



HAL
open science

Both sheep and goats can solve inferential by exclusion tasks

Josselin Duffrene, Odile Petit, Bernard Thierry, Raymond Nowak, Valérie Dufour

► **To cite this version:**

Josselin Duffrene, Odile Petit, Bernard Thierry, Raymond Nowak, Valérie Dufour. Both sheep and goats can solve inferential by exclusion tasks. *Animal Cognition*, 2022, 25 (6), pp.1631-1644. 10.1007/s10071-022-01656-y . hal-03753254

HAL Id: hal-03753254

<https://hal.inrae.fr/hal-03753254>

Submitted on 14 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Both sheep and goats can solve inferential by exclusion tasks**

2 Josselin Duffrene¹, Odile Petit², Bernard Thierry², Raymond Nowak², Valérie Dufour²

3

4 ¹University of Liège

5 ²Physiologie de la Reproduction et des Comportements, CNRS, IFCE, INRAE, Université de
6 Tours, PRC, F-37380, Nouzilly, France

7

8 Corresponding author: Valérie Dufour, valerie.dufour@inrae.fr

9 Orcid: 0000-0003-1766-9598

10

11

12 **Acknowledgments**

13 We thank the director and the keepers of the animal sanctuary of the Friedel farm, and the town
14 council of the city of Illkirsch-Graffenstaden for allowing this study.

15

16 Abstract

17 Despite the domestication of sheep and goats by humans for several millennia, we still lack
18 comparative data on their cognitive capacity. Comparing the cognitive skills of farm animals
19 can help understand the evolution of cognition. In this study, we compared the performances of
20 sheep and goats in inference by exclusion tasks. We implemented two tasks, namely a cup task
21 and a tube task, to identify whether success in solving the task could be attributed to either low
22 level mechanisms (avoiding the empty location strategy) or to deductive reasoning (if two
23 possibilities A and B, but not A, then it must be B). In contrast to a previous study comparing
24 goats and sheep in a cup task, we showed that both species solved the inferential condition with
25 high success rates. In the tube task, performances could not be explained by alternative
26 strategies such as avoiding the empty cup or preferring the bent tube. When applying a strict
27 set of criteria concerning responses in all conditions and controlling for the potential effects of
28 experience, we demonstrate that two individuals, a goat and a sheep, fulfil these criteria. This
29 suggests that sheep and goats are able to make inferences based on deductive reasoning.

30

31 Keywords: Domestic animal cognition, Cup task, Tube task, Deductive reasoning, Inference

32

33

34 **Introduction**

35 Most of the cognition studies carried out over the last decades have focused on non-human
36 primates, corvids and canids, all of which are considered key species to understand the
37 evolution of cognition in species closely related to humans or in more distant species that are
38 expected to have particularly sophisticated cognitive abilities (Shettleworth 2009). However,
39 the study of species such as farm animals can help to understand the evolution of cognition,
40 because domestication also acts as a driving force that may have influenced overall levels of
41 cognitive skills and/or emotional control (Boissy and Erhard 2014). For instance, with cattle,
42 the significant brain reduction observed in highly selected breeds of dairy and beef cows (*Bos*
43 *taurus*) compared to extinct wild bovids (aurochs, *Bos primigenius*), shows that exposure to
44 human contact and artificial selection affect brain size (Balcarcel et al. 2021). Similar brain
45 reduction has been reported in other domestic species such as dogs, pigs and sheep, and could
46 be related to the loss of the cognitive skills necessary to survival in the wild (Ebinger 1974;
47 Price 2002). Relative brain size can serve as a proxy for the cognitive abilities of species.
48 However, the cognitive abilities of animals can best be measured through testing them in
49 experimental tasks.

50 Inferential reasoning tasks are useful tools for comparing cognitive performances across
51 species. Inferential reasoning involves associating perceivable and absent elements to solve a
52 problem (Premack 1995). One type of inferential reasoning, called transitive inference, is the
53 capacity to derive a relation between two elements, based on premises about other elements (If
54 $A > B$, and $B > C$, then $A > C$). Another type of inference, namely inference by exclusion, is the
55 ability to select an option by systematically excluding other alternative options. Two types of
56 hypothesis have been proposed to explain the differences reported between related species in
57 inference by exclusion tasks. The first type favours the role of social factors in the evolution of
58 inference skills. Group living may have fostered the emergence of transitive ability, as group
59 members are more likely to survive in a hierarchically ranked society if they accurately
60 understand this type of organization (Schusterman et al. 2000). It could also be that species
61 differ in their levels of social complexity, as observed in Tonkean macaques (*Macaca*
62 *tonkeana*), which are more frequently engaged in cooperative and collective behaviours than
63 rhesus macaques (*M. mulatta*), and perform better in inference by exclusion tasks (Petit et al.
64 2015). The second type of hypothesis emphasizes the role of ecological factors. For instance,
65 cognitive specialization such as caching could account for better skills in inferential tasks in
66 caching birds compared to non-caching birds (Schloegl et al. 2009) – although this is not
67 entirely supported by other experimental work (Shaw et al. 2013). Other authors argue that the

68 better performances of goats (*Capras hircus*) compared to sheep (*Ovis aries*) in the inferential
69 domain are due to differences in foraging selectivity (Nawroth et al. 2014). A higher foraging
70 selectivity in goats compared to sheep would indeed reflect more cognitive flexibility in the
71 former compared to the latter.

72 To test inference by exclusion, researchers usually investigate the response of animals
73 using a “cup task”. In this task, subjects face two upturned opaque cups, one of which covers a
74 food item. Subjects are shown the empty cup, and are then offered a choice between the two
75 cups. They must infer from the absence of a cue from under a given cup (e.g., cup A) that the
76 other location (cup B) should be investigated (Völter and Call 2017). Note that this task can
77 involve a visual modality (the cue is a visible reward) and/or an auditory modality (the cue is
78 the sound produced by the presence of a reward). In principle, this task should tap into the
79 deductive reasoning skills of animals (if two possibilities A or B, but not A, then it must be B,
80 Watanabe and Huber 2006; O’Hara et al. 2016) and require high-level cognition mechanisms
81 (Call 2004; Watanabe and Huber 2006; Penn and Povinelli 2007; Danel et al. 2021). However,
82 the task could also be solved with the simple “avoid the empty location” decision rule (Paukner
83 et al. 2009; Jelbert et al. 2015), which involves lower-level cognitive mechanisms. More than
84 thirty different species (primates, birds, canids, dolphins, small ruminants) have been shown to
85 successfully solve inference by exclusion tasks (Call and Carpenter 2001; Call 2004, 2006; Aust
86 et al. 2008; Sabbatini and Visalberghi 2008; Schmitt and Fischer 2009; Paukner et al. 2009;
87 Schloegl et al. 2009; Hill et al. 2011; Nawroth et al. 2014; Völter and Call 2017; Danel et al.
88 2021), but it is generally not clear whether this was through deductive reasoning or by learning
89 to avoid the empty location (Jelbert et al. 2015). An alternative way to evaluate if decisions are
90 made based on reasoning or alternative mechanisms is the tube task initially developed by Call
91 and Carpenter (2001) for studies in primates. This task has also been used for birds by Shloegl
92 and colleagues (2009), and was recently adapted by Jelbert and colleagues (Jelbert et al. 2015).
93 In this task, subjects are presented with two tubes, one straight and one bent (Figure 1). If the
94 food is located in the bent part of the tube (thus invisible), subjects only see an empty straight
95 tube and an empty part of the bent tube (the part facing them). They must therefore infer that
96 the only plausible place for the food is in the bent section of the bent tube, i.e., the part to which
97 they have no visual access. Individuals must therefore choose one of the two options, and
98 avoiding the empty location is no longer an effective strategy. Compared to the cup task, the
99 tube task reduces the risk of adopting an alternative strategy such as “avoid the empty tube”,
100 and is thus more suitable for testing inference by exclusion skills.

