

Est.  
1841

YORK  
ST JOHN  
UNIVERSITY

Tecwyn, Emma ORCID logoORCID:  
<https://orcid.org/0000-0002-2343-2282> and Buchsbaum, Daphna  
What factors really influence domestic dogs' (Canis familiaris)  
search for an item dropped down a diagonal tube? The tubes task  
revisited. PsyArXiv.

Downloaded from: <https://ray.yorks.ac.uk/id/eprint/7494/>

The version presented here may differ from the published version or version of record. If  
you intend to cite from the work you are advised to consult the publisher's version:  
<https://doi.org/10.31234/osf.io/erz kf>

Research at York St John (RaY) is an institutional repository. It supports the principles of  
open access by making the research outputs of the University available in digital form.  
Copyright of the items stored in RaY reside with the authors and/or other copyright  
owners. Users may access full text items free of charge, and may download a copy for  
private study or non-commercial research. For further reuse terms, see licence terms  
governing individual outputs. [Institutional Repository Policy Statement](#)

# RaY

Research at the University of York St John

For more information please contact RaY at [ray@yorks.ac.uk](mailto:ray@yorks.ac.uk)

1 ACCEPTED FOR PUBLICATION IN *JOURNAL OF COMPARATIVE PSYCHOLOGY*

2

3

4

5 What factors really influence domestic dogs' (*Canis familiaris*) search for an item dropped down  
6 a diagonal tube? The tubes task revisited

7

8 Emma C. Tecwyn \*, Daphna Buchsbaum

9 Department of Psychology, University of Toronto

10

#### Author Note

11 This research was funded by the Natural Sciences and Engineering Research Council of Canada  
12 and the Canada Foundation for Innovation

13 \* Correspondence: TecwynE@cardiff.ac.uk

14 Emma C Tecwyn is now at School of Psychology, Cardiff University.

15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36

Abstract

It has been suggested that domestic dogs—like young human children—have a ‘gravity bias’; they expect an unsupported object to fall straight down, regardless of any obstacles that redirect or halt its path. In the diagonal tube task this bias is revealed by a persistent tendency to search the incorrect location directly beneath the top of the tube the item is dropped into, rather than the correct location attached to the bottom of the tube. We presented dogs (N=112) with seven different versions of the diagonal tube task, replicating and extending previous research, to examine what factors influence their search behavior for an object dropped down a diagonal tube, and investigate their physical reasoning skills more generally. Contrary to previous claims, we found no evidence for dogs exhibiting a persistent, or even a trial 1, gravity bias. However, in line with previous reports, dogs were also unable to search correctly for the reward, even when it could be heard rolling through the tube, though they succeeded when the tube was transparent (Exp. 1a-c). Exp. 2 suggested that dogs might search on the basis of proximity, but Exp. 3a-b ruled this out and showed that they prefer to commence searching at the center of the apparatus. Finally, when potential sources of bias were eliminated from the task (Exp. 4), dogs’ performance was improved, but still not above chance, suggesting that they are unable to reason about the tube’s physical-causal mechanism. We conclude that, on current evidence, the gravity bias might be unique to some primate species.

*Keywords:* comparative cognition; domestic dog; gravity bias; physical reasoning; tubes task

## DOGS' GRAVITY BIAS REVISITED

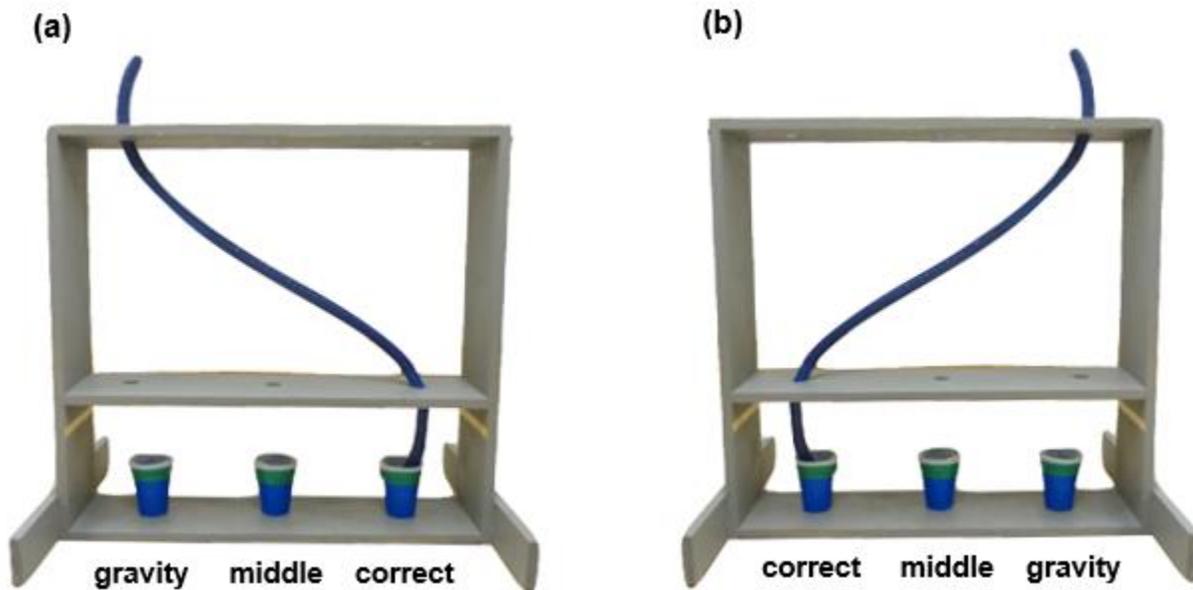
37 What factors really influence domestic dogs' (*Canis familiaris*) search for an item dropped down  
38 a diagonal tube? The tubes task revisited

39 As human adults we possess sophisticated knowledge about the physical world. We know, for  
40 example, that objects continue to exist even when they move out of sight; that a solid object  
41 cannot pass through another solid object; and that gravity causes unsupported objects to fall (e.g.,  
42 Baillargeon, 2002). Understanding how inanimate objects behave and interact with one another  
43 is also important for young humans, as well as being ecologically relevant for many non-human  
44 species. For example, it is extremely useful to be able to track and re-locate objects that move  
45 out of sight, and all terrestrial species experience evidence of the effect of gravity on falling  
46 objects, so it is feasible that similar physical reasoning mechanisms might be widely shared  
47 among species. The developmental and evolutionary origins of our rich physical reasoning  
48 abilities have thus long been of interest to researchers in the fields of cognitive development and  
49 comparative cognition. Do young children and non-human animals (hereafter, animals) reason  
50 about objects in the same way as human adults, or are there fundamental differences?

51 The tubes task (Hood, 1995) has been used widely in the fields of cognitive development  
52 and comparative cognition to investigate children's and animals' physical reasoning abilities –  
53 specifically, how different-aged children and different species reason about the way objects  
54 behave under the influence of gravity (see Tecwyn & Buchsbaum, in press for a review). The  
55 version of the task typically used with children consists of three intertwined opaque tubes  
56 positioned vertically in a frame, each attached to a non-aligned cup at the base of the apparatus  
57 (e.g., Baker, Gjersoe, Sibielska-Woch, Leslie, & Hood, 2011; Bascandziew & Harris, 2010;  
58 Freeman, Hood, & Meehan, 2004; Hood, Wilson, & Dyson, 2006; Jaswal, 2010; Joh & Spivey,  
59 2012). The task as typically used with animals is a simplified versions of this, and involves just a

## DOGS' GRAVITY BIAS REVISITED

60 single diagonally-configured tube (Fig. 1). We will refer to this simplified version as the  
61 'diagonal tube task', and given the comparative focus of the current study, the majority of this  
62 introduction will focus on how individuals perform in this version of the task. Even in the  
63 diagonal tube task that involves only one tube there are typically three possible search locations  
64 at the bottom of the apparatus: the *correct* location connected to the bottom of the diagonal tube;  
65 the *gravity* location, which is aligned directly beneath the release-point of the reward into the top  
66 of the tube, and the *middle* location which is positioned between the other two locations (Fig. 1).  
67 In accordance with the principles of gravity, solidity and containment, when dropped into the top  
68 of the tube, the item travels (invisibly) down through the tube and ends up in the cup attached to  
69 its bottom end. Participants typically first undergo some pre-training to introduce the different  
70 search locations without the tube in place in the apparatus. The experimenter then puts the tube  
71 in place and drops an item (typically a ball for children, a food reward for animals) down the  
72 tube. The participant must then search for the item.



73

74 **Figure 1** Diagonal tube task apparatus showing the opaque tube in a top left-bottom right configuration (a)  
75 and a top right-bottom left configuration (b). The gravity, middle and correct search locations are indicated  
76 for each setup

77           While this is seemingly a straightforward task to solve, young children tend to perform  
78 poorly. Interestingly, when instructed to search for the dropped item in the diagonal tube task, 2-  
79 year-olds make a surprising, non-random error: rather than searching in the correct location  
80 connected to the bottom of the single diagonal tube, they tend to search the gravity location (Fig.  
81 1). Furthermore, they do not search the gravity location only in their first trial when they are  
82 naïve to the task; they continue to do so across repeated trials, despite receiving feedback  
83 regarding the correct location, which remains fixed across trials (number of 2-year-olds  
84 searching gravity location in the diagonal tube task: trial 1: 9/10; trial 2: 8/10; trial 3: 9/10;  
85 Experiment 4 pre-test, Hood, 1995).

86           According to Hood and colleagues (1995; 1998; 2006) “the gravity error is characterized  
87 by repeated search in the box below despite adequate trials with feedback” (p. 304, Hood et al.,  
88 2006). Specifically, young children’s perseverative searching of the gravity location  
89 demonstrates resistance to counterevidence and suggests that they possess a naïve theory about  
90 the influence of gravity on unsupported objects that is challenging to overcome. The search error  
91 seems to be specific to objects moving under the influence of gravity, as children are less likely  
92 to make a comparable error of searching the aligned location in a version of the task involving  
93 upwards motion (Hood, 1998), or where the apparatus is horizontally configured (Hood, Santos,  
94 & Fieselman, 2000). This implies that children’s search error does not reflect a more general  
95 straight trajectory bias, or a proximity bias, though it should be noted that a proximity bias has  
96 never been directly ruled out in a vertical, gravity-based version of the tubes task.

97           It has been suggested that children develop a naïve theory about gravity during their first  
98 year of life due to repeated exposure to objects falling straight down (e.g., Spelke, Breinlinger,  
99 Macomber, & Jacobson, 1992). Given that this ‘straight down’ assumption is usually correct and

## DOGS' GRAVITY BIAS REVISITED

100 therefore typically a useful heuristic to follow, this belief can be difficult to abandon (Hood et al.,  
101 2006), and may therefore interfere with children's ability to succeed at tasks that require the  
102 theory to be ignored (e.g., when an object is dropped down a diagonal tube). Further evidence  
103 that this theory is resistant to counterevidence is the fact that even after young children  
104 participate in a transparent version of the diagonal tube task—which they are able to pass—when  
105 subsequently re-tested with the opaque version they revert to searching the gravity location  
106 (Hood, 1995).

107         How is it that children are able to overcome their gravity bias at around 4 years of age?  
108 Follow-up studies suggest that sufficient inhibitory control and causal knowledge are both  
109 important factors. Dividing the attention of 4- to 5-year olds who would normally succeed with  
110 the 3-intertwined-tube setup by dropping two balls simultaneously causes them to revert to a  
111 gravity bias, suggesting that the bias persists but is typically suppressed by this age (Hood et al.,  
112 2006). Modifying the apparatus to highlight the tube's physical-causal mechanism improves the  
113 performance of children who would usually display a gravity bias (e.g. Bascandziev & Harris,  
114 2011; Joh et al., 2011; Joh et al., 2012). Relatedly, even 2-year-olds do not show a gravity bias in  
115 the table/shelf task (which would be revealed by reliable searching beneath the solid shelf; Hood  
116 et al., 2000), where the physical-causal structure of the task is arguably much simpler than the  
117 tubes task (no diversion of trajectory; no containment).

118         Several studies have explored how non-human species perform in the tubes task, with the  
119 aim of discovering whether the gravity bias is unique to humans, or whether it is also seen in  
120 other species, and could potentially represent an evolutionarily ancient naïve theory based on the  
121 physics of life on earth (e.g. Hood, Hauser, Anderson, & Santos, 1999). Cotton-top tamarins  
122 (*Sanguinus oedipus oedipus*) were the first non-human species to be tested, and the results of this

## DOGS' GRAVITY BIAS REVISITED

123 study are the basis of claims that monkeys show a comparable gravity bias to young children.  
124 While 7/9 individuals searched the gravity location in their first trial, the bias did not  
125 compellingly persist across multiple trials – in trial 2, only 2/9 individuals searched the gravity  
126 location (Hood et al., 1999), and so whether this meets the ‘challenging to overcome’ criterion of  
127 a naïve theory is debatable. Three subjects succeeded at the task across 16 trials, and the errors  
128 made by the other six subjects were distributed evenly between the gravity and middle locations.  
129 Therefore, while this study provides evidence that tamarins’ initial search may have been  
130 influenced by gravity, their behavior across trials does not bear the hallmarks of a naïve theory,  
131 given that an initial bias was easily overcome by several individuals, and errors were as likely to  
132 be directed at the middle location as the gravity location. Hood et al. (1999) also noted that  
133 several tamarins developed a preference to search the middle location, which they suggested was  
134 due to a lack of differentiation between the gravity and middle locations, both of which were in  
135 closer proximity to where the reward was dropped from than the correct location.

136 In a separate study, cotton-top tamarins with prior experience of a horizontal version of  
137 the diagonal tube task did not exhibit a gravity bias when subsequently tested with the standard  
138 vertical version of the task (Hauser, Williams, Kralik, & Moskovitz, 2001), suggesting that any  
139 gravity bias is not particularly robust in this species. As was the case for the tamarins tested by  
140 Hood et al. (1999), Hauser and colleagues (2001) also noted that tamarins in both the vertical and  
141 horizontal versions of their diagonal tube task developed a preference to search the middle  
142 location, with the authors suggesting that this may have been due to them approximating the  
143 position of the invisible food item.

