J. (2017). Causal and inferential reasoning in animals. In J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.), *APA handbook of comparative psychology: Perception, learning, and cognition* (pp. 643–671). Washington: American Psychological Association. https://doi.org/10.1037/0000012-029

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wynne, C. D. L. (2021). Dogs' (*Canis lupus familiaris*) behavioral adaptations to a human-dominated niche: A review and novel hypothesis. *Advances in the Study of Behavior, 53*, 97–162. https://doi.org/10.1016/bs.asb.2021.03.004

Zeithamova, D., Schlichting, M. L., & Preston, A. R. (2012). The hippocampus and inferential reasoning: Building memories to navigate future decisions. *Frontiers in Human Neuroscience, 6*, 70. https://doi.org/10.3389/fnhum.2012.00070