101 A recent study by Nawroth et al. (2014) compared the performances of sheep and goats in an
102 inference by exclusion task that used a cup task involving a visual clue (Nawroth et al. 2014).
103 In a first version of the task, sheep and goats performed poorly (less than 50% of success in the
104 inference by exclusion condition), but when the experiment was repeated with new conditions
105 designed to control for the possible occurrence of local enhancement (i.e. drawing attention to
106 a particular cup by manipulating it differently than the other), goats reached 65% success levels.
107 This result is better than that obtained in sheep, which performed at random (Nawroth et al.
108 2014). According to the authors, this could be explained by a difference in cognitive flexibility
109 and food selectivity. Indeed, sheep are generalist grazers while goats are more selective. The
110 need to search for unexpectedly distributed food could impose a greater cognitive challenge
111 (De Petrillo and Rosati 2020). However, the results of a study exploring this question in two
112 species of lemur (Ruffed lemur, *Varecia variegata* and Sifaka, *Propithecus verreauxi*) did not
113 support this hypothesis (De Petrillo and Rosati 2020). Alternative explanations can be proposed
114 for the improved success of goats between the first experiment and the second. For example,
115 goats may have learned to solve the task through repeated testing (Völter and Call 2017). This
116 type of resolution involves low-level cognitive mechanisms rather than deductive reasoning,
117 but the cup task does not allow differentiation between the mechanisms involved. Sheep may
118 simply be unable to solve this type of task (even through low-level cognitive mechanisms).
119 However, this failure may not be related to differences in feeding selectivity, but could rather
120 be due to alternative factors such as shyness and lack of experience with cognitive tests, and/or
121 different living conditions compared to the goats compared in this study. Indeed, the goats used
122 by Nawroth et al. (2014) had already participated in cognitive studies, and the two species were
123 tested at different sites with very different general living conditions. To fully compare these
124 two species in terms of inference ability and confirm that goats perform better than sheep, it is
125 necessary to replicate the experiment of Nawroth et al. (2014) by studying goats and sheep
126 raised under similar conditions and with the same level of expertise for experimental testing
127 procedures. The studies should also use a tube task to verify whether the decisions are explained
128 by low-level (i.e., avoidance of the empty location) or inferential strategies.

129

130 Sheep and goats are small ruminants that are morphologically, physiologically and
131 phylogenetically close. Both species have undergone domestication for a period of 10,000 years
132 (Larson and Fuller 2014) and have a long history of dependence on humans, who have selected
133 individuals to obtain animals that are as docile, robust and cohesive as possible. Goats and
134 sheep exhibit contrasting behaviours; goats are often considered smarter and more inquisitive

135 than sheep, which are seen to be shy and fearful animals that tend to follow others (Miranda-de
136 la Lama and Mattiello 2010; Houpt 2018). These differences have not yet been rigorously
137 measured. In the domain of physical cognition, it is known that sheep have good spatial memory
138 (Hunter et al. 2015), and that goats can form categories based on abstract symbols, solve a
139 detour task and understand object permanence (Nawroth et al. 2019). However, studies directly
140 comparing their behaviour and cognition are scarce (Lyons et al. 1993; Greaves and
141 Wedderburn 1995). In the domain of social cognition, we know that sheep distinguish between
142 familiar faces (Kendrick et al 1996) and naïve lambs can learn to drink from an artificial teat
143 faster when exposed to expert lambs (Veissier and Stefanova 1993), but there is little knowledge
144 about other socio-cognitive skills. Goats can locate food based on human cues such as touching
145 or pointing, but not by following the direction of gazing; although they evaluate human
146 attentional states and can take the perspective of others into account (Kaminski et al. 2006;
147 Nawroth 2017), they mostly learn from their own experience to solve a food puzzle despite
148 being exposed to other, experienced individuals (Briefer et al. 2014).

149 We still lack comparative data on the behaviour and cognitive abilities of sheep and
150 goats, despite having lived in proximity with them for eleven millennia. To investigate the
151 inferential skills of these two species, we conducted two inferential by exclusion reasoning
152 experiments involving individuals from both species living in similar conditions, location and
153 routines. We used a standard inference by exclusion task (with cups that were either lifted or
154 shaken, Nawroth et al. 2014) and an adapted version of the tube task (Jelbert et al. 2015) to
155 control for alternative resolution strategies to deductive reasoning. If our results are consistent
156 with those of Nawroth et al.'s (2014) study and confirm common beliefs about the respective
157 abilities of sheep and goats, we can expect better performances on inference by exclusion tasks
158 in goats. There is however a paucity of quantitative comparisons between the two species and
159 we should remain open to any other possible outcome. If the cognitive abilities of sheep and
160 goats are similar, we should not find significant contrasts in their performances in inference by
161 exclusion tasks.

162

163 **Methods**

164 1) Subjects and conditions

165 Ten goats aged 1-12 years (4 females, 6 males) and ten sheep aged 2-10 years (2 females, 8
166 males) participated in the study (Table 1). The sheep and goats were housed together in the
167 Friedel animal park in Illkirch-Graffenstaden, France, which is an animal sanctuary open to the
168 public. They lived in groups with other animals (ponies, chickens, pigs, peacocks, cows). All

169 had free access to a wooded enclosure, with the possibility of sheltering at several sites with
170 roofs and straw bedding. They were fed every morning (hay, straw, wheat flour), but could also
171 graze on grass in the enclosure, and received varied food enrichment (fresh fruits and
172 vegetables, branches) at other times of the day. Water was available ad libitum. The sheep and
173 goats had no prior experience of participating in cognitive tests. At the time of testing, the
174 enclosure was not open to the public.

175 The tests were conducted in one of the shelters where the animals sleep at night. It is 7 x
176 7 m square, is covered with a roof and has wooden board fences that are 1.20 m high, thus
177 allowing an individual standing on hind legs to look outside. The animals were familiar with
178 this site, which partially prevented them from seeing their counterparts whilst continuing to
179 hear them. Sheep and goats were tested every afternoon from February 2021 to May 2021, from
180 1:30 to 5:00 pm. It was not always possible to test all 20 individuals every day. On one occasion,
181 towards the end of the study, two subjects were tested twice on the same day, with at least two
182 hours between each test session.

183

184 2) Material and food choice tests

185 The experimenter was seated behind a table (Height: 50 cm, Length: 52 cm, Width: 80 cm)
186 equipped with a grid to reduce access to the table by the subject. Subjects could insert their
187 snout through the bars of the grid when facing a choice test. The two containers (cups or tubes)
188 were placed on the table, approximately 30 cm apart (Figure 1a). For the “cup task”, the study
189 material was composed of opaque cups, transparent cups, larger opaque cups that could cover
190 the smaller cups, and a sheet of cardboard to hide the handling of the cups by the experimenter
191 (Figure 1b). For the “bent tube task”, the material was composed of bent tubes, an opaque over-
192 cup, straight tubes, and a sheet of cardboard to hide the handling of these items by the
193 experimenter (Figure 1c).

194 In a preliminary step, food preference tests were conducted (1) to familiarize subjects
195 with the set up, the test and the presence of a human, and (2) to determine the most palatable
196 reward for both species. The food choice tests were performed with two transparent cups,
197 similar to those used later on in the “cup protocol”. We conducted a session of ten choices
198 between a piece of carrot and a piece of apple, then a session of ten choices between a piece of
199 carrot and a piece of lettuce. All items were equally attractive for the sheep, but the slice of
200 carrot (about 3 mm thick and 2 cm as diameter) was retained as the reward for the tests after
201 goats showed a slight tendency to prefer carrots to other items.

202

203 3) Experimental design and testing procedure

204 We randomly divided subjects into two groups, each composed of 5 sheep and 5 goats. The first
205 group was tested with Nawroth et al.'s (2014) "cup protocol" controlling for solving by local
206 enhancement, then with the "bent tube protocol" controlling for task solving by empty container
207 avoidance. The second group was tested first with the "bent tube protocol" and then with the
208 "cup protocol".

209

210 Cup task

211 Training period

212 For individuals experiencing the cup task as their first experimental condition, we conducted
213 three preliminary training phases before the actual task began. The first training phase consisted
214 of 16 trials. The first four trials of this phase were shaping trials, during which a carrot slice
215 was placed in the middle of the table then covered with a single transparent cup. This cup was
216 covered with an opaque over-cup. As soon as the subject "pointed" correctly (i.e., as soon as
217 he/she positioned his/her head and inserted his/her snout in front of the over-cup), the over-cup
218 and the cup were lifted and the subject was given the carrot slice. The following 12 trials trained
219 the animals to point through the grid at various locations. We placed a carrot in the middle (4
220 trials), on the right-hand (4 trials) or on the left-hand (4 trials) side of the table, and covered this
221 carrot with a transparent cup, followed by an opaque over-cup. We brought the cups closer to
222 the individual to allow him/her to point, and then receive the reward. This familiarized the
223 subjects with the experimental material, as well as with the two new possible response locations
224 – left and right. It should be noted that with only one option on the table, individuals always
225 pointed correctly at the right location (i.e., opposite the over-cup). Subjects were then rewarded.
226 This training phase was completed when the individual positioned him/herself correctly within
227 few seconds and pointed with ease without unwanted behaviours (attempting to walk around
228 the table, repeatedly walking from one side of the table to the other, pointing towards the reward
229 but from a distance instead of directly in front of it, pushing the table, head-butting the table,
230 etc.).