144 Another callitrichid species—common marmosets (*Callithrix jacchus*)—did not exhibit a  
145 gravity bias even in trial 1 when they were naïve to the diagonal tube task (4/7 individuals

## DOGS' GRAVITY BIAS REVISITED

146 searched the gravity location, Cacchione & Burkart, 2012), but it is difficult to draw any firm  
147 conclusions from this small sample of individuals. In their first block of 16 trials marmosets'  
148 searches were randomly distributed between the three locations, but when they erred they were  
149 significantly more likely to search the gravity location than the middle location (though this  
150 difference disappeared in their second block of trials). Interestingly, when presented with a  
151 looking-time version of the task, marmosets looked significantly longer when the reward was  
152 revealed to have ended up in the gravity container than when it was revealed to have ended up in  
153 the correct container. Thus, while they were not able to search correctly at above-chance level, it  
154 is possible that they were implicitly aware of the role of the tube in constraining the reward's  
155 movement (Cacchione & Burkart, 2012; see Lee & Kuhlmeier, 2013, for similar findings with  
156 2-year-old children).

157         A study by Cacchione and Call (2010) presented all four species of non-human great ape  
158 with the diagonal tube task, and found that they did not exhibit a gravity bias (only 8/22 subjects  
159 searched the gravity location in trial 1) – in fact, they were able to locate the food item at above-  
160 chance levels from their first trial (Cacchione & Call, 2010). However, analysis of the errors  
161 made by the apes showed that they were significantly more likely to search the gravity location  
162 than the middle location, suggesting that apes may indeed hold naïve beliefs about gravity, but  
163 unlike 2-year-old children, they are usually able to suppress acting on the basis of this belief  
164 when it is inappropriate (as in the case of the diagonal tube). The findings of an earlier study by  
165 Tomonaga and colleagues (2007) fits with the idea that great apes might have a dormant gravity  
166 bias. Their task used a different measure to the other studies described here (prediction before the  
167 dropping event rather than search afterwards), and involved two crossed tubes rather than a  
168 single diagonal tube. In this context, both juvenile and adult chimpanzees selected the gravity

169 option at above-chance levels, and a further experiment ruled out that their choices were based  
170 on proximity. While the apparent presence of a gravity bias in this study versus the lack of a  
171 reliable gravity bias in Cacchione and Call's (2010) task might be explained by the different  
172 response measures used, it is also possible that apes are able to solve the single diagonal tube  
173 task, but reveal a gravity bias when the task is more complex because more tubes are intertwined,  
174 which is known to increase children's preference for the gravity location (Hood, 1995; Lee &  
175 Kuhlmeier, 2013).

176         Taken together, these studies provide mixed evidence for the existence of a gravity bias in  
177 non-human primates. Cotton-top tamarins showed a significant gravity bias in trial 1, but this did  
178 not persist across trials and they were equally likely to search the gravity and middle locations  
179 overall (Hood et al., 1999). Marmosets searched randomly initially, but were more likely to  
180 search the gravity location when they made a mistake, at least in their first block of trials  
181 (Cacchione & Burkhardt, 2012). Great apes were able to solve the single diagonal tube task, but  
182 were more likely to search the gravity location than the middle location when they erred  
183 (Cacchione & Call, 2010), and chimpanzees showed a gravity bias when they had to predict  
184 where a reward would appear when it was dropped into one of two crossed tubes (Tomonaga et  
185 al., 2007).

186         Only one study to date has investigated whether a non-primate species exhibits gravity-  
187 biased search in the diagonal tube task. When domestic dogs (*Canis familiaris*) were presented  
188 with the diagonal tube task by Osthaus and colleagues (2003), they searched the correct location  
189 significantly less often than in a control task with a straight up and down tube (where the gravity  
190 location and correct location were the same). In the diagonal tube task, 8/16 dogs searched the  
191 gravity location in their first trial. Although dogs chose the gravity location more frequently than

## DOGS' GRAVITY BIAS REVISITED

192 the two alternatives in trial 1, their performance did not differ significantly from random search  
193 (two-tailed binomial test: 0.33 chance of searching gravity location;  $p = 0.19$ ; not reported in the  
194 original paper). The number of dogs searching the gravity location decreased rapidly across  
195 trials, and in trial 16 only 2/16 dogs made a gravity error. Concurrently, the number of dogs  
196 searching in the correct location increased across trials: from 3/16 in trial 1 to 10/16 in trial 16.  
197 Across all trials several dogs searched in the middle location (5/16 in trial 1, and 4/16 in trial 16);  
198 in fact, overall, more than 40% of searches were directed at the middle location. The authors  
199 suggested that searching the middle might represent a strategy that dogs adopt when they are  
200 uncertain about the reward's location. In a follow-up experiment where the middle location was  
201 removed as a search option (Experiment 3, Osthaus et al., 2003), 6/8 dogs searched the gravity  
202 location in their first trial, but they learned to search the correct location even more quickly than  
203 they did in the experiment where the middle location was available as a search option.

204       Taken together, these data provide no evidence for a group-level gravity bias in dogs that  
205 persists across trials (i.e., that could constitute a naïve theory of gravity), and suggestive but non-  
206 significant evidence for a possible initial gravity bias present in trial 1. However, the authors  
207 conclude that “Dogs, like toddlers and non-human primates, display a gravity bias”, though they  
208 acknowledge that “dogs can learn to overcome this [gravity bias]” (p. 497, Osthaus et al., 2003).  
209 Based on the findings of this single study, several authors have gone on to report that dogs  
210 exhibit a persistent gravity bias (e.g. Bascandziev & Harris, 2011; Cacchione & Call, 2010; Joh  
211 et al., 2011; Kunder, Reyes, Taglang, Baruch, & German, 2010; Range, Möslinger, & Virányi,  
212 2012; Tomonaga, Imura, Mizuno, & Tanaka, 2007). However, having examined the data  
213 presented in Osthaus et al. (2003), we do not believe there are grounds for such a strong

214 conclusion. It is therefore puzzling that the claim that dogs having a robust and persistent gravity  
215 bias comparable to that of human toddlers is so pervasive in the literature.

216         Given that on existing evidence, whether and to what extent dogs exhibit gravity-biased  
217 search in the diagonal tube task remains unclear, the first aim of the current study was to re-  
218 examine dogs' performance in the diagonal tube task, to establish whether dogs, like young  
219 children, show a gravity bias (Experiment 1a). An additional aim was to use the diagonal tube  
220 task to investigate dogs' physical-causal reasoning abilities more generally, which remain  
221 relatively understudied in comparison with their socio-cognitive skills, as well as in comparison  
222 with the physical-causal reasoning abilities of other non-human taxa such as primates and  
223 corvids. As a first step to address this deficit, we replicated previous diagonal tube task  
224 experiments that have been conducted with apes (Cacchione & Call, 2010) to investigate how  
225 auditory (Experiment 1b) and visual (Experiment 1c) information about the tube's causal  
226 mechanism influences dogs' performance in the diagonal tube task.

227         Seeing as several dogs in Experiments 1a–c exhibited a tendency to search the middle  
228 location, as was the case in Osthaus et al.'s (2003) previous study with dogs, in Experiment 2 we  
229 replicated our Experiment 1a but with the middle location removed, to see how dogs' search  
230 shifted when searching the middle location was no longer an option. This also provided a  
231 replication of Osthaus et al.'s (2003) Experiment 3, but with a larger sample of dogs (16 vs. 8).  
232 Based on the results of our Experiment 2, in Experiments 3a and 3b we probed whether dogs'  
233 search might indeed be influenced by a gravity bias in some situations, or whether their behavior  
234 might in fact be better explained by proximity between the reward's release point and the search  
235 locations. These experiments represent novel versions of the diagonal tube task, as while the role  
236 of proximity has been indirectly explored in comparisons of the vertical tubes task to the

237 horizontal tubes task (Hauser et al., 2001; Hood et al., 2000) and the version involving upwards  
238 motion (Hood, 1998), to our knowledge it is the first time that the gravity location and most  
239 proximal location to where the reward is dropped from have been de-confounded in a vertical  
240 version of the single diagonal tube task in any species (though see Experiment 2 of Tomonaga et  
241 al., 2007 for a test of the proximity bias in a two-tube version of the task).

242 Finally, in Experiment 4 we presented dogs with a version of the diagonal tube task  
243 described in Gomez (2005) in which they could not search on the basis of any of the biases that  
244 might have guided them in Experiments 1–3 (namely gravity, middle or proximity), to see  
245 whether this would enable them to succeed, as would be predicted if they do understand the  
246 causal mechanism of the tube, but are unable to inhibit searching on the basis of some bias.

247 Given that an important aspect of Hood's 'naïve theory' account of children's gravity bias  
248 (e.g., Hood, 1995; 1998; Hood et al., 2006) is that the bias is resistant to counterevidence—that  
249 is, it persists across repeated trials in spite of counterevidence—in all experiments we examined  
250 both how dogs performed in trial 1, but also whether and how performance changed across  
251 repeated trials. It is possible that dogs (and other animals) exhibit an initial gravity bias, but  
252 unlike for young children this bias does not persist across trials. If this were the case then such a  
253 bias would not be a candidate for a naïve theory of gravity, which would suggest that any bias is  
254 qualitatively different from that shown by young children.

## 255 **General Methods**

### 256 **Subjects**

257 All test subjects were pet dogs whose owners volunteered to participate in the study. Dog  
258 owners were recruited via email, local advertisements and local dog training facilities, and  
259 subsequently completed a questionnaire. In order to participate, dogs could not have a prior  
260 history of aggression towards humans and had to be in generally good health (including no

261 known issues with their vision or hearing). There were no breed or age restrictions, though all  
262 dogs but one were at least 6 months old (see Table S1 for further subject details including breed).  
263 Dogs participated either in the Canine Cognition Lab at the University of Toronto or in a similar  
264 sized space at a dog training facility in the Toronto area. Each dog only participated in one of the  
265 experiments.

266 **Materials**

267 The apparatus used was based on Hood's (1995) tubes task for children, and subsequent  
268 tubes tasks adapted for use with animals (e.g., Cacchione & Call, 2010; Hood et al., 1999;  
269 Osthaus et al., 2003, Fig. 1). It consisted of a wooden frame (height: 80 cm, width: 79 cm, depth  
270 18.5 cm) with orthogonal 'feet' (length: 54.5 cm) for stability and a mid-section at a height of 29  
271 cm to hold the bottom of the tube in place. There were three holes (3 cm diameter, 25.5 cm apart)  
272 in the top of the frame and the mid-section above the cups that the tube could be passed through.  
273 The possible search locations were opaque paper cups (height: 11 cm; diameter: 8 cm) that all  
274 had inaccessible treats hidden in the bottom to control for odor cues and were padded with cotton  
275 wool and soft fabric to mask the sound of treats dropping into them.

276 Our apparatus differed from that used by Osthaus et al. (2003), in that their search  
277 locations at the bottom consisted of three adjacent boxes without any clear separation between  
278 them. We reverted to a setup more similar to the original Hood (1995) apparatus as we felt that  
279 the lack of clear physical separation between search locations may have been confusing for dogs,  
280 and indeed it has been suggested that subjects' tendency to search the middle location in previous  
281 studies may have been due to spatial confusion of the gravity and middle locations (Hood et al.,  
282 1999).

283           Across all experiments a single tube (diameter: 2.75 cm) was positioned diagonally in the  
284 apparatus. The start and end points of the diagonal tube within the frame—and hence the length  
285 of the tube—varied between experiments, as did the number and position of the search locations  
286 (see individual experiment sections and Fig. 2 and 4 for details). In all of the experiments except  
287 for Experiment 1b a small piece (~ 1 cm<sup>3</sup>) of freeze-dried liver treat that moved inaudibly  
288 through tube was used. To further eliminate any sound an electric fan was on in the room  
289 throughout the testing session to provide white noise. The fan was on from when the dog entered  
290 the testing space so they had time to become accustomed to the sound before starting the task. In  
291 Experiment 1b a similarly sized but harder and heavier liver-based treat was used and the fan  
292 was not turned on.

## 293 **Procedure**

### 294 **Warm-up.**

295           The aim of the warm-up was to ensure that dogs felt comfortable in the testing space, and  
296 that they would interact with the cups to indicate their choice of search location during the  
297 experiment. Upon arrival in the testing area, dogs were given approximately five minutes to  
298 explore the space off-leash while the owner completed an informed consent form. After this  
299 initial exploration period, dogs were introduced to the cups by the main experimenter, who  
300 placed one cup on the ground, then showed the dog a treat and dropped it into the cup and  
301 encouraged the dog to retrieve it by giving a release command (“OK!” unless the owner  
302 suggested an alternative). Some dogs spontaneously knocked the cup over and retrieved the treat;  
303 for dogs that touched the cup with their muzzle or paw but did not knock it over, the  
304 experimenter tipped the treat out for the dog to eat. This was repeated until the dog  
305 touched/knocked over the cup a total of three times. After this initial off-leash warm-up period,

306 dogs were put on leash and handled by a second experimenter (handler). Owners were present  
307 during testing but were seated at the side of the room behind the dog's starting position (see Fig.  
308 S1) and were asked not to interact with their dog during the experiment.

309 The study was approved by the University of Toronto's University Animal Care  
310 Committee (UACC). All procedures were in accordance with Ontario's Animals for Research  
311 Act, and the federal Canadian Council on Animal Care and complied with the APA Ethical  
312 Standards for Use of Animals in Research. All sessions were video recorded.

313 **Cup pre-training trials.**

314 The aim of the cup pre-training trials was to introduce dogs to searching for treats in the  
315 cups whilst they were positioned in the apparatus, and to both measure and reduce the influence  
316 of any prior location biases during the test phase. The cups were in position in the bottom of the  
317 apparatus (the number and location of cups varied between experiments, see individual  
318 experiment sections and Fig. 2 and 4 for details) and the tube was not present.

319 The main experimenter knelt behind the apparatus and the handler held the dog on leash  
320 in front of the apparatus at a distance of approximately 160 cm. The experimenter showed the  
321 dog a treat, moved it back and forth above the frame mid-section to ensure the dog was tracking  
322 it, then dropped it through one of the holes into the cup underneath (Fig. S2a, Video S1). The  
323 experimenter then put her hands behind her back, stared at a fixed point on the wall behind the  
324 dog, then gave a release command (e.g. "OK!") and the dog was allowed to search exhaustively  
325 for the treat. A choice was defined as a dog making physical contact with a cup with their muzzle  
326 or paw (sniffing a cup, staring at a cup or lying down in front of a cup did not constitute a  
327 choice). As in the warm-up, once the dog had touched the cup the experimenter tipped the treat  
328 out of the cup for the dog if necessary. Once a dog had chosen a cup it was removed by the

329 experimenter. This dropping of treats into cups was repeated in a pseudorandom order (with the  
330 constraint that the treat was not dropped into the same location in more than two consecutive  
331 trials) until the dog successfully located the treat on their initial search on six consecutive trials  
332 (an equal number of times from each location). The individual cups were randomly interchanged  
333 between trials so the same cup did not always appear in the same location.

334       Following Osthaus et al. (2003) we set the maximum number of cup pre-training trials to  
335 30; however, if a dog showed a persistent location/side bias (defined as 12 consecutive searches  
336 of the same location) we administered the following training: treats were no longer dropped into  
337 the preferred cup, and pseudorandomly dropped into the other cups until dogs got 6 consecutive  
338 trials correct; then we reverted to all 3 cups and they had to get another 6 in a row correct.

339 Therefore in a few cases the total number of trials including these training trials went above 30  
340 (see individual experiment results for the range of pre-training trials required to reach criterion).

341       **Tube familiarization.**

342       The aim of the tube familiarization was to demonstrate the tube mechanism to the dogs.  
343 While this step was not included in Osthaus et al.'s (2003) study, previous child studies (e.g.  
344 Hood 1995) and some non-human primate studies (e.g. Cacchione & Burkhardt, 2012) have  
345 incorporated this step. Given that dogs likely have little experience of items travelling through  
346 hollow tubes it seemed an important step to include. The unconnected tube was held aloft so it  
347 formed a loose S-shape, with the bottom end approximately at the dog's head height (see Fig.  
348 S2b). The experimenter showed the dog a treat, then dropped it into the top of the tube so it  
349 rolled out of the bottom of the tube and onto the ground, and the dog was allowed to retrieve it.  
350 This was repeated until the dog spontaneously retrieved the treat (i.e. the experimenter did not  
351 need to indicate the treat's location on the ground) on three consecutive occasions.

352           **Test trials.**

353           Each dog participated in 12 test trials presented in a single block immediately after the  
354 tube familiarization. While Osthaus et al. (2003) presented dogs with 16 trials per condition,  
355 pilot work suggested that the dogs in our study began to lose attention and/or become satiated  
356 after around 12 trials (having already completed the cup pre-training and tube familiarization). In  
357 line with previous studies with non-human animals, in all experiments dogs were randomly  
358 assigned to one of two possible diagonal configurations of the tube, which were a mirror image  
359 of one another (e.g. top left-bottom right, or top right-bottom left; Fig. 1). The experimenter  
360 inserted the tube into the frame in full view of the dog and it remained in this position for all of  
361 the test trials.

362           For each test trial the experimenter knelt behind the apparatus and the handler held the  
363 dog on leash in front of the apparatus at a distance of approximately 160 cm. The experimenter  
364 showed the dog a treat, moved it back and forth across the top of the frame until the dog tracked  
365 it, then dropped it into the top of the tube, showed the dog her empty hand, placed her hands on  
366 her lap, stared straight ahead at a fixed point behind the dog, then gave the dog a release  
367 command (e.g., “OK!”) to search for the treat (see Video S2). The dog was allowed to search  
368 exhaustively until they located the treat. We allowed exhaustive search to match previous studies  
369 with dogs (Osthaus et al., 2003), apes (Cacchione & Call, 2009) and monkeys (Hood et al.,  
370 1999). Pilot work also revealed that dogs quickly stopped participating (they refused to search) if  
371 they were only allowed to search one location and chose incorrectly, which meant they were not  
372 rewarded for that trial. While one might imagine that allowing exhaustive search reduces the  
373 incentive for the subject to make an initial correct choice (because they ultimately get a reward  
374 anyway), previous work has demonstrated that this is not the case for monkeys at least: in two

375 tubes task studies, performance did not differ according to whether subjects were allowed to  
376 search exhaustively, or only allowed to search a single location and therefore went unrewarded if  
377 they chose incorrectly (Hauser et al., 2001).

378 **Data coding & analysis.**

379 In both the cup pre-training trials and the test trials we scored the location that dogs  
380 searched first. We coded searches as correct or incorrect. For the 3-cup versions of the task  
381 (Experiments 1a-1c), if dogs searched incorrectly then their search was further coded as directed  
382 at the middle or gravity location. To investigate performance in trial 1 of each experiment we  
383 used Chi-square goodness of fit tests (3-cup versions) or binomial tests (2-cup versions) to see  
384 whether the distribution of dogs' search differed from random. We used mixed effects logistic  
385 regression models that assumed a fixed slope across subjects (including a random slope term did  
386 not significantly improve fit for any experiment) to examine successful performance across all  
387 trials in each individual experiment, and to look for change in performance over trials (lme4  
388 package version 1.1.13; Bates, Maechler, Bolker, & Walker, 2015) in the R environment (Version  
389 3.3.3; R Development Core Team, 2017). We used the same approach to examine the nature of  
390 dogs' errors, and look for any change in the nature of errors made across trials. We also  
391 compared overall performance in each experiment to chance. For the 3-cup versions of the task  
392 (Experiments 1a-1c) we adjusted the intercept to account for testing against 0.33 (as opposed to  
393 the standard 0.5 in the 2-cup versions) and used that to calculate an adjusted  $z$  statistic and obtain  
394 the correct  $p$ -value. We used binomial tests to examine the performance of individual dogs. For  
395 all 3-cup versions (Exp. 1a-1c), where we were interested in seeing whether individuals dogs  
396 either searched correctly, in the middle, or at the gravity location significantly more often than  
397 expected by chance we used a Bonferroni correction for multiple comparisons, so for these tests

398 alpha was 0.017 (0.05/3). Finally, mixed effects logistic regression models were also used to  
399 compare performance between-subjects across all 3-cup versions (Exp. 1a-1c) and all 2-cup  
400 versions (Exp. 2-4) of the task. As for the individual experiment analyses, we assumed a fixed  
401 slope across subjects because including a random slope term did not significantly improve fit for  
402 either comparison. All tests were two-tailed and alpha was 0.05.

403 A second coder scored the test trials of a randomly selected six dogs per experiment from  
404 video footage to assess inter-observer reliability. Cohen's kappa for which location the subject  
405 searched first on each trial was 0.98 (excellent agreement between coders).

### 406 **Experiment 1: Replicating previous versions of diagonal tube task**

#### 407 **Experiment 1a: The classic diagonal tube task**

408 In Experiment 1a, we presented dogs with the classic version of the diagonal tube task  
409 used in comparative studies, where a reward is dropped down an opaque diagonal tube, and  
410 travels invisibly and inaudibly through it into the cup at its bottom end. In this version, no direct  
411 perceptual information (either visual or auditory) regarding the reward's location is available  
412 after it disappears into the top of the tube (Cacchione & Burkart, 2012; Hood et al., 1999;  
413 Osthaus et al., 2003). The aim was to generate additional data to address the widespread claim in  
414 the literature that dogs share a naïve theory of gravity with young children, despite limited  
415 empirical evidence to support this.

#### 416 **Subjects**

417 Sixteen dogs (4 male, 12 female; mean age =  $66 \pm 10$  months) participated in Experiment  
418 1a (Table S1). Three additional dogs were tested but excluded because they did not reach  
419 criterion in the cup pre-training trials (1), or because they failed to complete the test trials (2).

420 **Materials**

421 The tube was opaque and configured top left to bottom right or top right to bottom left  
422 (Fig. 1 and 2a). Freeze-dried liver treats were used for the test trials.

423 **Results**

424 On average, dogs required  $12.2 \pm 2.3$  trials (mean  $\pm$  SEM; range = 6 – 44; median = 9.5)  
425 to reach criterion in the cup pre-training trials. In their first test trial dogs did not show a bias to  
426 search any particular location (chance:  $16/3 = 5.33$  dogs searching correctly;  $\chi^2 = 2.38$ ,  $df = 2$ ,  $p$   
427 = 0.304); rather, they searched randomly for a treat that travelled invisibly and inaudibly down a  
428 diagonal tube. Half of the dogs (8/16) searched the middle location, five searched the gravity  
429 location, and three searched the correct location (Fig. 2a). Therefore, we found no evidence that  
430 dogs' search behavior was guided by gravity when they were naïve to the task.

431 Across the 12 test trials, 52% of searches were directed to the middle location (mean<sub>middle</sub>  
432 =  $6.2 \pm 0.7$  trials), 32% of searches were directed to the correct location (mean<sub>correct</sub> =  $3.8 \pm 0.7$   
433 trials), and just 16% of searches were directed to the gravity location (mean<sub>gravity</sub> =  $2.1 \pm 0.6$   
434 trials; Fig. 3a). A mixed effects logistic regression model revealed that, as a group, dogs'  
435 tendency to search the correct location did not differ significantly from the 33% expected by  
436 chance ( $z = 0.61$ ,  $p = 0.54$ ).

437 The mixed effects logistic regression model revealed a significant improvement in  
438 performance across trials (Fig. 2a and Fig. 3b; trial log odds = 0.17,  $z = 3.192$ ,  $p = 0.001$ , Table  
439 S2). We ran a separate mixed effects logistic regression model to examine the search errors that  
440 dogs made, which revealed that, across the 12 test trials, on trials where dogs erred they were  
441 significantly more likely to search the middle location than the gravity location ( $z = 3.71$ ,  $p <$

442 0.001; Fig. 3a). Across trials, the number of gravity searches decreased significantly relative to  
443 the number of middle searches (trial log odds = -0.16,  $z = -2.27$ ,  $p = 0.022$ ; Fig. 2a).

444 We also examined individual performance, and whether individual dogs had a preference  
445 for any of the search locations. After correcting for multiple comparisons (Bonferroni correction;  
446  $\alpha = 0.017$ ; a dog had to search the same location in at least 9/12 trials to produce a  $p$ -value <  
447 0.017 in a binomial test) only 1/16 dogs searched correctly significantly more often than  
448 expected by chance across the 12 test trials (10/12 trials correct;  $p < 0.001$ ). Two out of 16 dogs  
449 had a significant middle-location preference (10/12 – 11/12 trials;  $p < 0.001$ ) and no dogs  
450 exhibited a significant preference for searching the gravity location (maximum number of gravity  
451 searches = 8/12, Table S3).

## 452 Discussion

453 When no perceptual cues were available, like children (Hood, 1995), monkeys (Hood et  
454 al., 1999; Hauser et al., 2001) and dogs (Osthaus et al., 2003) tested previously, dogs as a group  
455 failed to locate a reward dropped down a diagonal tube, either in trial 1 or across 12 trials.  
456 However, there was no evidence that they searched on the basis of a gravity bias—when dogs  
457 erred, they were significantly more likely to search the middle location than the gravity location.  
458 In this respect their performance differed from that of tamarins, who exhibited a trial 1 gravity  
459 bias (Hood et al., 1999), and children, who seemed to show a gravity bias that was difficult to  
460 overcome, even after several repeated trials (Hood, 1995). The performance of dogs in the  
461 current experiment also differed from that of great apes, who were able successfully locate the  
462 reward at above chance levels within 9 trials (Cacchione & Call, 2010). Apes have previously  
463 demonstrated superior physical-causal reasoning skills compared with dogs (Bräuer, Kaminski,  
464 Riedel, Call, & Tomasello, 2006), so it is possible that they were better able to understand the

465 role of the tube in constraining the path of the reward. In the following two experiments we  
466 explored whether highlighting the tube's physical-causal mechanism by making the reward's  
467 passage through the tube audible (Experiment 1b) or visible (Experiment 1c) would improve  
468 dogs' performance in the diagonal tube task.

#### 469 **Experiment 1b: Auditory cues available**

470 In Experiment 1b we investigated whether being able to hear the reward travelling  
471 through the tube (but still not hear it landing in the cup) would enable dogs to perform better in  
472 the task, either because they could acoustically track the reward travelling through the tube, or  
473 because the sound provided some information regarding the tube's causal mechanism. Great apes  
474 tested with a comparable version searched randomly in their first trial, but were able to  
475 successfully locate the reward at above chance levels within their first block of nine trials  
476 (Cacchione & Call, 2010). Dogs have not previously been presented with this version of the  
477 diagonal tube task.

#### 478 **Subjects**

479 Sixteen dogs (10 male, 6 female; mean age =  $42 \pm 9$  months) participated in Experiment  
480 1b (Table S1). Three additional dogs were tested but excluded because they did not complete the  
481 cup pre-training trials (1), or because their choice of search location was unclear in the test trials  
482 (2).

#### 483 **Materials**

484 The tube was opaque and positioned from top left to bottom right, or top right to bottom  
485 left (Fig. 2b). A hard, heavier liver-based treat that made a rattling noise as it rolled down the  
486 tube was used as the food reward and the electric fan was turned off.

487 **Results**

488 On average, dogs required  $12.9 \pm 2.6$  trials (mean  $\pm$  SEM; range = 6 – 35; median = 6) to  
489 reach criterion in the cup pre-training trials. In the first test trial, dogs did not show a bias for  
490 searching any particular location ( $\chi^2 = 0.88$ ,  $df = 2$ ,  $p = 0.644$ ; middle: 7/16; gravity: 5/16;  
491 correct: 4/16; Fig. 2b).