231 In a second training phase, we familiarized the subjects with the fact that they had to
232 choose between two locations, only one of which was baited. We conducted 10 trials in which
233 a carrot was placed either on the left or right-hand side of the table. We ensured that the subject
234 had seen the location of the carrot, which could take between 2 and 5 seconds, then covered it
235 with a transparent cup and placed a similar empty cup – that did not cover a reward – on the
236 opposite side of the table. Both cups were then covered with an opaque over-cup. We

237 simultaneously moved both cups closer to the individual so that he/she could make a choice.
238 Each side was baited in a pseudo-randomized way in order to place the carrot on each side five
239 times. The individual made his/her choice by pointing correctly through the grid and was
240 rewarded only if he/she had pointed at the right location. The cups and over-cups were then
241 removed from the table before starting the next trial. We considered that the individual had
242 completed this training phase after obtaining a 80 % success rate for two consecutive sessions.

243 In the third training phase, we familiarized the subjects with the fact that the cups – and
244 not just the over-cups – could either be transparent or opaque. This training proceeded as before,
245 except that the transparent smaller cups were replaced by opaque ones. Once the individual had
246 obtained an 80% success rate for two consecutive sessions, we considered that he/she had
247 completed this phase. The individual results for each of the training phases are shown in Table
248 1.

249 For individuals that had already been tested in the tube task first, we conducted just one
250 session of the first phase of training (Training 4), then proceeded as previously described for
251 the two last phases of training (which were then labelled training phase 5 and 6).

252

253 Testing period

254 For the testing period, the baiting protocol was identical to the last phase of training (i.e., a
255 carrot slice covered with transparent or opaque cups, which were covered with opaque over-
256 cups), except that a sheet of cardboard was used to prevent the individual from observing the
257 baiting. The sheet of cardboard was then removed, and the experimenter simultaneously lifted
258 both of the over-cups and showed the cups to the subjects under four different conditions, each
259 providing different degrees of information (Figure 2): (1) Complete information (“both”): the
260 two cups were transparent and the food was clearly visible; (2) Direct information (“baited”):
261 only the baited cup was transparent and the food was clearly visible; (3) Inferential information
262 (“inferential”): the unbaited cup was transparent, so the subject had to infer that the food was
263 in the opaque cup; (4) No information (“control”): both cups were opaque and the subject was
264 not provided with any visual information about the location of the reward. As mentioned in the
265 introduction, the inference condition does not necessarily imply inferential reasoning as it can
266 be solved with lower cognitive level strategies. In total, the subjects participated in 80 trials. 10
267 sessions of 8 trials each were conducted, presenting each of the 4 conditions twice (with the
268 carrot placed equally often on each side) within a session. Cups could not be consecutively
269 baited more than twice on the same side of the table.

270

271 Bent tube task

272 Training period

273 For individuals experiencing the tub task as their first experimental condition, we conducted
274 three preliminary training phases before the actual experiment began. The first training phase
275 consisted of 16 trials similar to those described for the cup task above. The first four trials of
276 this phase were shaping trials, during which we first showed the carrot to the subject, placed
277 the carrot in the middle of the table, then moved a straight opaque tube towards the reward and
278 inserted the carrot inside the tube in the same location. The tube was then covered with an
279 opaque over-cup, and as soon as the individual pointed correctly, the over-cup was lifted and
280 the carrot slice was taken from the tube and given to the subject. In the following 12 trials, we
281 placed a carrot slice inside the straight tube, located in the middle (4 trials), on the right (4 trials)
282 or on the left (4 trials), covered the tube with an opaque over-cup, then let the subject point. It
283 should be noted that with only one option on the table (i.e. centre, left or right), individuals
284 always pointed correctly at the right location (facing the over-cup). They were then rewarded.
285 This phase was complete when the individual positioned him/herself correctly within few
286 seconds, and pointed with ease without unwanted behaviours.

287 In the second training phase, we conducted 10 trials in which a carrot slice was placed on
288 either the left or the right side of the table. We made sure that the subject had observed the
289 location of the reward, which could take between 2 to 5 seconds; we then put it inside the
290 straight tube and placed another (empty) straight tube on the opposite side of the table. We
291 brought them closer to the individual so that he/she could make a choice. Each side was baited
292 in a pseudo-randomized way in order to place the carrot on each side of the table five times.
293 Once the individual reached an 80 % success rate for two consecutive sessions, we considered
294 that he/she had completed this training phase.

295 In the third training phase, we familiarized subjects with the fact that the tube could be
296 bent. We conducted 12 trials in which we first placed both the straight and bent tubes on the
297 table. The straight tube was placed so that the inside was not visible. The bent tube was placed
298 so that only one part of the tube was visible. We then inserted the carrot slice into one of the
299 tubes – in the case of the bent tube, it was placed in the part that was not directly visible. Both
300 tubes were then rotated in a single movement, so that whatever tube the reward was hidden
301 in, it became visible to the individual. This exposed the individual to the notion that the food
302 could be there without being visible, but it did not train the individual to make a choice by
303 inference, since the food was visible at the time of choice. Each type of tube was baited equally
304 often in a semi-randomised manner to ensure that the carrot was available with the same

305 frequency on the right and left sides of the table. Once the individual had reached an 80 %
306 success rate for two consecutive sessions, we considered that he/she had completed this phase.
307 The individual results for each of these training phases are given in Table 1.

308 For individuals who had already been tested in the cup task first, we conducted only one
309 session of the first phase of training (now labelled training 4), then proceeded as previously
310 described for the two last phases of training (which were then labelled training phase 5 and 6).

311
312

313 Testing period

314 For the testing period, the baiting protocol was identical to Phase 3 of training, except that a
315 cardboard screen prevented the subject from observing the baiting. Once the cardboard was
316 lifted, the subject was exposed to both types of tubes under 10 different configurations, offering
317 a total of 5 conditions providing different levels of information (Figure 2). In the Visible
318 condition (configurations 1 to 4), the carrot slice was always placed in a visible part of the tubes.
319 In the first inferential condition (configuration 5), the straight tube was visible and empty, and
320 the carrot was in the bent tube, with no part of the bent tube visible. In the second inferential
321 condition (configuration 6), the straight tube was empty, and the carrot was placed in the
322 invisible part of the bent tube, with the alternative part of the bent tube visible and empty. In
323 the incomplete information condition (configurations 7 and 8), the straight tube was presented
324 horizontally and only the straight part of the bent tube was visible. In the no information
325 condition (configurations 9 and 10), the straight and the bent tubes were presented horizontally
326 and no part was visible (Figure 2). In total, the subjects took part in 120 trials. 12 sessions of
327 10 trials each were conducted with the requirement to present each of the 10 configurations
328 twice (with the carrot placed an equal number of times on each side) within a set of two sessions.
329 Neither side could be baited more than twice in a row. A single experimenter (JD) conducted
330 all phases of this study (training and testing), and all choices could be classified as correct or
331 incorrect without ambiguity. Inter-observer reliability tests were therefore unnecessary.

332
333

333 4) Statistics

334 In a first analysis (GLMM1), we compared the performances of the two species in both tasks.
335 We sought to identify which of the following variables best explained the number of correct
336 responses: Species (goat, sheep), Task type (cup, bent tube), Task order (cup-tube, when the
337 cup task was the first task experienced, or, tube-cup when the tube task was the first task

338 experienced), Side (left, right), Experience within a task (first half or second half of a given
339 task), and Condition (both, baited, inferential, control). For Condition, the 10 configurations of
340 the tube task were grouped as follows to match the conditions of the cup task: Configurations
341 1 and 2 were grouped as an equivalent of the “both” condition of the cup task. Configurations
342 3 and 4 were grouped in the “baited” conditions. Configurations 5 and 6 were grouped in the
343 “inferential” condition. Configurations 7-10 were grouped in the “control” condition. The full
344 model also took into account the possible interactions between the factors Condition x Order of
345 testing (to detect a potential increase in performance in the inferential condition in the second
346 task compared to the first) and between the factors Condition x Species (to detect a potential
347 difference between species in the inferential condition). We added subjects as a random factor.
348 We first ran a GLMM model with a logit link function and binomial distribution (package *lme4*,
349 in *R*). This model was compared with a null model. The post-hoc analysis of main effects was
350 conducted using package *LSmean* with a Bonferroni correction (Lenth 2016).