492 Across the 12 test trials, 42% of searches were directed to the middle location (mean<sub>middle</sub>  
493 =  $5.0 \pm 0.8$  trials), 36% of searches were directed to the correct location (mean<sub>correct</sub> =  $4.4 \pm 0.1$   
494 trials), and just 22% of searches to the gravity location (mean<sub>gravity</sub> =  $2.6 \pm 0.6$  trials; Fig. 3b).  
495 The mixed effects logistic regression model revealed that, as a group, dogs' tendency to search  
496 the correct location did not differ from the 33% expected by chance ( $z = -0.41$ ,  $p = 0.68$ ; Fig. 3a).  
497 However, their performance improved significantly across trials (trial log odds = 0.21,  $z = 3.60$ ,  $p$   
498  $< 0.001$ ; Fig. 2b; Fig. 3b; Table S2). Examining the search errors that dogs made revealed that, as  
499 in Experiment 1a, dogs were significantly more likely to search the middle location than the  
500 gravity location ( $z = 2.42$ ,  $p = 0.015$ ; Fig. 3a). Across trials, the number of gravity searches  
501 relative to the number of middle searches did not change significantly (trial log odds = -0.03,  $z =$   
502 -0.48,  $p = 0.632$ ; Fig. 2b).

503 We also examined individual performance, and whether individual dogs had a preference  
504 for any of the search locations. After correcting for multiple comparisons (Bonferroni correction;  
505  $\alpha = 0.017$ ; a dog had to search the same location in at least 9/12 trials to produce a  $p$ -value  $<$   
506 0.017 in a binomial test) four dogs searched correctly significantly more often than expected by  
507 chance across twelve trials (9/12 – 11/12 trials correct;  $p < 0.012$ , see Table S4 for apparatus  
508 configuration information for these dogs). Two dogs showed a significant middle-location  
509 preference (11/12 – 12/12 trials;  $p < 0.001$ ) and as in Experiment 1a, no individual dogs

510 exhibited a significant gravity location preference (maximum number of gravity searches = 7/12,  
511 Table S3).

## 512 **Discussion**

513 When acoustic cues were available, dogs still failed to locate the reward at above chance-  
514 level, either in trial 1 or across 12 trials. Great apes tested previously with an acoustic diagonal  
515 tube task searched randomly in trial 1 like dogs in the present experiment, but unlike dogs, apes  
516 performed above chance across a 9-trial session (Cacchione & Call, 2010). Apparently, apes  
517 were more able than dogs to utilize the sound cue, either by tracking the reward's movement  
518 through the tube to the correct location, or because the sound highlighted the tube's causal  
519 mechanism. This fits with previous research suggesting that, compared to other species, dogs are  
520 relatively poor at using physical-causal cues to locate food (apes: Brauer et al., 2006; wolves:  
521 Lampe et al., 2017).

## 522 **Experiment 1c: Transparent tube**

523 Hood (1995) found that if the tube was translucent so that it was possible to observe the  
524 movement of the object dropped down it, then 2.5-year-old children were able to successfully  
525 locate the item. In Experiment 1c we investigated whether dogs—who have not previously been  
526 tested with a transparent version of the diagonal tubes task—would be able to solve the diagonal  
527 tube task if they were able to *see* the reward moving through the tube, i.e. when the reward was  
528 *visibly* displaced by gravity and constrained by the tube. This is important because if dogs do not  
529 succeed in this version of the task, this might suggest that there are other task demands limiting  
530 their performance. For example, an inability to search correctly in a transparent version could be  
531 due to lack of motivation, some physical constraint of the apparatus (e.g., dogs avoid searching  
532 the cup with the tube attached because it is harder to access), or an object permanence/working

533 memory failure, such that once the object is out of sight dogs are completely unable to reason  
534 about its location (though success in the cup pre-training trials makes this unlikely).

### 535 **Subjects**

536 Sixteen dogs (9 male, 7 female; mean age =  $67 \pm 9$  months) participated in Experiment 1c  
537 (Table S1). Five additional dogs were tested but excluded because they did not reach criterion in  
538 the cup pre-training (2), they did not complete the cup pre-training trials (1), they did not  
539 complete the test trials (1) or because their choice of search location was unclear in the test trials  
540 (1).

### 541 **Materials**

542 The tube was transparent so the reward could be seen sliding through it and was  
543 positioned either top left-bottom right, or top right-bottom left (Fig. 2c). As in Experiment 1a, the  
544 light freeze-dried liver treats were used and the electric fan was switched on to mask any residual  
545 sound.

### 546 **Results**

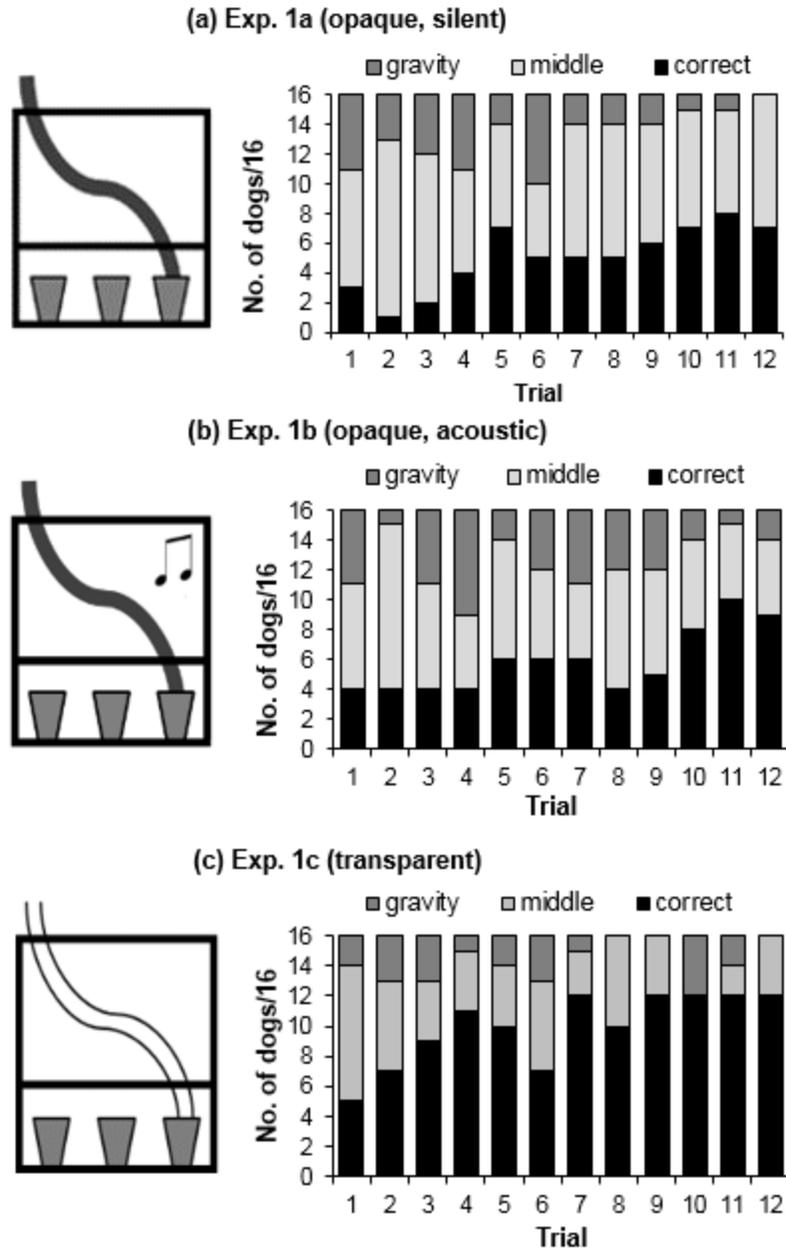
547 On average, dogs required  $10.4 \pm 1.1$  trials (mean  $\pm$  SEM; range = 6 – 19; median = 9.5)  
548 to reach criterion in the cup pre-training trials. As in Experiments 1a and 1b, dogs did not show a  
549 bias for searching any particular location in their first trial ( $\chi^2 = 4.63$ ,  $df = 2$ ,  $p = 0.099$ ), though  
550 again, more dogs searched the middle location (9/16) than the gravity location (2/16) or the  
551 correct location (5/16).

552 Across the 12 test trials, 58% of searches were directed to the correct location (mean<sub>correct</sub>  
553 =  $7.0 \pm 0.8$  trials), 28% of searches were directed to the middle location (mean<sub>middle</sub> =  $3.4 \pm 0.7$   
554 trials), and just 14% of searches to the gravity location (mean<sub>gravity</sub> =  $1.6 \pm 0.4$  trials; Fig. 2a). A  
555 mixed effects logistic regression model revealed that, as a group, dogs searched the correct

## DOGS' GRAVITY BIAS REVISITED

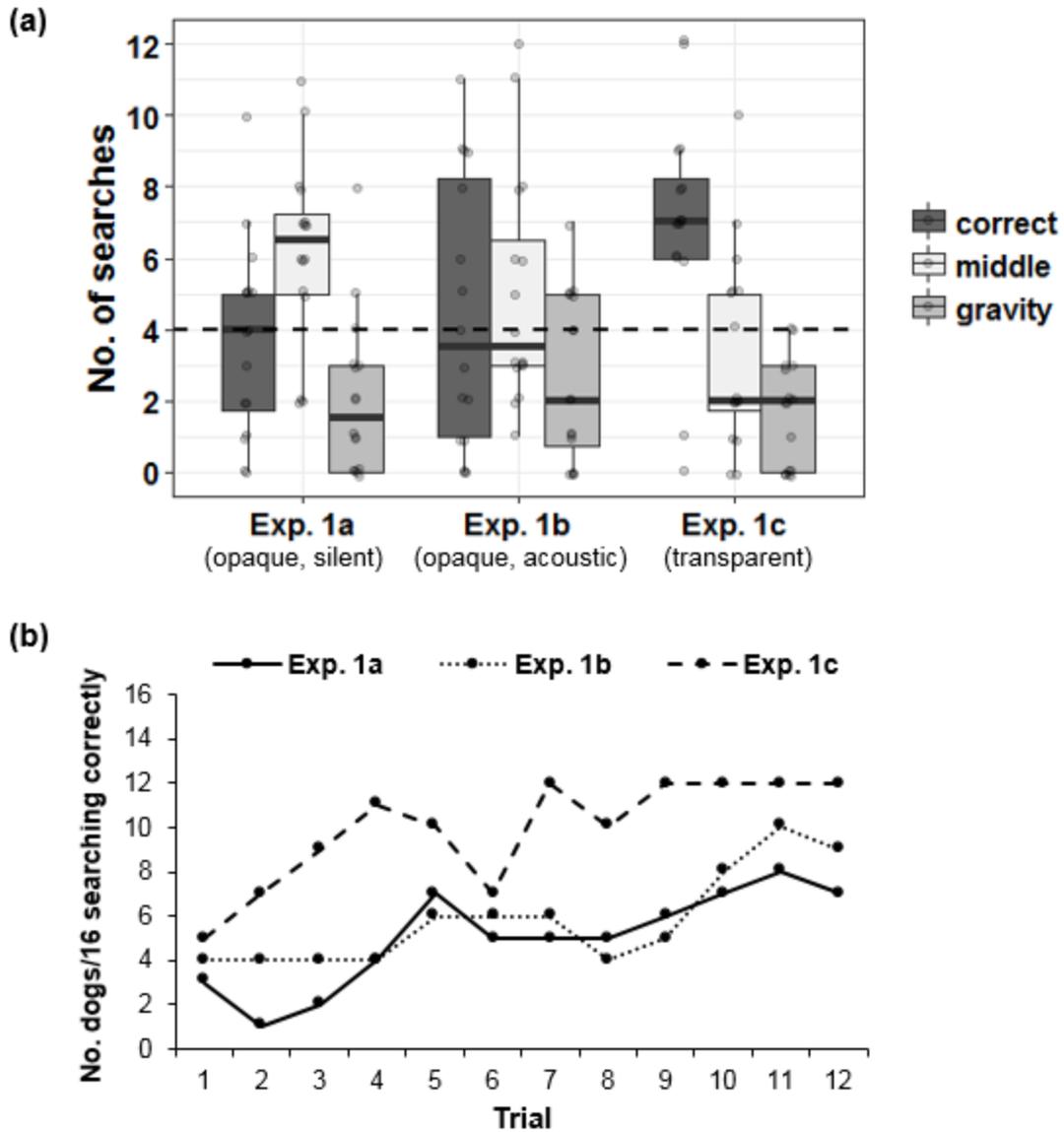
556 location more often than the 33% expected by chance ( $z = 3.14, p = 0.002$ ; Fig. 2a). The mixed  
557 effects logistic regression model revealed a significant effect of trial on performance (log odds =  
558 0.18;  $z = 3.59, p < 0.001$ ), so, as in Experiments 1a and 1b, dogs were more likely to search  
559 correctly across trials (Fig. 2c and Fig. 3b; Table S2). When dogs made search errors, as in  
560 Experiments 1a and 1b, they were more likely to be directed to the middle location than the  
561 gravity location ( $z = 2.245, p = 0.025$ ; Fig. 3a). Across trials, the number of gravity searches  
562 decreased relative to the number of middle searches, but not significantly so (trial log odds = -  
563 0.11,  $z = -1.73, p = 0.08$ ; Fig. 2c).

564 We also examined individual performance, and whether individual dogs had a preference  
565 for any of the search locations. After correcting for multiple comparisons (Bonferroni correction;  
566  $\alpha = 0.017$ ; a dog had to search the same location in at least 9 trials to produce a  $p$ -value  $< 0.017$   
567 in a binomial test), 4/16 dogs searched correctly significantly more often than expected by  
568 chance across twelve trials (9/12 - 12/12 trials correct;  $p < 0.012$ , see Table S4 for apparatus  
569 configuration information for these dogs); 1/16 dogs had a significant middle-location preference  
570 (10/12 trials;  $p < 0.001$ ) and no dogs exhibited a significant gravity location preference  
571 (maximum number of gravity searches = 4/12, Table S3).



572

573 **Figure 2.** Schematic representation of the configuration of the apparatus and the number of dogs searching  
 574 each location across trials 1-12 in Experiments 1a-c where there were always three search locations. Black  
 575 indicates correct cup; light-grey indicates middle cup; mid-grey indicates gravity cup  
 576



577

578 **Figure 3.** (a) Box plot showing the distribution of dogs' searches of the correct, middle and gravity locations  
 579 in Experiments 1a (opaque, silent), 1b (opaque, acoustic) and 1c (transparent, silent). The dashed horizontal  
 580 line represents the expected number of searches per location if search was random. (b) Comparison of the  
 581 number of dogs out of 16 that searched the correct location in each trial in Experiments 1a, 1b and 1c  
 582

583 **Discussion**

584 When the displacement of the reward through the tube was visible, dogs, like 2-year-old  
 585 children (Hood 1995) tended to succeed at searching correctly for it across the 12 test trials  
 586 (though dogs did still make errors, especially in early trials). Like dogs, cotton-top tamarins that

587 participated in a transparent tube version of the task searched randomly in trial 1, but only 2/5  
588 tamarins performed above chance across 10 trials (Hauser et al., 2001). Importantly, Exp. 1c  
589 shows that solving the diagonal tube task is within the capabilities of dogs if they have sufficient  
590 perceptual information, i.e., poor performance in opaque versions is not due to a lack of  
591 motivation, physical constraints imposed by the apparatus, or a working memory/object  
592 permanence failure.

### 593 **Comparison of performance in Experiments 1a – 1c and interim discussion**

594         Across Experiments 1a–c, dogs were presented with a situation where a treat was  
595 dropped down a diagonal tube, and there were three possible search locations at the bottom of  
596 the apparatus corresponding to correct, middle and gravity locations. As well as replicating  
597 Osthaus et al's (2003) study with dogs (Experiment 1a), we manipulated the availability of  
598 auditory (Experiment 1b) and visual (Experiment 1c) information to dogs, and have thus  
599 replicated previous studies with non-human primates (Cacchione & Call, 2010) and human  
600 children (Hood, 1995), to facilitate comparison of performance between species in the diagonal  
601 tube task.

602         Mixed effects logistic regression that assumed a fixed slope across subjects was used to  
603 compare dogs' ability to search correctly between Experiments 1a, 1b and 1c (Table S5). Dogs  
604 were significantly more likely to search the correct location when the tube was transparent (Exp.  
605 1c), compared with when no perceptual cues were available (log odds = 1.47,  $z = 2.52$ ,  $p =$   
606 0.012), as well as when only acoustic cues were available (log odds = 1.31,  $z = 2.23$ ,  $p = 0.026$ ;  
607 Fig. 3b). Performance did not differ between Exp. 1a and Exp. 1b (log odds = 0.16,  $z = 0.27$ ,  $p =$   
608 0.790; Fig. 3b). Thus, being able to see the reward's trajectory helps dogs to identify its end  
609 location, but being able to hear it travelling through the tube does not.

610 We found no evidence for a gravity bias in Experiment 1a–1c; the gravity location was  
611 the least-searched option in all three experiments. This was true whether we considered  
612 performance in trial 1, across all 12 trials, or at an individual level. On the basis of these findings  
613 (and indeed the results of Osthaus et al.'s (2003) Experiment 1 diagonal tube condition) we  
614 conclude that, contrary to previous claims in the literature, dogs' search behavior is not primarily  
615 guided by a gravity bias in the diagonal tube task.

616 In Experiments 1a and 1b the middle cup was the most searched location and several  
617 individual dogs showed a significant middle-bias. Even in Experiment 1c where dogs succeeded  
618 at locating the reward overall, the middle was the second most common choice across trials. This  
619 preference for commencing searching in the middle was noted of the dogs in the diagonal tube  
620 condition of Osthaus et al.'s (2003) Experiment 1, and has also been recorded for some cotton-  
621 top tamarins (Hauser et al., 2001; Hood et al., 1999) and common marmosets (Cacchione &  
622 Burkhardt, 2012). It has previously been suggested that this tendency might be due to subjects  
623 confusing the middle location with the gravity location due to their adjacent spatial proximity  
624 (Hood, 1999). This seems plausible where there is no clear separation between adjacent search  
625 locations, as has been the case in many non-human animal versions of the diagonal tube task,  
626 including Osthaus et al. (2003). However, we deliberately modified our apparatus from Osthaus  
627 et al.'s (2003) to provide clear separation between the three search locations (and make the setup  
628 more similar to previous child studies), yet dogs still showed a tendency to search the middle  
629 location. Another possible explanation for searching the middle location is that, if dogs have  
630 some notion of the correct search location, and also a (weak?) gravity bias, then their tendency to  
631 search the middle location might reflect a kind of naïve averaging of competing biases. We  
632 explore this option in Experiment 3, but first, in Experiment 2, we replicate Osthaus et al.'s

633 (2003) Experiment 3 with a larger sample, to re-visit how dogs' search shifts when searching the  
634 middle location is not an option. Will they be more successful at locating the reward, or will they  
635 be more likely to search the gravity location when the middle option is removed?

636 **Experiment 2: No middle search location**

637 In Experiments 1a-1c, in trials where dogs erred they were significantly more likely to  
638 search the middle location than the gravity location. This was also the case for dogs in Osthaus et  
639 al. (2003), and a tendency to search the middle location has also been reported for cotton-top  
640 tamarins (Hood et al., 1999) and marmosets (Hauser et al., 2001). This raises the possibility that,  
641 rather than a gravity bias, dogs (and possibly monkeys) have some sort of bias to search the  
642 middle location. Alternatively, perhaps several competing biases influence dogs' search behavior;  
643 it is possible that dogs do have a weak gravity bias, but that this is masked by a stronger bias to  
644 search the middle location. Therefore, of interest is how dogs re-distribute their search when the  
645 middle option is removed; i.e., is the tendency to search the middle masking an ability to solve  
646 the task, or potentially masking a gravity bias? Osthaus et al. (2003) tested 8 dogs with a  
647 comparable version of the task; in Experiment 2 we replicate this experiment with a larger  
648 sample of 16 dogs.

649 **Subjects.**

650 Sixteen dogs (10 male, 6 female; mean age =  $45 \pm 9$  months) participated in Experiment  
651 2 (Table S1). Two additional dogs were tested but excluded because they did not complete the  
652 test trials (1), or because the session was disrupted by outside noise (1).

653 **Materials.**

654 The apparatus used in Experiment 2 was identical to Experiment 1a, except for that the  
655 middle cup was not present during cup pre-training trials or test trials, so there were only two

656 possible search locations, both in the cup pre-training trials and the test trials (gravity and  
657 correct; Fig. 4a). The light freeze-dried liver treats were used and the electric fan was switched  
658 on to mask any residual sound.

### 659 **Results.**

660 On average, dogs required  $8.1 \pm 0.9$  trials (mean  $\pm$  SEM; range = 6 – 18; median = 6) to  
661 reach criterion in the cup pre-training trials. In the first test trial, 6/16 dogs searched the correct  
662 location, which did not differ from chance (chance:  $16/2 = 8$  dogs searching correctly; binomial  
663 test:  $p = 0.454$ ).

664 Across the 12 test trials, 63% of searches were directed to the gravity location (mean<sub>gravity</sub>  
665 =  $7.5 \pm 0.8$  trials), and 37% of searches were directed to the correct location (mean<sub>correct</sub> =  $4.5 \pm$   
666  $0.8$  trials; Fig. 5a). A mixed effects logistic regression model revealed that, while dogs as a group  
667 were more likely to search the gravity location than the correct location, this did not quite reach  
668 significance ( $z = -1.86$ ,  $p = 0.063$ ; Fig. 5a). Dogs' performance improved significantly across the  
669 session (trial log odds = 0.12,  $z = 2.39$ ,  $p = 0.017$ ; Fig. 4a; Fig. 5b; Table S2).

670 We also examined individual performance, and whether individual dogs had a preference  
671 for either of the search locations. One dog searched correctly significantly more often than  
672 expected by chance across twelve trials (11/12 trials correct,  $p = 0.006$ ); and 6/16 dogs exhibited  
673 a significant preference to search the gravity location (10/12 – 12/12 trials; binomial test:  $p <$   
674  $0.039$ , see Table S4 for apparatus configuration information for these dogs).

### 675 **Discussion.**

676 These results suggest that, when there is no middle location to search, dogs' tendency to  
677 search the middle location gets shifted to the gravity location. In contrast to our findings in this  
678 experiment, dogs tested with a comparable setup in Osthaus' et al's (2003) Experiment 3 seemed

679 to shift to searching the correct location, though only 8 dogs were tested so direct comparison of  
680 findings is challenging. Common marmosets on the other hand shifted to searching the gravity  
681 location when the middle cup was removed (Cacchione & Burkhardt, 2012). On the basis of our  
682 results in the present experiment, should we therefore conclude that dogs have a weak gravity  
683 bias that is masked by a stronger preference to search the middle location?

684         Our Experiment 2 results do indeed raise the possibility that gravity might influence  
685 dogs' search, at least in certain contexts. However, with the middle location removed, the gravity  
686 cup is quite clearly the most proximal of the two cups to the top of the tube—i.e. the location  
687 from which the reward is dropped (and therefore last seen by the dog). It is possible that with the  
688 middle cup removed, this proximity relationship becomes more salient, and thus becomes the  
689 key factor guiding dogs' search. This possibility is particularly important to explore with dogs,  
690 given that there is evidence that proximity to reward influences their choices in other physical  
691 problem-solving tasks (e.g. string-pulling, Osthaus et al., 2005). Indeed Hood et al. (1999)  
692 suggest that tamarins perhaps did not differentiate the gravity and middle locations, because both  
693 are closer to the reward's drop-off point than the correct location—thus implying a potential role  
694 for proximity. However, the role of proximity in the diagonal tubes task has to our knowledge  
695 never been explicitly tested. In Experiment 3 we de-confound gravity and proximity, with the  
696 aim of establishing whether our findings in Experiment 2 are due to dogs exhibiting a bias to  
697 search on the basis of gravity, or whether in fact proximity might be guiding their search.

### 698         **Experiment 3: Teasing apart the influence of gravity, proximity and middle biases**

699         To attempt to tease apart whether dogs' search in Experiment 2 was influenced by gravity  
700 or proximity, in Experiment 3 we pit gravity *against* proximity, by configuring the apparatus so  
701 that the gravity location is a greater distance from the top of the tube (where the reward is last

## DOGS' GRAVITY BIAS REVISITED

702 seen by the dog) than the correct location (see Fig. 4b and 4c). To our knowledge these versions  
703 of the diagonal tube task have not previously been presented to any species.

704 In Experiment 3a, an opaque tube was configured either top left-shelf middle, or top  
705 right-shelf middle (Fig. 4b). Because in this configuration, the correct, proximal location was  
706 also in the center of the apparatus, and we know from Experiment 1 that dogs tend to search the  
707 middle (although here the 'middle' location was on top of the shelf rather than the base of the  
708 apparatus, and was not in the middle in the sense of being the central of three cups), in  
709 Experiment 3b we presented dogs with a version of the task where the spatial relationships  
710 between the tube and the search locations were the same as in Experiment 3a, but the entire  
711 configuration was shifted, so that the correct search location was no longer in the center of the  
712 apparatus.

713 These two experiments together enable us to make a series of predictions regarding how  
714 dogs should perform, depending on the relative influence of different factors (gravity, proximity,  
715 middle) on their search behavior. First, if dogs' search is primarily influenced by gravity, then  
716 they should perform similarly poorly (below chance) in Experiments 3a and 3b, because the  
717 gravity location is incorrect in both cases. Second, if search is instead primarily guided by  
718 proximity, dogs should be equally successful (above chance) in Experiments 3a and 3b, because  
719 the most proximal location is the correct search location in both cases. Finally, if some sort of  
720 middle bias has the strongest influence on where dogs search, then performance should be better  
721 in Experiment 3a (where the correct location is in the center) than in Experiment 3b (where the  
722 incorrect/gravity location is in the center).

723 **Experiment 3a: Gravity vs. Proximity/Middle**724 **Subjects.**

725 Sixteen dogs (6 male, 10 female; mean age =  $51 \pm 9$  months) participated in Experiment  
726 3a (Table S1). Four additional dogs were tested but excluded because they did not reach criterion  
727 in the cup pre-training trials (1), or because they did not complete the cup pre-training trials (1)  
728 or the test trials (2).

729 **Materials.**

730 The configuration of the apparatus used in Experiment 3a is shown in Fig. 4b. There  
731 were two possible search locations both in the cup pre-training trials and the test trials: a gravity  
732 location that was either on the bottom left or right, and a correct location that was in the center,  
733 but on top of the mid-section of the frame, so that it was also the most proximal location to the  
734 starting point of the reward. The light freeze-dried liver treats were used and the electric fan was  
735 switched on to mask any residual sound.

736 **Results.**

737 On average, dogs required  $9.4 \pm 0.8$  trials (mean  $\pm$  SEM; range = 6 – 14; median = 9) to  
738 reach criterion in the cup pre-training trials. In the first test trial, 11/16 dogs searched the correct  
739 location, which did not differ from chance (chance: 8 dogs searching correctly; exact binomial  
740 test:  $p = 0.21$ ).

741 Across the 12 test trials, 73% of searches were directed to the correct location (mean<sub>correct</sub>  
742 =  $8.8 \pm 0.8$  trials; Fig. 5a), and just 27% of searches to the gravity location (mean<sub>gravity</sub> =  $3.3 \pm 0.8$   
743 trials). Dogs as a group searched the correct location significantly more often than expected by  
744 chance ( $z = 2.867$ ,  $p = 0.004$ ; Fig. 5a); that is, they were more likely to search the correct,  
745 proximal location than the gravity location. The mixed effects logistic regression model revealed

## DOGS' GRAVITY BIAS REVISITED

746 no change in performance across trials (trial log odds = 0.06,  $z = 1.00$ ,  $p = 0.316$ ; Fig. 4b; Fig.  
747 5b; Table S2).

748 We also examined individual performance, and whether individual dogs had a preference  
749 for either of the search locations. Half of the dogs (8/16) searched the correct/middle location  
750 significantly more often than expected by chance across twelve trials (10/12 – 12/12 trials  
751 correct; binomial test:  $p < 0.039$ , see Table S4 for apparatus configuration information for these  
752 dogs). Only one dog exhibited a significant preference to search the gravity location (12/12 trials,  
753  $p < 0.001$ , Table S3).