351 A second analysis (GLMM2) focused in more detail on the responses of the two species
352 in the tube task in order to investigate the type of cognitive strategy (deductive/low level)
353 involved. We then performed a similar analysis, with the full model including the variables
354 Species, Task order, Experience within a task, and Configuration. The full model also took into
355 account the possible interactions between the factors Configuration x Task order (to detect a
356 potential increase in performance in the inferential configurations according to whether the tube
357 task was the second protocol experienced by subjects or the first) and between the factors
358 Configuration x Species (to detect a potential difference between species in the inferential
359 configurations).

360 A third analysis (GLMM3) assessed the effect of the conditions (inferential 5, inferential
361 6, incomplete information, no information) on the variable “Preference for the bent tube”, with
362 Subjects as random factors. We predicted that if success in inferential conditions was due to a
363 preference for the bent tube, there should be no differences in the choice of the bent tube
364 between the two inferential conditions and the two control conditions.

365 Finally, we evaluated individual response in each task based on a success criterion that
366 would be significantly different from random choices, as estimated by a binomial two-tailed
367 test (see Jelbert et al. 2015; Danel et al. 2021 for similar procedure) in the inferential conditions,
368 but not in the control conditions (where individuals have no way to predict the correct location).
369 For the result to be statistically significant, individuals had to display success rates of 75 % or
370 more in the cup task (15 tests out of 20 to reach significance, with $p < 0.05$). In the tube task, a
371 success rate of 83 % or more was required to reach this criterion (10 out of 12 tests). We can

372 expect the following performances in the tube task: (1) Reasoning by exclusion: if subjects
 373 solve the task by making inferences, they should perform well in inferential configurations 5
 374 and 6 (more than 83 % correct choices), and they should choose at random in incomplete or no
 375 information configurations (configurations 7 to 10). (2) Avoiding empty tubes: configurations
 376 5 and 6 are both inferential configurations but, unlike configuration 6, configuration 5 is
 377 solvable with a strategy based on avoiding empty tubes: subjects using this strategy should
 378 therefore perform significantly above chance for configuration 5, but at random for
 379 configuration 6; they should also show 100 % success in configuration 7, 0% in configuration
 380 8, and 50 % in configurations 9 and 10. (3) Preference for bent tubes: subjects may show a
 381 preference for the bent tubes, which contained rewards more often than the straight tubes (by
 382 experimental design, six times out of ten): in this case, subjects should succeed in both
 383 inferential configurations, never succeed at configurations 7 and 9, and always succeed in
 384 configurations 8 and 10.

385

386 **Results**

387 The first analysis (GLMM1) compared the performances of each species in both tasks and under
 388 the different conditions. The results of the GLMM indicated that the number of correct
 389 responses was affected by the variables Condition ($F_{(3)} = 261, p < 0.001$), Task order ($F_{(1)} = 9.5,$
 390 $p = 0.002$), and the interaction Condition x Task order ($F_{(3)} = 14, p < 0.001$). The variables
 391 Species ($F_{(1)} = 0.3, p = 0.6$), Experience within the task ($F_{(1)} = 0.4, p = 0.5$), Type of task ($F_{(1)}$
 392 $= 0.8, p = 0.4$), Side ($F_{(1)} = 0.3, p = 0.6$), and the interaction Species x Condition ($F_{(3)} = 0.8, p$
 393 $= 0.5$) did not significantly affect the number of correct responses. This model was better than
 394 the null model (likelihood ratio test: GLMM1 model vs null model: $Df = 14, \text{Chi-square} =$
 395 $1478.7, p < 0.001$). Performances were better in the second task to which subjects were exposed
 396 (cup-tube or tube-cup) compared to their first task ($M_{(\text{first task})} = 74\%$ of success, $M_{(\text{second task})} =$
 397 79%) (see Online resource, Supplementary Table 1 for pairwise comparisons). The subjects
 398 performed better in the “both” ($M_{(\text{both})} = 96\%$), “baited” ($M_{(\text{baited})} = 97\%$), and “inferential”
 399 ($M_{(\text{inferential})} = 89\%$) conditions compared to the control conditions ($M_{(\text{control})} = 42\%$) (Online
 400 resource Supplementary Table 1). For the Task order x Condition interaction, a pairwise
 401 comparison with a Bonferroni correction indicated that in the first task (cup-tube or tube-cup),
 402 performances in the inferential condition were high ($M_{(\text{inferential})} = 81\%$), but were significantly
 403 lower than in the “baited” and “both” conditions ($M_{(\text{both})} = 95\%$; $M_{(\text{baited})} = 95\%$) (Online
 404 resource Supplementary Table 1). Performances in the “inferential” condition in the first task
 405 were also significantly lower than in the second task ($M_{(\text{inferential-task 1})} = 81\%$; $M_{(\text{inferential-task 2})} =$

406 96 %) (Supplementary Table 1). In the second task (cup-tube or tube-cup), performances
 407 between inferential and baited or both conditions did not differ significantly from each other
 408 and were close to 100 % of success ($M_{(\text{inferential})} = 96 \%$; $M_{(\text{both})} = 98 \%$; $M_{(\text{baited})} = 99 \%$, Online
 409 resource Supplementary Table 1, Figure 3). At the individual level, note that for their first four
 410 inferential trials four goats and two sheep responded correctly in only half of the trials (in goats:
 411 Eden, Oscar, Raspoutine and Riesling; in sheep: Nanook and Schtroumpf). Except for Oscar,
 412 these individuals were also the worst performers in their first task (either cup or tube task) in
 413 inferential trials (Table 2). Four goats and six sheep responded correctly in 75 % of the trials
 414 (goats: Elvis, Oreo, Petit Oscar and Prisca; sheep: Candy, Domino, Hector, Igor, Neige and
 415 Popeye). Two goats and two sheep responded with 100 % of success (goats: Ella and Satine;
 416 sheep: Bidule and Praline).

417 We then focused on the responses in the tube task, designed to discriminate between low-
 418 and high-level cognitive solutions for the inferential task. If high-level cognitive mechanisms
 419 were involved, responses at the inferential configurations 5 and 6 should be similar from
 420 responses to the “visible configurations” (configurations 1 to 4, where 100% of success was
 421 expected) and different from the control configurations (configurations 7 to 10, where random
 422 choices were expected). We ran the same GLMM as above on the responses in the tube task,
 423 but removed the “Type of task” variable (GLMM2). The full model was better than the null
 424 model (likelihood ratio test: GLMM2 vs null model: $Df = 31$, Chi-square = 1024.7, $p < 0.001$).
 425 The variable Configuration ($F_{(9)} = 42.7$, $p < 0.001$), and the interaction Configuration x Task
 426 order ($F_{(9)} = 2.5$, $p = 0.007$) influenced the number of correct responses. The variables Species
 427 ($F_{(1)} = 0.05$, $p = 0.81$), Experience within the task ($F_{(1)} = 0.19$, $p = 0.65$), Side ($F_{(1)} = 0.0006$, p
 428 $= 0.98$), and the interaction Species x Configuration ($F_{(9)} = 0.49$, $p = 0.88$) did not affect the
 429 number of correct responses significantly.