### 754 **Discussion.**

755 Based on the results of Experiment 3a, we can already eliminate the first option outlined  
756 above—that in Experiment 2, when the middle cup was removed, dogs' search was primarily  
757 influenced by gravity. If that were the case then dogs should have performed badly in this  
758 version of the task (i.e., they should have searched the gravity location), when in fact their  
759 performance was above chance. However, the results of this experiment alone cannot tell us  
760 whether dogs are searching on the basis of proximity—it is also possible that dogs in this  
761 experiment were searching the 'middle' location, in the sense that the correct cup was in the  
762 absolute center of the apparatus (on a horizontal plane). In Experiment 3b we aimed to establish  
763 whether dogs' search is more strongly influenced by proximity to the reward's starting point, or a  
764 preference for searching at the center of the apparatus.

765 **Experiment 3b: Gravity/Middle vs. Proximity**766 **Subjects.**

767 Sixteen dogs (7 male, 9 female; mean age =  $57 \pm 11$  months) participated in Experiment  
768 3b (Table S1). Four additional dogs were tested but excluded because they didn't reach criterion  
769 in the cup pre-training trials (2), or because they did not complete the test trials (2).

770 **Materials.**

771 The configuration of the apparatus used in Experiment 3b is shown in Fig. 4c. The  
772 configuration was the same as for Experiment 3a in terms of the spatial relationships between the  
773 search locations and the reward's starting point (i.e. there was a gravity location and a more  
774 proximal correct location), but the entire configuration was shifted within the frame of the  
775 apparatus, so that the gravity location was bottom middle, and the correct location on top of the  
776 mid-section of the frame was either on the left or the right. The light freeze-dried liver treats  
777 were used and the electric fan was switched on to mask any residual sound.

778 **Results.**

779 On average, dogs required  $10.75 \pm 1.24$  trials (mean  $\pm$  SEM; range = 6 – 21; median = 9)  
780 to reach criterion in the cup pre-training trials. In the first test trial, 4/16 dogs searched the  
781 correct location, which did not differ significantly from chance (chance: 8 dogs searching  
782 correctly; binomial test:  $p = 0.077$ ).

783 Across the 12 test trials, 35% of searches were directed to the correct location (mean<sub>correct</sub>  
784 =  $4.2 \pm 1.0$  trials; Fig. 5b), and 65% of searches to the gravity location (mean<sub>gravity</sub> =  $7.8 \pm 1.0$   
785 trials). Thus, while dogs tended to search incorrectly, overall performance did not quite reach  
786 significance ( $z = -1.895$ ,  $p = 0.058$ ; Fig. 5b). As in Experiment 3a, the mixed effects logistic

787 regression model revealed no change in performance across trials (trial log odds = 0.08,  $z = 1.50$ ,  
788  $p = 0.132$ ; Fig. 4c; Fig. 5b; Table S2).

789 We also examined individual performance, and whether individual dogs had a preference  
790 for either of the search locations. One dog searched correctly significantly more often than  
791 expected by chance across twelve trials (12/12 trials,  $p < 0.001$ ); and 8/16 dogs exhibited a  
792 significant preference to search the gravity/middle location (10/12 – 12/12 trials; binomial test:  $p$   
793  $< 0.039$ , Table S3, see Table S4 for apparatus configuration information for these dogs).

#### 794 **Discussion.**

795 This shift from above-chance performance in Experiment 3a to close-to-below-chance  
796 performance in Experiment 3b, despite the fact that the spatial relationship between the tube and  
797 the two search locations was the same in both cases, demonstrates that above all else, dogs'  
798 search is directed to the center of the apparatus. This result is in line with Osthaus et al.'s (2003)  
799 Experiment 4, which showed that dogs searched the 'gravity location' more often when it was  
800 located bottom-middle, and also searched correctly more often when the correct location was  
801 bottom-middle. This finding also enables us to rule out several previously posited explanations  
802 for why individuals tend to search the middle location in 3-cup versions of the diagonal tube  
803 task. First, it eliminates the possibility that dogs search the middle location because they confuse  
804 it spatially with the gravity location, as suggested by Hood et al., (1999), as in our Experiments  
805 3a and 3b the middle and gravity locations are clearly physically separated, both horizontally and  
806 vertically. Therefore, it seems infeasible that dogs could confuse the two locations spatially.  
807 Second, it also rules out the possibility that dogs are performing some sort of naïve averaging  
808 that leads them to search in the center, because here there are only two available search options.  
809 Finally, it also excludes the suggestion that individuals search the middle because the middle and

810 gravity locations are both closer to reward's dropping point than the correct location (Hood et al.,  
811 1999), as this was not true in our Experiment 3b, where the correct location was in closer  
812 proximity to the reward's dropping point than the middle location.

813 **Experiment 4: Does removing sources of bias reveal successful performance?**

814 While we have found no evidence of dogs exhibiting gravity-biased search, it appears  
815 likely that their performance in the diagonal tube task is limited by a preference to commence  
816 searching at the center of the apparatus. It is possible that contextually inappropriate responses  
817 elicited by the setup of the task (e.g., an inability to inhibit searching particular preferred  
818 locations) is masking dogs' physical-causal knowledge and ability to succeed at the task (e.g.,  
819 Gómez, 2005). Therefore, in our final experiment, we investigated how dogs would perform in a  
820 version of the diagonal tube task described in Gomez (2005), in which all potential sources of  
821 bias examined in the previous experiments are eliminated.

822 Specifically, in Experiment 4 there was no gravity location, no middle location, and no  
823 most proximal location because the two search locations were equidistant from the reward's  
824 starting point; i.e., there was no plausible physical reason to choose the distractor cup (Fig. 4d).  
825 According to Southgate and Gomez's unpublished data described in Gomez (2005), when  
826 presented with this version of the diagonal tube task, macaques were still unable to successfully  
827 locate the reward. We were interested in whether eliminating these potential sources of search-  
828 bias might either reveal understanding of the physical-causal structure of the task in dogs, or at  
829 least enable them to better attend to relevant cues (i.e. the location of the cup connected to the  
830 bottom of the tube).

831 **Subjects.**

832 16 dogs (10 male, 6 female; mean age =  $41 \pm 7$  months) participated in Experiment 4  
833 (Table S1). No dogs had to be excluded from this experiment.

834 **Materials.**

835 The tube was opaque and positioned either top middle-bottom right, or top middle-  
836 bottom left. As in Experiment 2, the middle cup was not present so there were only two possible  
837 search locations, both in the cup pre-training trials and the test trials (correct and incorrect). This  
838 meant that as well as being no middle location, there was also no gravity location. The light  
839 freeze-dried liver treats were used and the electric fan was switched on to mask any residual  
840 sound.

841 **Results.**

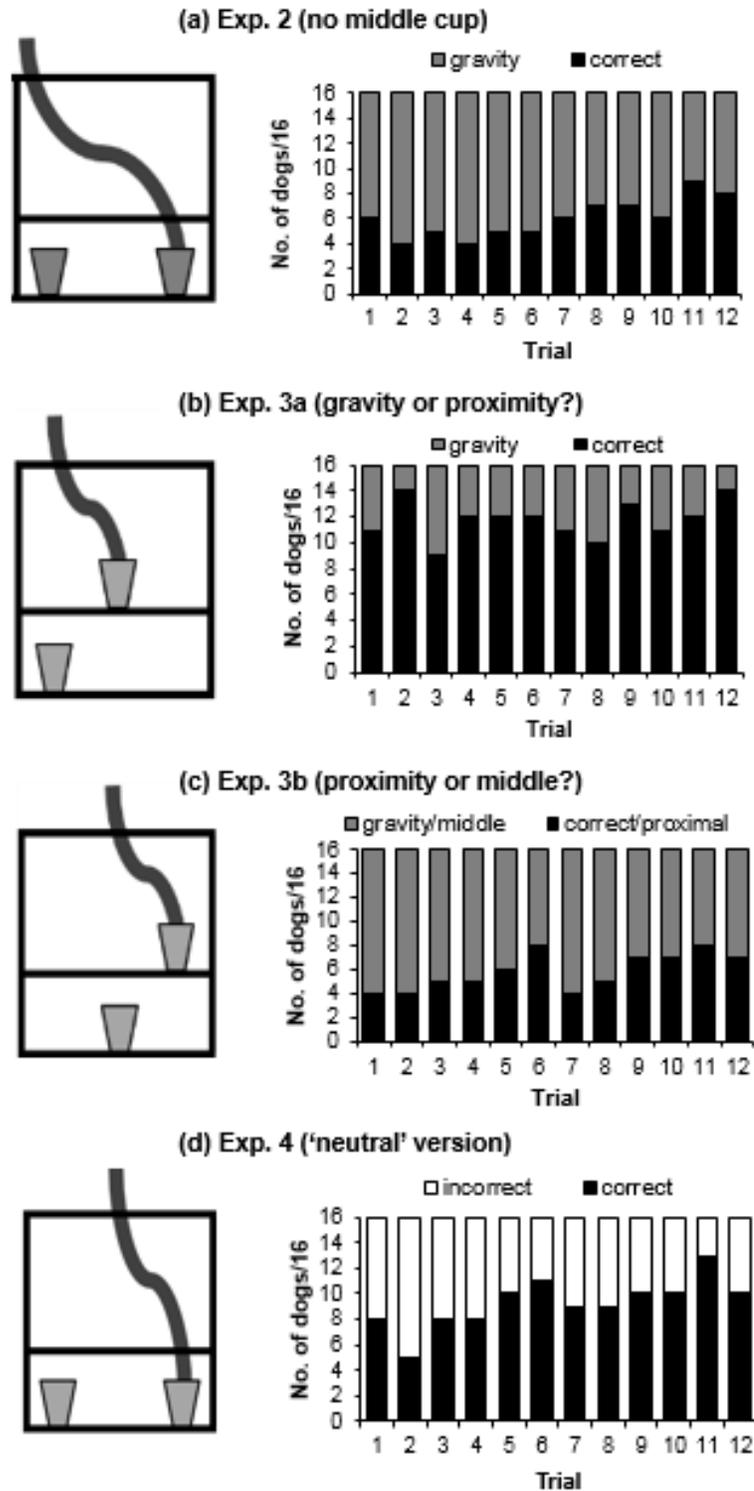
842 On average, dogs required  $7.1 \pm 0.4$  trials (mean  $\pm$  SEM; range = 6 – 10; median = 6) to  
843 reach criterion in the cup pre-training trials. Dogs did not show a bias for searching any  
844 particular location in their first trial; 8/16 dogs searched the correct location and 8/16 dogs  
845 searched the incorrect location (chance: 8 dogs searching correctly; binomial test:  $p = 1.00$ ).

846 Across the 12 test trials, 58% of searches were directed to the correct location (mean<sub>correct</sub>  
847 =  $6.9 \pm 1.0$  trials; Fig. 5a), and 42% of searches to the incorrect location (mean<sub>incorrect</sub> =  $5.1 \pm 1.0$   
848 trials). Dogs as a group failed to search the correct location significantly more often than  
849 expected by chance (mean<sub>correct</sub> =  $z = 1.08$ ,  $p = 0.28$ ; Fig. 5a), though according to the mixed  
850 effects logistic regression model, their performance improved significantly across the session  
851 (trial log odds = 0.17,  $z = 3.00$ ,  $p = 0.003$ ; Fig. 4d; Fig. 5b; Table S2).

852 We also examined individual performance, and whether individual dogs had a preference  
853 for either of the search locations. Six dogs searched correctly significantly more often than

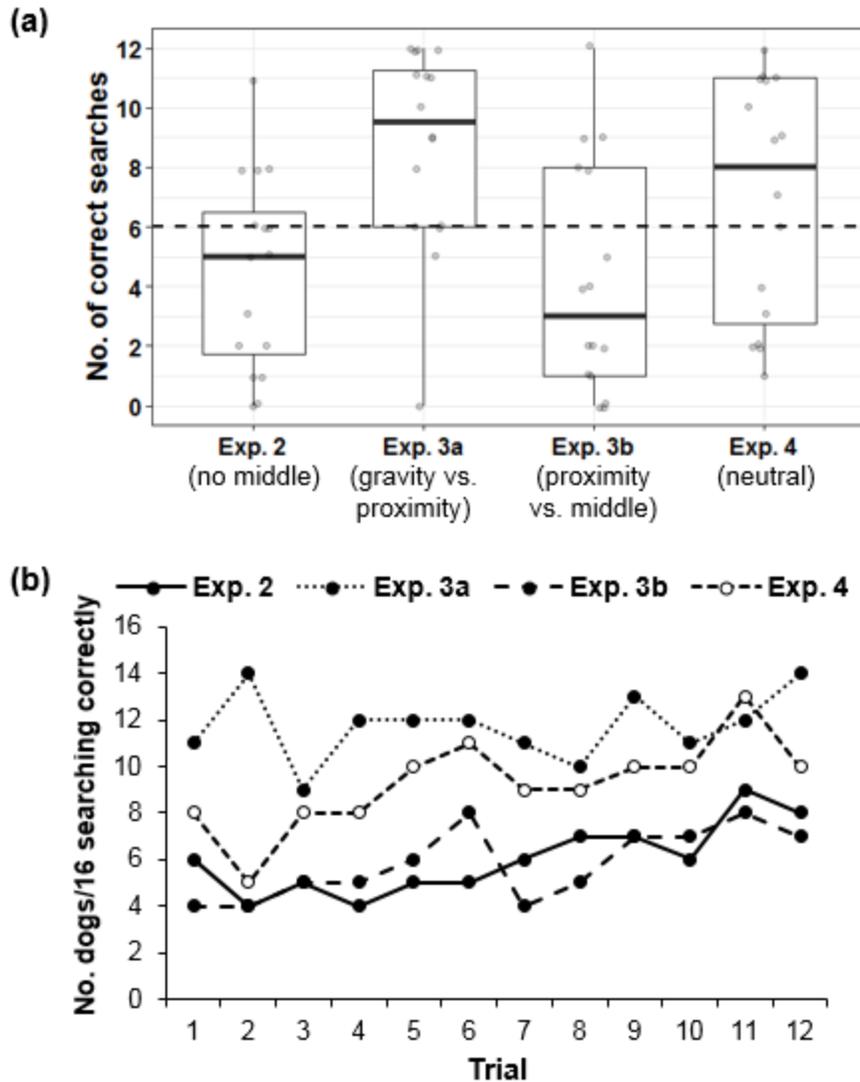
## DOGS' GRAVITY BIAS REVISITED

854 expected by chance across twelve trials (10/12 – 12/12 trials, binomial test:  $p < 0.039$ ). Four  
855 dogs exhibited a significant preference for the incorrect location (10/12-11/12 trials,  $p < 0.039$ ,  
856 Table S3, see Table S4 for apparatus configuration information for these dogs).  
857



858

859 **Figure 4.** Schematic representation of the configuration of the apparatus and the number of dogs searching  
 860 each location across trials 1-12 in Experiments 2-4 where there were always two search locations. Black  
 861 indicates correct cup; light-grey indicates middle cup; white indicates incorrect cup, which in Exp. 4 was  
 862 neither in the gravity nor middle location  
 863



864

865 **Figure 5.** (a) Box plot showing the number of searches directed at the correct location in Experiments 2  
 866 (no middle location), 3a (gravity vs. proximity), 3b (gravity vs. middle) and 4 ('neutral' version). The  
 867 dashed horizontal line represents the expected number of searches per location if search was random. (b)  
 868 Comparison of the number of dogs out of 16 that searched the correct location in each trial in Experiments  
 869 2, 3a, 3b and 4  
 870

871 **Discussion.**

872 By removing the gravity and central locations, and making both search options equally  
 873 proximal to the reward's dropping point we eliminated potential cues that could be influencing  
 874 dogs' search behavior. If search biases were masking dogs' actual knowledge of the physical-  
 875 causal structure of the task in previous experiments, then we would have expected them to

876 succeed here. This was not the case—although the majority of searches were directed to the  
877 correct location, overall performance was not better than chance. However, performance was  
878 improved relative to some of our other experiments (Fig. 5, see next section for model  
879 comparing these experiments), providing some evidence that eliminating sources of bias may  
880 have helped dogs to some extent; potentially by enabling them to focus on the relevant cue of the  
881 tube.