430 For the Configuration x Task order interaction, a pairwise comparison with a Bonferroni
 431 correction indicated that subjects had a lower success rate during configuration 5 if they
 432 experienced the tube task first, compared to when it was the second task they experienced
 433 (configuration 5, $M_{(\text{tube task first})} = 73 \%$, $M_{(\text{tube task second})} = 98 \%$) (Online resource supplementary
 434 Table 2). If they experienced the tube task first, success rates were also significantly lower for
 435 configuration 5 compared to all configurations where visual information was accessible
 436 ($M_{(\text{configurations 1 to 4})} = 98 \%$, $M_{(\text{configuration 5})} = 73 \%$) (Online resource supplementary Table 2).
 437 However, success rate for configuration 5 was significantly higher than success rates for the
 438 control configurations except for configuration 8 ($M_{(\text{configuration 7})} = 32 \%$, $M_{(\text{configuration 8})} = 54 \%$,
 439 $M_{(\text{configuration 9})} = 41 \%$, $M_{(\text{configuration 10})} = 47 \%$; Supplementary Table 2). When experiencing the

440 tube task as the second task (i.e., after the cup task), the performances of both configurations 5
441 and 6 were similar to the performances in configurations where visual information was given
442 (configurations 1 to 4) and significantly different from all control configurations. For
443 configuration 6, regardless of the order of the task (tube task first or second), success rate was
444 not significantly different from the configurations where visual information was given
445 (configurations 1 to 4) (Figure 4).

446 To examine whether subjects solved the task due to a preference for the bent tubes, we
447 performed a third GLMM analysis (GLMM3) to compare their preference for the bent tube in
448 inferential configurations and control conditions. This model was better than the null model
449 (likelihood ratio test: GLMM3 model vs null model: $Df = 3$, Chi-square = 159.45, $p < 0.001$).
450 The results showed an effect of condition ($F_{(3)} = 43.5$, $p < 0.001$). Choices of the bent tube did
451 not differ between configurations 5 ($M_{(\text{configuration } 5)} = 86\%$) and 6 ($M_{(\text{configuration } 6)} = 88\%$). The
452 bent tube was chosen significantly more often in both inferential configurations compared to
453 the other control conditions ($M_{(\text{incomplete information})} = 62\%$, $M_{(\text{no information})} = 51\%$) (Online resource
454 supplementary Table 3). A significant difference also appeared between the two control
455 conditions, with the bent tube chosen slightly more often in the incomplete information
456 condition than in the no information condition, however the choices of the bent tube remained
457 close to 50 % in these two conditions.

458

459 Individual responses to predictions

460 We summarize the results of each individual in Table 2. The goal was to establish whether some
461 individuals met all the criteria required to fulfil the high-level cognitive resolution of the
462 inference tasks, or if the subjects showed alternative strategies. Most subjects successfully met
463 the 75 % success criteria in the cup task. One goat (Raspoutine) and one sheep (Schtroumpf)
464 did not meet this criterion (with 65% of correct choices in the inferential condition) and another
465 goat (Elvis) succeeded above chance level in the inferential condition but did not chose
466 randomly at the control condition. The cup task was the first task for these three individuals. In
467 the tube task, 6 sheep and 6 goats met the 83 % success criteria, i.e. with success at both
468 inferential configurations. However, 4 of these individuals (3 sheep, 1 goat) had to be removed
469 because at least one of their controls differed significantly from a 50 % choice, as assessed with
470 a binomial test. A total of 3 sheep and 5 goats fulfilled all criteria in the tube task.

471 It must also be noted that mean success levels of performances at control conditions were
472 significantly below 50% (individuals made the wrong choice in 65% of the control trials, exact
473 binomial test, $p < 0.001$). In the cup condition, only 4 out of 20 individuals reached success

474 rates of 50% or more at the control condition, but most of the individuals' choices remain close
475 to 50% (mean = 42%). In the tube task, for control configuration 7, only 1 in 20 individuals
476 reached a 50% success rate, and the mean percentage of success was very low (30%). If we
477 combine this information with the observation that subjects were often below 50% in
478 configuration 9 (13 subjects), and above 50% in configuration 8 (15 subjects), there might be a
479 possible explanation for this response pattern, i.e. a tendency to avoid tubes that were presented
480 sideways (i.e., when the side of the tube is facing them rather than the opening). This potential
481 bias is not necessarily present in all individuals, but means we should only retain those
482 individuals whose performance in the control configurations did not differ significantly from
483 chance, as we did above.

484 We also detected an effect of the task order, as inferential performances were better in the
485 second task compared to the first one (regardless of the type of task they were exposed to first,
486 cup or tube). To limit the risk of a learning bias, we refined the criteria to only include
487 individuals who experienced the tube task first. In this case, one sheep (Hector) and one goat
488 (Oreo) still achieved 83 % success in both inferential configurations, and showed responses at
489 control configurations that did not differ significantly from random. Note that 8 individuals (5
490 sheep, 3 goats) tested with the tube task first met the 83 % criteria in configuration 6 (the high-
491 level cognition condition), but only 2 of them meet the criteria in both configurations 5 and 6.

492 When tested with the tube as a second task, all individuals of each species met the 83 %
493 criteria in both conditions 5 and 6. Among them, 2 sheep and 4 goats also had correct controls,
494 i.e. not different from random.

495

496

497 **Discussion**

498 Our results show that both sheep and goats solved the inferential conditions of the cup and tube
499 tasks with high percentages of success, and performed slightly better in their second task. Their
500 performances in the tube task cannot be explained by alternative strategies such as avoiding the
501 empty cup or preferring the bent tube. If we apply a strict set of criteria requiring unbiased
502 responses at controls and success in both inferential conditions in the tube task when
503 experienced first, we find that two individuals, a goat (Oreo) and a sheep (Hector), fulfil these
504 criteria. This suggests that sheep and goats are able to make inferences based on deductive
505 reasoning.

506 The results for sheep and goats in our study are generally better than those reported by
507 Nawroth et al. (2014). In their study, both species were tested in a cup task and only two goats

508 out of 11 (and none of the six sheep) solved the inferential condition. In comparison, all sheep
509 and goats solved the inferential condition of the cup task in our study (whether tested first or
510 second), and there were no differences between species. None of the sheep and goats in our
511 study had previous experience of cognitive testing. This was true of the sheep in Nawroth's
512 study but not for the goats. However, our sheep and goats were very familiar with humans and
513 could choose to approach or avoid park visitors. They also had good control of their daily
514 activities, with freedom of movement from 11 am to 5 pm daily. In cognitive terms, these
515 enriched welfare conditions can be expected to be more stimulating than standard husbandry
516 conditions (Franks 2018). The goats studied by Nawroth et al. (2014) were housed in a research
517 centre with various enrichments and had previous experience with cognitive tests, but the sheep
518 were housed in a more conventional breeding facility, where animals may have had limited
519 control over their environment. Differences between environmental conditions – enriched
520 versus standard – could explain the differences reported between sheep and goats in Nawroth's
521 experiments, and more generally the poorer performances of their sheep compared to ours. If
522 this hypothesis is confirmed, it could have strong implications for the management of well-
523 being in breeding facilities, where the housing conditions for livestock should be enriched to
524 provide opportunities for animals to express a wider range of cognitive skills.

525 One of the main problems with inferential tasks is ensuring that subjects do not use
526 alternative strategies to solve the task (Paukner et al. 2009; Jelbert et al. 2015; Völter and Call
527 2017). In the standard cup task, individuals can learn to avoid the transparent cup. Using the
528 tube task is a way to control for this possibility, as both tubes are visibly empty in the inferential
529 conditions. Avoiding the empty tube is not, therefore, a strategy that can explain success in our
530 tube task. An alternative explanation could be that individuals learned the following rule:
531 “choose the bent tube if it is next to an empty straight tube”. This alternative explanation would
532 require individuals to learn this rule in the course of the experiment and is unlikely, as the
533 GLMM2 detected no effect of testing experience within the tube task, indicating that there were
534 no significant differences in performances between the first half and the second half of the trials.
535 However, while 8 of 10 individuals tested with the tube task first succeeded in inferential
536 configuration 6, only 2 of these successful individuals also performed above chance on
537 inferential configuration 5. It is unclear why configuration 5 proved difficult for them. This
538 effect was not detected when the tube task was the second task to be experienced. Poor
539 performance on this inferential condition cannot be explained by a particular bias. However,
540 we cannot exclude that individuals who performed accurately in condition 6 and slightly less
541 accurately in condition 5 may have experienced minor biases that were not expressed

542 consistently throughout the study. Nevertheless, considering responses to both inferential
543 configurations and control configurations did not identify a particular strategy of response that
544 would preclude reasoning. In New Caledonian crows (*Corvus moneduloides*), 7 out of 8 birds
545 chose the correct location above chance level in configuration 5, but only 3 birds out of 8 did
546 so in configuration 6 (Jelbert et al. 2015). Most bird performances could be explained by mixed
547 strategies, some of which included the avoidance of empty tubes, but two birds showed results
548 that were consistent with high-level inferential reasoning. These performances are similar to
549 those reported for sheep and goats. Note that in the initial versions of the tube task, subjects
550 were allowed to check the contents of the tube before picking them up (Schloegl et al. 2009;
551 Völter and Call 2017). Crucially, if the individuals first visually check the empty straight tube,
552 subjects using inferential reasoning should infer the correct location of the food in the
553 alternative tube and choose this second location without checking inside. This occurred in
554 chimpanzees (*Pan troglodytes*) but was infrequent, as they still checked about 60 % of the time
555 (Call and Carpenter 2001). Corvids (*Corvus corvax* and *Corvus monedula*) checked between
556 80 and 90 % of the time, and psittacines such as keas (*Nestor notabilis*) checked 95 % of the
557 time (Schloegl et al. 2009; Schloegl 2011). As applied by Jelbert et al. (2015), our procedure
558 used bent tubes to control for the “avoiding the empty tube” strategy. Subjects could see the
559 part of the tube they were facing, but we did not allow them to look inside at the invisible parts
560 of the tube. Indeed, it would have been difficult to let the sheep and goats move around the
561 tubes, and given their wide vision field angle, it would have been impossible to detect if or
562 when they had indeed looked inside and checked.

563 According to Nawroth et al. (2014), the cognitive differences between sheep and goats
564 could be explained by the adaptive specialization hypothesis (de Kort and Clayton 2006), which
565 posits that species-specific adaptations to a particular feeding ecology may lead to the evolution
566 of different cognitive abilities (Paukner et al. 2009; Schloegl et al. 2009). A study by Hosoi and
567 colleagues (Hosoi et al. 1995) also showed that goats avoid high-fibre food if they have the
568 choice to opt for lower-fibre food, which is not the case for sheep. Thus, goats have more dietary
569 flexibility than sheep, and this may be underpinned by higher cognitive flexibility in goats.
570 Goats also adjusted faster to a modified detour task than sheep did, which provides an additional
571 argument in favour of a greater cognitive flexibility in goats compared to sheep (Raoult et al.
572 2021). Our results do not support this hypothesis because both species performed equally well,
573 even in a more controlled inferential task, the tube task. Contrary to expectations, we did not
574 find that goats performed better than sheep. One might also assume that domestication has
575 levelled the cognitive abilities of both species towards lower performances. We know that the

576 reduction in brain size reported in several domestic species has also occurred in sheep (Ebinger
577 1974; Price 2002; Balcarcel et al. 2021). However, brain size reduction may not have an impact
578 on cognitive abilities such as inferential reasoning. In fact, some authors suggest that the limbic
579 system has suffered the most drastic reduction (Balcarcel et al. 2021), which would affect
580 emotion regulation more than high-level executive functions. However, performances in our
581 tests were quite good and were comparable to many species tested. Pigs also experienced brain
582 size reduction (Ebinger 1974) and nevertheless often demonstrate advanced cognitive skills in
583 both the physical and social domain (Held et al. 2002; Mendl et al. 2010). Additional studies
584 are needed to extend our knowledge of how domestication affects cognitive and emotional skills
585 in farm animals.

586 It should be noted that small ruminants are still often maintained with access to outdoor
587 areas and fields with many conspecifics; i.e., they have a fairly enriched environment. Intensive
588 breeding in environmentally poor facilities is a relatively recent practice compared to 10,000
589 years of extensive breeding. The cognitive skills of sheep and goats may not therefore be very
590 different from those of their feral and wild counterparts (e.g., mouflon: *Ovis aries musimon*;
591 ibex: *Capra aegagrus*). Further studies should assess how far they may have diverged in other
592 aspects than reasoning abilities. In addition, we might expect “breed” effects in the expression
593 of cognitive skills, with breeds that have been subject to the highest level of selection by humans
594 being the less likely to perform well. Our study groups were composed of several breeds, some
595 hardy (i.e. regional races, able to breed without human help) and some not (races more recently
596 selected to increase the production of meat or milk). Although this combination gave us a good
597 overview of the general abilities of sheep and goats, it did not allow us to test for breed effects.
598 It will be necessary to evaluate the effect of strong *versus* low genetic selection on the
599 performances of small ruminants. Indeed, strong artificial selection combined with poor welfare
600 conditions could affect performances.

601 An enriched environment characterized by physical and social complexity is conducive
602 to a higher degree of behavioural diversity in livestock (van de Weerd and Day 2009). These
603 solicitations and enrichments could induce positive emotional states in farm animals, improving
604 their well-being (Dantzer 2002; Greiveldinger et al. 2007; Manteuffel et al. 2009; Boissy and
605 Erhard 2014). Cognitive challenges, whether natural or artificial, combined with rewards and
606 positive reinforcement, also elicit positive emotional states (Dantzer 2002; Greiveldinger et al.
607 2007). Since small ruminants such as sheep and goats can successfully use inferential reasoning,
608 we should strive to enrich their environment in a manner that will give them the possibility to
609 express their cognitive skills.

610

611 Funding and competing interest

612 The authors did not receive financial support from any organization for the submitted work.

613 The authors have no competing interests to declare that are relevant to the content of this article.

614

615

616

617 Bibliography

618 Aust U, Range F, Steurer M, Huber L (2008) Inferential reasoning by exclusion in pigeons,
619 dogs, and humans. *Anim Cogn* 11:587–597. <https://doi.org/10.1007/s10071-008-0149-0>

620 Balcarcel AM, Veitschegger K, Clauss M, Sánchez-Villagra MR (2021) Intensive human
621 contact correlates with smaller brains: differential brain size reduction in cattle types.

622 *Proc R Soc B Biol Sci* 288:20210813. <https://doi.org/10.1098/rspb.2021.0813>

623 Boissy A, Erhard HW (2014) How Studying Interactions Between Animal Emotions,
624 Cognition, and Personality Can Contribute to Improve Farm Animal Welfare. In:

625 *Genetics and the Behavior of Domestic Animals*. Elsevier, pp 81–113

626 Briefer EF, Haque S, Baciadonna L, McElligott AG (2014) Goats excel at learning and
627 remembering a highly novel cognitive task. *Front Zool* 11:20.

628 <https://doi.org/10.1186/1742-9994-11-20>

629 Call J (2004) Inferences about the location of food in the great apes (*Pan paniscus*, *Pan*
630 *troglodytes*, *Gorilla gorilla*, and *Pongo pygmaeus*). *J Comp Psychol* 118:232–241.

631 <https://doi.org/10.1037/0735-7036.118.2.232>

632 Call J (2006) Inferences by exclusion in the great apes: The effect of age and species. *Anim*
633 *Cogn* 9:393–403. <https://doi.org/10.1007/s10071-006-0037-4>

634 Call J, Carpenter M (2001) Do apes and children know what they have seen? *Anim Cogn*
635 3:207–220. <https://doi.org/10.1007/s100710100078>

636 Danel S, Chiffard-Carricaburu J, Bonadonna F, Nesterova AP (2021) Exclusion in the field:
637 wild brown skuas find hidden food in the absence of visual information. *Anim Cogn*

638 24:867–876. <https://doi.org/10.1007/s10071-021-01486-4>

639 Dantzer R (2002) Can farm animal welfare be understood without taking into account the
640 issues of emotion and cognition? *J Anim Sci* 80:E1–E9.