### 882 **Comparison of performance in Experiments 2 – 4 and interim discussion**

883 In Experiments 2–4, dogs were presented with versions of the diagonal tube task where a  
884 treat was dropped down a diagonal tube, and there were two possible search locations.  
885 Experiment 2 was a replication of Osthaus et al.'s (2003) Experiment 3 but with a larger sample  
886 of dogs, and Experiments 3a, 3b and 4 were novel variations of the diagonal tube task for dogs,  
887 designed to further probe what factors guide dogs' search, and explore how dogs perform when  
888 these potential sources of bias are eliminated from the testing setup.

889 We used mixed effects logistic regression that assumed a fixed slope across subjects to  
890 compare dogs' ability to search correctly between Experiments 2, 3a, 3b and 4 (Table S6). Dogs  
891 were significantly more likely to search the correct location when it was positioned in the middle  
892 of the apparatus and most proximal to the point where it was last seen (Exp. 3a), compared with  
893 in Experiment 2 where there was no middle cup (log odds = 2.23,  $z = 3.49$ ,  $p < 0.001$ ) and  
894 compared with Experiment 3b, when the incorrect/gravity location was in the middle (log odds =  
895 2.42,  $z = 3.58$ ,  $p < 0.001$ ; Fig. 5a and 5b). Dogs also performed better in Experiment 4 where  
896 potential sources of bias were eliminated than in Experiment 3b (log odds = 1.46,  $z = 2.21$ ,  $p =$   
897 0.027). There were no other significant differences between experiments in terms of dogs' ability

## DOGS' GRAVITY BIAS REVISITED

898 to search correctly, though there was a pattern of greater success in Experiment 4 compared to  
899 Experiment 2 (log odds = 1.27,  $z = 1.94$ ,  $p = 0.052$ ; Fig. 5b and 5c).

900         When the middle search location was removed (Experiment 2), rather than improving  
901 performance, dogs' search shifted to the gravity location, which suggested that in addition to  
902 having a preference to search the middle, search behavior might also be influenced (to a lesser  
903 extent) by gravity, or potentially proximity. In Experiment 3a, dogs were able to locate the  
904 reward significantly more often than expected by chance, which when considered in isolation,  
905 lent support to the idea that proximity, not gravity might be guiding dogs' search. However, when  
906 the same configuration was shifted within the frame of the apparatus so that the correct (still  
907 most proximal) location was on the left or right and the gravity location was now in the center of  
908 the apparatus (Experiment 3b), dogs no longer succeeded at locating the reward: again, they  
909 directed their search to the central location. Dogs' performance did not change across trials in  
910 either of these experiments—in Experiment 3a they performed consistently well and in  
911 Experiment 3b they performed consistently badly—reflecting their tendency to perseveratively  
912 search the middle location in both experiments. This finding for Experiment 3b in particular  
913 suggests that their preference to search in the center is difficult to overcome—even despite never  
914 being reinforced for searching centrally in Experiment 3b they continued to do so across repeated  
915 trials. Taken together, this suggests that when additional information regarding the reward's  
916 movement/the tube's mechanism is lacking, dogs default to searching in the center of the  
917 apparatus.

918         In Experiment 4, dogs' performance was significantly improved relative to Experiment  
919 3b but not better than chance. This suggests that eliminating potential sources of bias may go  
920 some way to improving dogs' search for a reward invisibly displaced down a diagonal tube, but

921 does not reveal successful performance, i.e., it is not the case that search biases are masking  
922 dogs' true knowledge of the physical-causal structure of the task (Gomez, 2005).

923 **General Discussion**

924 The tubes task has been used widely in the fields of cognitive development and  
925 comparative cognition to investigate children's and animals' physical reasoning abilities,  
926 specifically regarding their expectations about the influence of gravity on unsupported objects.  
927 By carefully manipulating the availability of perceptual cues (Experiments 1a-c) and the relative  
928 positions of various components of the apparatus (Experiments 2–4) we have revisited previous  
929 versions of the diagonal tube task and presented dogs with several novel versions of the task in  
930 an attempt to elucidate what factors really guide their search for a reward dropped down a  
931 diagonal tube.

932 Dogs as a group were generally unable to solve the diagonal tube task across 12 trials,  
933 though in most experiments their performance gradually improved over the course of the session  
934 suggesting that they would learn to succeed eventually, though likely via reinforcement rather  
935 than understanding anything about the physical-causal structure of the task. This is in keeping  
936 with the findings of Osthaus et al. (2003), who likewise reported that dogs were initially  
937 unsuccessful in the diagonal tube task, but learned to locate the reward across a limited number  
938 of trials. The results of Experiment 4—where we eliminated the potential for dogs to search on  
939 the basis of a gravity, middle or proximity bias—provide support for dogs' lack of causal  
940 understanding, because if it were the case that successful performance was being masked in other  
941 versions by an inability to suppress some search bias(es), dogs should have succeeded here.

942 A lack of ability to reason about the constraints imposed by the tube to locate hidden food  
943 fits with dogs' performance in other physical-causal reasoning tasks, where they have been

944 outperformed by great apes (Bräuer et al., 2006) and wolves (Lampe, Bräuer, Kaminski, &  
945 Virányi, 2017). Solving the diagonal tube task by reasoning about its physical-causal structure  
946 requires knowledge of object permanence, invisible displacement, object solidity and gravity, as  
947 well as the ability to elicit an appropriate search response (Tecwyn & Buchsbaum, in press).  
948 Although there is some evidence that dogs may possess an implicit understanding of object  
949 solidity based on looking-time experiments (Pattison, Miller, Rayburn-Reeves, & Zentall, 2010),  
950 studies that have investigated whether they can accurately search for invisibly displaced objects  
951 have proven inconclusive (e.g., Collier-Baker, Davis, & Suddendorf, 2004; Fiset & Leblanc,  
952 2007; Miller, Rayburn-Reeves, & Zentall, 2009), with dogs only compellingly passing specific  
953 simplified versions of invisible displacement tasks (e.g., Miller et al., 2009; Zentall & Pattison,  
954 2016).

955         While dogs generally failed to search correctly in the diagonal tubes task, their errors  
956 were not of the same nature as those observed in children. Specifically, we found no evidence  
957 that dogs exhibit a gravity bias in the diagonal tube task, either across trials or in trial 1. In fact,  
958 in all of the 3-cup versions of the task (Experiments 1a–c), when dogs searched incorrectly they  
959 were significantly more likely to search the middle location than the gravity location. Even in  
960 experiments where dogs did mainly search the gravity location (Exp. 2 where there was no  
961 middle cup, and Exp. 3b where the gravity cup was in the middle), the distribution of their  
962 searches did not differ from chance. It is possible that the incorporation of a tube familiarization  
963 phase in the present study could have diminished dogs' gravity bias relative to that reported by  
964 Osthaus and colleagues (e.g., our Experiment 1a: 6/16 trial 1 gravity searches; Osthaus et al.'s  
965 (2003) comparable Experiment 1 diagonal condition: 8/16 gravity searches). However, given that

## DOGS' GRAVITY BIAS REVISITED

966 overall our results generally replicated those of Osthaus et al., the tube familiarization appears  
967 not to have had a great impact on performance.

968         Why would dogs not exhibit a gravity bias? After all, they are subject to the same laws of  
969 physics as young children, and both species exist in a world where objects do typically fall  
970 straight down. Further, it seems likely that dogs have much experience of seeing objects (e.g.,  
971 food, balls) being dropped onto the ground. One possibility is that even if dogs are able to predict  
972 that a dropped object will fall straight down, their cognition is fundamentally different to that of  
973 humans and they do not form a naïve theory on the basis of this information. This could also  
974 explain the qualitative difference in the gravity bias seen in children versus some other  
975 primates—perhaps only humans form and reason on the basis of a naïve theory of gravity, which  
976 results in perseverative searching of the gravity location. Other species (e.g., cotton-top tamarins)  
977 might predict that an unsupported item will fall straight down, but because they have not formed  
978 a robust theory about this, searching of the gravity location rapidly decreases after trial 1.  
979 Relatedly, this prediction may not transfer to a situation where the object immediately moves out  
980 of sight (as is the case when it is dropped into an opaque tube). An alternative possibility is that  
981 human infants learn about the properties and behavior of objects, including the effect of gravity  
982 on objects, through their own actions—we are all familiar with toddlers in high chairs repeatedly  
983 throwing things onto the floor. Dogs' anatomy does not afford the same opportunity to act on  
984 objects and therefore limits the extent to which they are able to learn from observing the effects  
985 of their own actions on these objects. Presenting human infants who have not yet started  
986 manually interacting with objects with a either an eye-tracking or looking time version of the  
987 diagonal tube task could enable investigation of this; if repetitive experience of acting on objects  
988 is critical for the development of a gravity bias, then these infants should not expect the object to

989 end up in the gravity location. Work by Spelke and colleagues (1992) suggests showing 4-month-  
990 olds that an item dropped behind an occluder has remained suspended in midair does not appear  
991 to violate their expectations, thus lending support to the idea that young infants might not have  
992 an expectation that dropped objects will fall straight down to the ground.

993         Adapting looking-based measures with dogs would also enable the investigation of one  
994 further possibility: that dogs in fact *do* have a gravity bias (or, indeed, they are able to correctly  
995 predict where the reward will end up, as has been found for marmosets; Cacchione & Burkhart,  
996 2012) but this is not revealed by their search behavior. Dissociations between looking-based and  
997 action-based measures have been found for the tubes task and other physical reasoning tasks in  
998 non-human primates (e.g., Cacchione & Burkhart, 2012; Santos & Hauser, 2002) as well as  
999 young children (e.g., Lee & Kuhlmeier, 2013). Action-based versions of the tubes task pose  
1000 executive demands, as well as requiring individuals to use “feedforward logic-causal inferences”  
1001 (Cacchione & Rakoczy, 2017), so it is feasible that dogs might predictively look to the gravity  
1002 location (or the correct location), but then proceed to search elsewhere.

1003         In all of the experiments where a cup was positioned in the center of the apparatus (Exp.  
1004 1a–c; Exp. 3a–b), the majority of dogs' searches were directed to that location. Why might dogs  
1005 have a preference to search initially in the middle? A tendency to search the middle has been  
1006 observed previously in dogs (Osthaus et al., 2003), as well as in two different monkey species  
1007 (Cacchione & Burkhart, 2012; Hauser et al., 2001; Hood et al., 1999). However, while the  
1008 authors of these studies speculated about potential reasons for a tendency to search the middle  
1009 location (e.g., spatial confusion between the gravity and middle locations (Hood et al., 1999);  
1010 approximation of the reward's position, (Hauser et al., 2001); search the middle when uncertain  
1011 (Osthaus et al., 2003)), previous work did not explore these possibilities experimentally. We took

1012 this on in our Experiment 3, the results of which suggested that dogs have a preference to  
1013 commence their search at the center of the apparatus, as opposed to spatially confusing the  
1014 gravity and middle locations, or engaging in naïve averaging of competing search preferences.  
1015 The explanation offered by Osthaus et al. (2003) remains plausible—that when dogs are  
1016 uncertain of the reward's location, they commence searching at the center of the apparatus.  
1017 Future work could explore whether this strategy is specific to the diagonal tube task (e.g. related  
1018 to the constraints of the frame) or a more general strategy under conditions of uncertainty, by, for  
1019 example, hiding a reward in one of an array of cups and recording dogs' search behavior. If  
1020 searching in the middle reflects a general strategy, dogs should also commence searching  
1021 centrally in this context. Experiment 3 also allowed us to rule out the possibility that dogs' search  
1022 might be influenced by proximity to the last place the reward was seen, which has never  
1023 previously been explored in any species in the vertical version of the tubes task.

1024         The diagonal tube task has been used to study the gravity bias and physical reasoning  
1025 abilities in human children and a range of animal species, and so we chose to use this task here in  
1026 order to replicate and extend this previous work. However, given that the tube is a very specific  
1027 causal mechanism that is likely unfamiliar to dogs (and to animals more generally), future work  
1028 should explore dogs' physical reasoning abilities using more ecologically plausible paradigms.  
1029 While what is known about domestic dog physical cognition suggests that the species might have  
1030 relatively poor skills in this domain, physical and causal reasoning abilities have not been studied  
1031 in dogs to the same extent as in other taxa (e.g. primates, corvids, parrots), and some of the more  
1032 basic tasks that have been used to investigate intuitions about fundamental object properties such  
1033 as solidity and support in other species have been bypassed in favor of more complex designs  
1034 (e.g., Müller, Riemer, Range, & Huber, 2014). For example, a search-based version of the table

1035 or shelf task (Cacchione, Call, & Zingg, 2009; Hood et al., 2000; Spelke et al., 1992), suitably  
1036 adapted for dogs, could be an appropriate means to investigate dogs' knowledge of solidity, as  
1037 well as providing an additional paradigm with which to examine whether dogs' search might be  
1038 guided by gravity, as seems to be the case for macaques presented with this task (Hauser et al.,  
1039 2001).