641 <https://doi.org/10.2527/animalsci2002.0021881200800ES10002x>

642 de Kort SR, Clayton NS (2006) An evolutionary perspective on caching by corvids. *Proc Biol*
643 *Sci* 273:417–423. <https://doi.org/10.1098/rspb.2005.3350>

- 644 De Petrillo F, Rosati AG (2020) Logical inferences from visual and auditory information in
645 ruffed lemurs and sifakas. *Anim Behav* 164:193–204.
646 <https://doi.org/10.1016/j.anbehav.2020.03.010>
- 647 Ebinger P (1974) A cytoarchitectonic volumetric comparison of the area gigantopyramidalis
648 in wild and domestic sheep. *Anat Embryol (Berl)* 147:167–175
- 649 Franks B (2018) Cognition as a cause, consequence, and component of welfare. In: Mench J
650 a. (ed) *Advances in Agricultural Animal Welfare*. Elsevier, pp 3–24
- 651 Greaves LA, Wedderburn ME (1995) Comparison of the behaviour of goats and sheep on an
652 eroded hill pasture. *Appl Anim Behav Sci* 42:207–216
- 653 Greiveldinger L, Veissier I, Boissy A (2007) Emotional experience in sheep: Predictability of
654 a sudden event lowers subsequent emotional responses. *Physiol Behav* 92:675–683.
655 <https://doi.org/10.1016/j.physbeh.2007.05.012>
- 656 Held S, Mendl M, Devereux C, Byrne RW (2002) Foraging pigs alter their behaviour in
657 response to exploitation. *Anim Behav* 64:157–165.
658 <https://doi.org/10.1006/anbe.2002.3044>
- 659 Hill A, Collier-Baker E, Suddendorf T (2011) Inferential Reasoning by Exclusion in Great
660 Apes, Lesser Apes, and Spider Monkeys. *J Comp Psychol* 125:91–103.
661 <https://doi.org/10.1037/a0020867>
- 662 Hosoi E, Swift DM, Rittenhouse LR, Richards RW (1995) Comparative foraging strategies of
663 sheep and goats in a T-maze apparatus. *Appl Anim Behav Sci* 44:37–45.
664 [https://doi.org/10.1016/0168-1591\(95\)00572-A](https://doi.org/10.1016/0168-1591(95)00572-A)
- 665 Houpt KA (2018) *Domestic animal behavior for veterinarians and animal scientists*. John
666 Wiley & Sons, Ltd.
- 667 Hunter DS, Hazel SJ, Kind KL, et al (2015) Do I turn left or right? Effects of sex, age,
668 experience and exit route on maze test performance in sheep. *Physiol Behav* 139:244–
669 253. <https://doi.org/10.1016/j.physbeh.2014.11.037>
- 670 Jelbert SA, Taylor AH, Gray RD (2015) Reasoning by exclusion in New Caledonian crows
671 (*Corvus moneduloides*) cannot be explained by avoidance of empty containers. *J Comp*
672 *Psychol* 129:283–290. <https://doi.org/10.1037/a0039313>
- 673 Kaminski J, Call J, Tomasello M (2006) Goats' behaviour in a competitive food paradigm:
674 Evidence for perspective taking? *Behaviour* 143:1341–1356.
675 <https://doi.org/10.1163/156853906778987542>
- 676 Larson G, Fuller DQ (2014) The evolution of animal domestication. *Annu Rev Ecol Evol Syst*
677 45:115–136. <https://doi.org/10.1146/annurev-ecolsys-110512-135813>

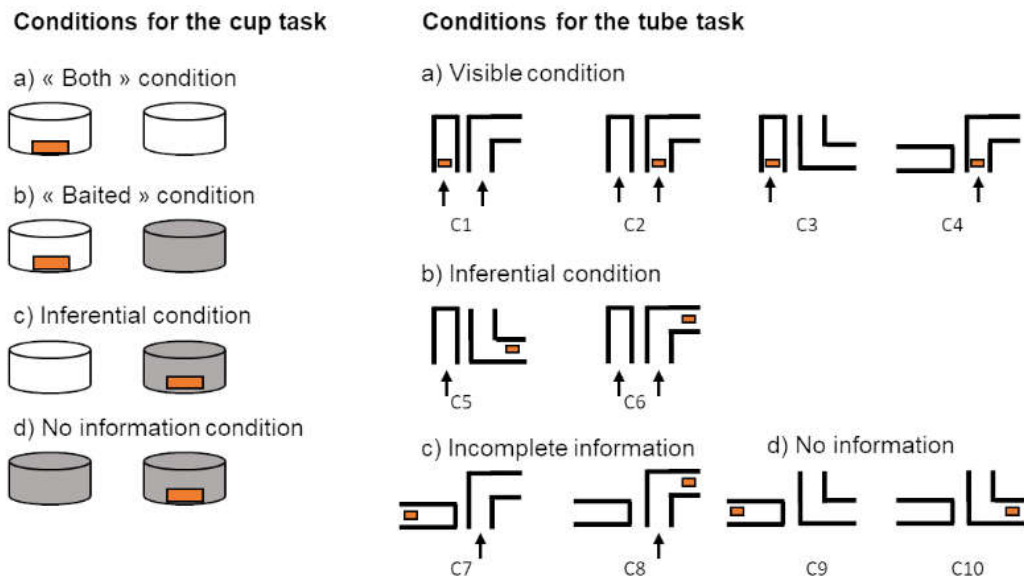
- 678 Lenth R V. (2016) Least-Squares Means: The R Package lsmeans. *J Stat Softw* 69:.
679 <https://doi.org/10.18637/jss.v069.i01>
- 680 Lyons DM, Price EO, Moberg GP (1993) Social grouping tendencies and separation-induced
681 distress in juvenile sheep and goats. *Dev Psychobiol* 26:251–259.
682 <https://doi.org/10.1002/dev.420260503>
- 683 Manteuffel G, Langbein J, Puppe B (2009) From operant learning to cognitive enrichment in
684 farm animal housing: Bases and applicability. *Anim Welf* 18:87–95
- 685 Mendl M, Held S, Byrne RW (2010) Pig cognition. *Curr Biol* 20:796–798.
686 <https://doi.org/10.1016/j.cub.2010.07.018>
- 687 Miranda-de la Lama GC, Mattiello S (2010) The importance of social behaviour for goat
688 welfare in livestock farming. *Small Rumin Res* 90:1–10.
689 <https://doi.org/10.1016/j.smallrumres.2010.01.006>
- 690 Nawroth C (2017) Invited review: Socio-cognitive capacities of goats and their impact on
691 human–animal interactions. *Small Rumin Res* 150:70–75.
692 <https://doi.org/10.1016/j.smallrumres.2017.03.005>
- 693 Nawroth C, Langbein J, Coulon M, et al (2019) Farm Animal Cognition—Linking Behavior,
694 Welfare and Ethics. *Front Vet Sci* 6:.. <https://doi.org/10.3389/fvets.2019.00024>
- 695 Nawroth C, Von Borell E, Langbein J (2014) Exclusion performance in dwarf goats (*Capra*
696 *aegagrus hircus*) and sheep (*Ovis orientalis aries*). *PLoS One* 9:1–8.
697 <https://doi.org/10.1371/journal.pone.0093534>
- 698 O’Hara M, Schwing R, Federspiel I, et al (2016) Reasoning by exclusion in the kea (*Nestor*
699 *notabilis*). *Anim Cogn* 19:965–975. <https://doi.org/10.1007/s10071-016-0998-x>
- 700 Paukner A, Huntsberry ME, Suomi SJ (2009) Tufted Capuchin Monkeys (*Cebus apella*)
701 Spontaneously Use Visual but Not Acoustic Information to Find Hidden Food Items. *J*
702 *Comp Psychol* 123:26–33. <https://doi.org/10.1037/a0013128>
- 703 Penn DC, Povinelli DJ (2007) On the lack of evidence that non-human animals possess
704 anything remotely resembling a “theory of mind.” *Philos Trans R Soc B Biol Sci*
705 362:731–744. <https://doi.org/10.1098/rstb.2006.2023>
- 706 Petit O, Dufour V, Herrenschildt M, et al (2015) Inferences about food location in three
707 cercopithecine species: an insight into the socioecological cognition of primates. *Anim*
708 *Cogn* 18:821–830. <https://doi.org/10.1007/s10071-015-0848-2>
- 709 Premack D (1995) Cause/induced motion: Intention/spontaneous motion. In: Changeux J,
710 Chavaillon J (eds) *Origins of the human brain*. Clarendon Press/Oxford University Press,
711 pp 286–309

- 712 Price EO (2002) *Animal Domestication and Behavior*. Cabi publishing, Wallingford
- 713 Raoult CMC, Osthaus B, Hildebrand ACG, et al (2021) Goats show higher behavioural
714 flexibility than sheep in a spatial detour task. *R Soc Open Sci* 8:rsos.201627.
715 <https://doi.org/10.1098/rsos.201627>
- 716 Sabbatini G, Visalberghi E (2008) Inferences about the location of food in capuchin monkeys
717 (*Cebus apella*) in two sensory modalities. *J Comp Psychol* 122:156–66.
718 <https://doi.org/10.1037/0735-7036.122.2.156>
- 719 Schloegl C (2011) What you see is what you get—Reloaded: Can jackdaws (*Corvus*
720 *monedula*) find hidden food through exclusion? *J Comp Psychol* 125:162–174.
721 <https://doi.org/10.1037/a0023045>
- 722 Schloegl C, Dierks A, Gajdon GK, et al (2009) What you see is what you get? Exclusion
723 performances in ravens and keas. *PLoS One* 4:.
724 <https://doi.org/10.1371/journal.pone.0006368>
- 725 Schmitt V, Fischer J (2009) Inferential Reasoning and Modality Dependent Discrimination
726 Learning in Olive Baboons (*Papio hamadryas anubis*). *J Comp Psychol* 123:316–325.
727 <https://doi.org/10.1037/a0016218>
- 728 Schusterman RJ, Reichmuth CJ, Kastak D (2000) How animals classify friends and foes. *Curr*
729 *Dir Psychol Sci* 9:1–6. <https://doi.org/10.1111/1467-8721.00047>
- 730 Shaw RC, Plotnik JM, Clayton NS (2013) Exclusion in corvids: The performance of food-
731 caching eurasian jays (*Garrulus glandarius*). *J Comp Psychol* 127:428–435.
732 <https://doi.org/10.1037/a0032010>
- 733 Shettleworth SJ (2009) *Cognition, evolution, and behavior*. Oxford University Press
- 734 van de Weerd HA, Day JE (2009) A review of environmental enrichment for pigs housed in
735 intensive housing systems. *Appl Anim Behav Sci* 116:1–20
- 736 Veissier I, Stefanova I (1993) Learning to suckle from an artificial teat within groups of
737 lambs: Influence of a knowledgeable partner. *Behav Processes* 30:75–82.
738 [https://doi.org/10.1016/0376-6357\(93\)90013-H](https://doi.org/10.1016/0376-6357(93)90013-H)
- 739 Völter CJ, Call J (2017) Causal and inferential reasoning in animals. In: *APA handbook of*
740 *comparative psychology: Perception, learning, and cognition*. American Psychological
741 Association, Washington, pp 643–671
- 742 Watanabe S, Huber L (2006) Animal logics: Decisions in the absence of human language.
743 *Anim Cogn* 9:235–245. <https://doi.org/10.1007/s10071-006-0043-6>
744
745



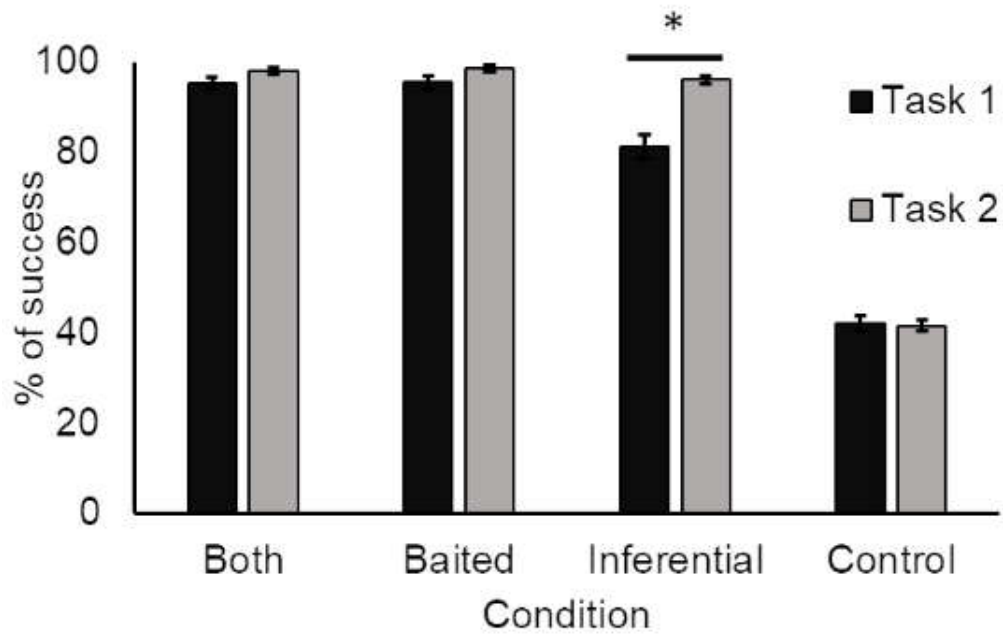
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760

Fig. 1: a. General set-up for both tasks (here, the tube task) with a sheep facing a grid, fixed on the table. To make a choice, individuals must insert their snout in front of the chosen location. **b.** Material used for the cup task, with small opaque (front left) or transparent (front right) cups, and larger opaque over-cups. **c.** Material used for the tube task, with a straight tube and a bent tube.



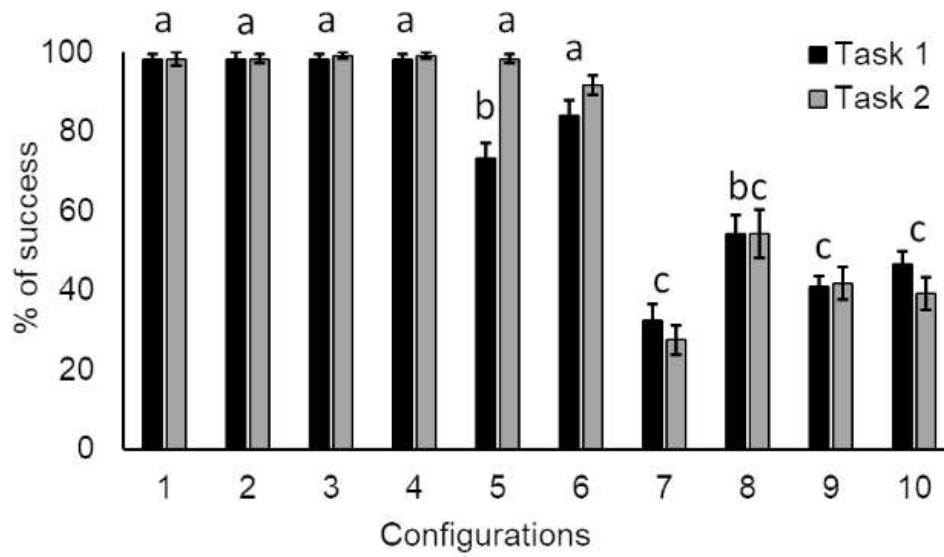
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782

Fig. 2: Illustration of the conditions for each task. In the cup task, when the carrot slice (orange square) is covered with an opaque cup (in grey), it is no longer visible. In the inferential condition c), the task can be solved by making the correct inference (if the carrot is not in the transparent cup then it must be in the opaque cup). It can also be solved by learning to avoid the empty cup. In the tube task, the visible condition consists of four configurations (from C1-C4) where the carrot is always visible. Percentage of success should be close to 100% in these configurations. The inferential conditions comprise the configurations C5 and C6. Configuration 5 can be solved by inferring that if the straight tube is empty, then the carrot should be in the bent tube. This configuration can also be solved by learning to avoid the straight empty tube. Compared to C5, C6 cannot be solved by using this “avoidance of empty locations” strategy. Indeed, both tubes are visibly empty. Thus succeeding at C6 is a better indicator of inferential reasoning than succeeding at C5. The incomplete information condition consists of two configurations (C7, C8) where the information provided (no food in the visible part of the bent tube) cannot help to solve the task. In the no information condition, the two configurations (C9, C10) do not provide any information on the location of the carrot.



783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795

Fig. 3: Percentage of success (+/- SE) according to conditions in the first analysis (GLMM1) which combines results for the cup task and for the tube task. This figure illustrates the effect of the interaction between task order and condition. Subjects perform less accurately on the inferential condition in their first task than in their second, regardless of the task order (cup-tube or tube-cup: half of the individuals started with the cup task, and half started with the tube task). Only the most relevant contrasts are represented (see Supplementary Table 1 for the list of contrasts).



796

797

798

799

800

801

802

803

804

Fig. 4: Percentage of success (+/- SE) according to configurations in the tube task alone (GLMM2). This figure illustrates the effect of the interaction between Configuration and Task order. Performance in configuration 5 is poorer than in configuration 6 when the tube task is the first task to which the subjects are exposed. Configurations with no letters in common differ significantly at $p < 0.05$ (see supplementary Table 2 for the list of contrasts).