1040         Finally, the fact that dogs' performance varied so much in our different versions of the  
1041 diagonal tube task setup should serve as an example of the value and importance of running  
1042 multiple experiments that carefully manipulate different factors that might influence behavior. If  
1043 we had only run Experiment 3a, we could have mistakenly concluded that dogs had a grasp of  
1044 the physical-causal structure of the task. If, on the other hand, we had only run Experiment 3b,  
1045 we could have—again mistakenly—concluded that dogs had a gravity bias. It is only when we  
1046 consider dogs' behavior across multiple experiments that a picture of what might really be  
1047 influencing their performance begins to emerge. As ever in animal cognition research, it is  
1048 critical to consider what other factors (in addition to the ones being investigated) might be  
1049 influencing behavior.

1050         In conclusion, across seven experiments we found no evidence that dogs spontaneously  
1051 grasp that the tube constrains the path of the reward and guides it to the cup attached to its  
1052 bottom end. However, our data also suggest that this failure is not primarily explained by a  
1053 gravity bias. Based on current evidence, it is possible that a gravity bias might be unique to some  
1054 primate species, or potentially (given the mixed evidence from non-human primate studies)  
1055 unique to young human children. To better understand the origins of the gravity bias and the  
1056 mechanisms underpinning it, additional groups should be tested with the diagonal tube task,  
1057 ideally using a developmental comparative approach in which evidence for a gravity bias is

1058 examined in immature and mature individuals, across species that differ with respect to their  
1059 causal knowledge and inhibitory control skills.

1060 **Acknowledgements**

1061 We thank Stephen Lea and an additional anonymous reviewer for their helpful feedback and  
1062 comments. Thanks to University College at the University of Toronto for providing our main  
1063 testing space, as well as All About Dogs, When Hounds Fly, and Good As Gold K9 School for  
1064 allowing us to use their facilities and helping to recruit participants. Several members of the U of  
1065 T Canine Cognition Lab assisted with data collection: Sam Clark, Nina Esmail, Julia Espinosa,  
1066 Aarushi Gupta, Sarah Marton-MacKay, Pingki Mazumder, Amanda Nickerson, and Madeline  
1067 Pelgrim. Madeline Pelgrim also performed the secondary coding of the data. Thanks to Dan  
1068 Goldwater for lending tools and Paul Coleman for help building the apparatus. We acknowledge  
1069 the support of the Natural Sciences and Engineering Research Council of Canada (NSERC),  
1070 [2016-05552], and of the Canada Foundation for Innovation.

## 1071 References

- 1072 Baillargeon, R. (2002). The acquisition of physical knowledge in infancy: A summary in eight  
1073 lessons. In U. Goswami (Ed.), *Blackwell Handbook of Childhood Cognitive Development*  
1074 (pp. 47–83). Blackwell Publishers Ltd. <https://doi.org/10.1002/9780470996652.ch3>
- 1075 Baker, S. T., Gjersoe, N. L., Sibielska-Woch, K., Leslie, A. M., & Hood, B. M. (2011). Inhibitory  
1076 control interacts with core knowledge in toddlers' manual search for an occluded object.  
1077 *Developmental Science*, 14(2), 270–279. <https://doi.org/10.1111/j.1467-7687.2010.00972.x>
- 1078 Bascandziev, I., & Harris, P. L. (2010). The role of testimony in young children's solution of a  
1079 gravity-driven invisible displacement task. *Cognitive Development*, 25(3), 233–246.  
1080 <https://doi.org/10.1016/j.cogdev.2010.06.002>
- 1081 Bascandziev, I., & Harris, P. L. (2011). Gravity is not the only ruler for falling events: Young  
1082 children stop making the gravity error after receiving additional perceptual information  
1083 about the tubes mechanism. *Journal of Experimental Child Psychology*, 109(4), 468–477.  
1084 <https://doi.org/10.1016/j.jecp.2011.03.010>
- 1085 Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models  
1086 using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01
- 1087 Bräuer, J., Kaminski, J., Riedel, J., Call, J., & Tomasello, M. (2006). Making inferences about  
1088 the location of hidden food: social dog, causal ape. *Journal of Comparative Psychology*,  
1089 120(1), 38–47. <https://doi.org/10.1037/0735-7036.120.1.38>
- 1090 Cacchione, T., & Burkart, J. M. (2012). Dissociation between seeing and acting: Insights from  
1091 common marmosets (*Callithrix jacchus*). *Behavioural Processes*, 89(1), 52–60.  
1092 <https://doi.org/10.1016/j.beproc.2011.10.010>

## DOGS' GRAVITY BIAS REVISITED

- 1093 Cacchione, T., & Call, J. (2010). Intuitions about gravity and solidity in great apes: the tubes  
1094 task. *Developmental Science*, 13(2), 320–330. [https://doi.org/10.1111/j.1467-](https://doi.org/10.1111/j.1467-7687.2009.00881.x)  
1095 [7687.2009.00881.x](https://doi.org/10.1111/j.1467-7687.2009.00881.x)
- 1096 Cacchione, T., Call, J., & Zingg, R. (2009). Gravity and solidity in four great ape species  
1097 (*Gorilla gorilla*, *Pongo pygmaeus*, *Pan troglodytes*, *Pan paniscus*): Vertical and horizontal  
1098 variations of the table task. *Journal of Comparative Psychology*, 123(2), 168–180.  
1099 <https://doi.org/10.1037/a0013580>
- 1100 Cacchione, T. & Rakoczy, H. (2017). *Comparative metaphysics: Thinking about objects in space*  
1101 *and time*. In J. Call (Ed.), *Handbook of Comparative Psychology*. (pp. 579-599). American  
1102 Psychological Association.
- 1103 Collier-Baker, E., Davis, J. M., & Suddendorf, T. (2004). Do dogs (*Canis familiaris*) understand  
1104 invisible displacement? *Journal of Comparative Psychology*, 118(4), 421-433.  
1105 <http://dx.doi.org/10.1037/0735-7036.118.4.421>.
- 1106 Fiset, S., & Leblanc, V. (2007). Invisible displacement understanding in domestic dogs (*Canis*  
1107 *familiaris*): the role of visual cues in search behavior. *Animal Cognition*, 10(2), 211–224.  
1108 <https://doi.org/10.1007/s10071-006-0060-5>
- 1109 Freeman, N. H., Hood, B. M., & Meehan, C. (2004). Young children who abandon error  
1110 behaviourally still have to free themselves mentally: a retrospective test for inhibition in  
1111 intuitive physics. *Developmental Science*, 7(3), 277–282. [https://doi.org/10.1111/j.1467-](https://doi.org/10.1111/j.1467-7687.2004.00346.x)  
1112 [7687.2004.00346.x](https://doi.org/10.1111/j.1467-7687.2004.00346.x)
- 1113 Gómez, J.-C. (2005). Species comparative studies and cognitive development. *Trends in*  
1114 *Cognitive Sciences*, 9(3), 118–125. <https://doi.org/10.1016/j.tics.2005.01.004>

## DOGS' GRAVITY BIAS REVISITED

- 1115 Hauser, M. D., Williams, T., Kralik, J. D., & Moskovitz, D. (2001). What guides a search for  
1116 food that has disappeared? Experiments on cotton-top tamarins (*Saguinus oedipus*).  
1117 *Journal of Comparative Psychology*, 115(2), 140–151. [https://doi.org/10.1037/0735-](https://doi.org/10.1037/0735-7036.115.2.140)  
1118 7036.115.2.140
- 1119 Hood, B. M. (1995). Gravity rules for 2- to 4-year olds? *Cognitive Development*, 10(4), 577–598.  
1120 [https://doi.org/10.1016/0885-2014\(95\)90027-6](https://doi.org/10.1016/0885-2014(95)90027-6)
- 1121 Hood, B. M. (1998). Gravity does rule for falling events. *Developmental Science*, 1(1), 59–63.  
1122 <https://doi.org/10.1111/1467-7687.00013>
- 1123 Hood, B. M., Hauser, M. D., Anderson, L., & Santos, L. (1999). Gravity biases in a non-human  
1124 primate? *Developmental Science*, 2(1), 35–41. <https://doi.org/10.1111/1467-7687.00051>
- 1125 Hood, B. M., Santos, L., & Fieselman, S. (2000). Two-year-olds' naïve predictions for horizontal  
1126 trajectories. *Developmental Science*, 3(3), 328–332. [https://doi.org/10.1111/1467-](https://doi.org/10.1111/1467-7687.00127)  
1127 7687.00127
- 1128 Hood, B. M., Wilson, A., & Dyson, S. (2006). The effect of divided attention on inhibiting the  
1129 gravity error. *Developmental Science*, 9(3), 303–308. [https://doi.org/10.1111/j.1467-](https://doi.org/10.1111/j.1467-7687.2006.00493.x)  
1130 7687.2006.00493.x
- 1131 Jaswal, V. K. (2010). Believing what you're told: Young children's trust in unexpected testimony  
1132 about the physical world. *Cognitive Psychology*, 61(3), 248–272.  
1133 <https://doi.org/10.1016/j.cogpsych.2010.06.002>
- 1134 Joh, A. S., Jaswal, V. K., & Keen, R. (2011). Imagining a way out of the gravity bias:  
1135 preschoolers can visualize the solution to a spatial problem. *Child Development*, 82(3),  
1136 744–750. <https://doi.org/10.1111/j.1467-8624.2011.01584.x>

## DOGS' GRAVITY BIAS REVISITED

- 1137 Joh, A. S., & Spivey, L. A. (2012). Colorful success: Preschoolers' use of perceptual color cues  
1138 to solve a spatial reasoning problem. *Journal of Experimental Child Psychology*, 113(4),  
1139 523–534. <https://doi.org/10.1016/j.jecp.2012.06.012>
- 1140 Kundery, S. M. A., Reyes, A. D. L., Taglang, C., Baruch, A., & German, R. (2010). Domesticated  
1141 dogs' (*Canis familiaris*) use of the solidity principle. *Animal Cognition*, 13(3), 497–505.  
1142 <https://doi.org/10.1007/s10071-009-0300-6>
- 1143 Lampe, M., Bräuer, J., Kaminski, J., & Virányi, Z. (2017). The effects of domestication and  
1144 ontogeny on cognition in dogs and wolves. *Scientific Reports*, 7(1), 11690.  
1145 <https://doi.org/10.1038/s41598-017-12055-6>
- 1146 Lee, V., & Kuhlmeier, V. A. (2013). Young children show a dissociation in looking and pointing  
1147 behavior in falling events. *Cognitive Development*, 28(1), 21–30.  
1148 <https://doi.org/10.1016/j.cogdev.2012.06.001>
- 1149 Miller, H. C., Rayburn-Reeves, R., & Zentall, T. R. (2009). What do dogs know about hidden  
1150 objects? *Behavioural Processes*, 81(3), 439–446.  
1151 <https://doi.org/10.1016/j.beproc.2009.03.018>
- 1152 Müller, C. A., Riemer, S., Range, F., & Huber, L. (2014). Dogs' use of the solidity principle:  
1153 revisited. *Animal Cognition*, 17(3), 821–825. <https://doi.org/10.1007/s10071-013-0709-9>
- 1154 Osthaus, B., Slater, A. M., & Lea, S. E. G. (2003). Can dogs defy gravity? A comparison with the  
1155 human infant and a non-human primate. *Developmental Science*, 6(5), 489–497.  
1156 <https://doi.org/10.1111/1467-7687.00306>
- 1157 Osthaus, B., Lea, S. E.G., & Slater, A. M. (2005). Dogs (*Canis lupus familiaris*) fail to show  
1158 understanding of means-end connections in a string-pulling task. *Animal Cognition*, 8(1),  
1159 37-47.

## DOGS' GRAVITY BIAS REVISITED

- 1160 Pattison, K. F., Miller, H. C., Rayburn-Reeves, R., & Zentall, T. (2010). The case of the  
1161 disappearing bone: dogs' understanding of the physical properties of objects. *Behavioural*  
1162 *Processes*, 85(3), 278–282. <https://doi.org/10.1016/j.beproc.2010.06.016>
- 1163 R Core Team (2017). R: A language and environment for statistical computing. R Foundation for  
1164 Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- 1165 Range, F., Möslinger, H., & Virányi, Z. (2012). Domestication has not affected the understanding  
1166 of means-end connections in dogs. *Animal Cognition*, 15(4), 597–607.  
1167 <https://doi.org/10.1007/s10071-012-0488-8>
- 1168 Santos, L. R., & Hauser, M. D. (2002). A non-human primate's understanding of solidity:  
1169 Dissociations between seeing and acting. *Developmental Science*, 5(2), F1-F7.  
1170 <https://doi.org/10.1111/1467-7687.t01-1-00216>
- 1171 Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge.  
1172 *Psychological Review*, 99(4), 605–632.
- 1173 Tecwyn, E. C., & Buchsbaum, D. (in press) Hood's gravity rules. In J. Vonk & T. Shackleford  
1174 (Eds). *Encyclopedia of Animal Cognition and Behavior*. Springer.
- 1175 Tomonaga, M., Imura, T., Mizuno, Y., & Tanaka, M. (2007). Gravity bias in young and adult  
1176 chimpanzees (*Pan troglodytes*): tests with a modified opaque-tubes task. *Developmental*  
1177 *Science*, 10(3), 411–421. <https://doi.org/10.1111/j.1467-7687.2007.00594.x>
- 1178 Zentall, T. R., & Pattison, K. F. (2016). Now you see it, now you don't: object permanence in  
1179 dogs. *Current Directions in Psychological Science*. 25(5), 357–362.  
1180 <https://doi.org/10.1177/0963721416664861>
- 1181