American Psychological Association

Journal of Comparative Psychology

0735-7036/19/$12.00

2019, Vol. 133, No. 1, 4–19
http://dx.doi.org/10.1037/com0000145

What Factors Really Influence Domestic Dogs' (*Canis familiaris*) Search for an Item Dropped Down a Diagonal Tube? The Tubes Task Revisited

Emma C. Tecwyn and Daphna Buchsbaum
University of Toronto

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, 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, 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 (Experiments 1a–c). Experiment 2 suggested that dogs might search on the basis of proximity, but Experiments 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 (Experiment 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

Supplemental materials: http://dx.doi.org/10.1037/com0000145.supp

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

The tubes task (Hood, 1995) has been used widely in the fields of cognitive development and comparative cognition to investigate children's and animals' physical reasoning abilities—specifically, how different-aged children and different species reason about the way objects behave under the influence of gravity (see Tecwyn & Buchsbaum, 2018, for a review). The version of the task typically used with children consists of three intertwined opaque tubes positioned vertically in a frame, each attached to a nonaligned cup at the base of the apparatus (Baker, Gjersoe, Sibielska-Woch, Leslie, & Hood, 2011; Bascandziev & Harris, 2010; Freeman, Hood, & Meehan, 2004; Hood, Wilson, & Dyson, 2006; Jaswal, 2010; Joh & Spivey, 2012). The task as typically used with animals is a simplified version of this and involves just a single diagonally configured tube (Figure 1). We will refer to this simplified version as the "diagonal tube task," and given the comparative focus of the current study, the majority of this introduction will focus on how individuals perform in this version of the task.

This article was published Online First November 1, 2018.

Emma C. Tecwyn and Daphna Buchsbaum, Department of Psychology, University of Toronto.

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

Correspondence concerning this article should be addressed to Emma C. Tecwyn, who is now at School of Psychology, Cardiff University, Tower Building, 70 Park Place, Cardiff CF10 3AT, United Kingdom. E-mail: tecwyne@cardiff.ac.uk

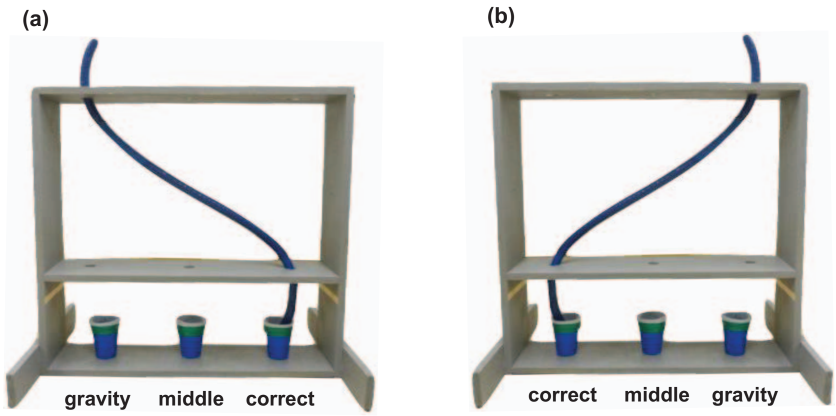

Figure 1. Diagonal tube task apparatus showing the opaque tube in a top left-bottom right configuration (a) and a top right-bottom left configuration (b). The gravity, middle, and correct search locations are indicated for each setup. See the online article for the color version of this figure.

Even in the diagonal tube task that involves only one tube there are typically three possible search locations at the bottom of the apparatus: the *correct* location connected to the bottom of the diagonal tube; the *gravity* location, which is aligned directly beneath the release point of the reward into the top of the tube; and the *middle* location, which is positioned between the other two locations (Figure 1). In accordance with the principles of gravity, solidity, and containment, when dropped into the top of the tube, the item travels (invisibly) down through the tube and ends up in the cup attached to its bottom end. Participants typically first undergo some pretraining to introduce the different search locations without the tube in place in the apparatus. The experimenter then puts the tube in place and drops an item (typically a ball for children and a food reward for animals) down the tube. The participant must then search for the item.

Although this is seemingly a straightforward task to solve, young children tend to perform poorly. Interestingly, when instructed to search for the dropped item in the diagonal tube task, 2-year-olds make a surprising, nonrandom error: Rather than searching in the correct location connected to the bottom of the single diagonal tube, they tend to search the gravity location (Figure 1). Furthermore, they do not search the gravity location only in their first trial when they are naïve to the task; they continue to do so across repeated trials, despite receiving feedback regarding the correct location, which remains fixed across trials (number of 2-year-olds searching gravity location in the diagonal tube task: Trial 1: nine out of 10; Trial 2: eight out of 10; Trial 3: nine out of 10; Experiment 4 pretest, Hood, 1995).

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

It has been suggested that children develop a naïve theory about gravity during their first year of life due to repeated exposure to objects falling straight down (Spelke, Breinlinger, Macomber, & Jacobson, 1992). Given that this "straight down" assumption is usually correct and therefore typically a useful heuristic to follow, this belief can be difficult to abandon (Hood et al., 2006), and may therefore interfere with children's ability to succeed at tasks that require the theory to be ignored (e.g., when an object is dropped down a diagonal tube). Further evidence that this theory is resistant to counterevidence is the fact that even after young children participate in a transparent version of the diagonal tube task—which they are able to pass—when subsequently retested with the opaque version they revert to searching the gravity location (Hood, 1995).

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

Several studies have explored how nonhuman species perform in the tubes task, with the aim of discovering whether the gravity bias is unique to humans, or whether it is also seen in other species, and could potentially represent an evolutionarily ancient naïve theory based on the physics of life on earth (Hood, Hauser, Anderson, & Santos, 1999). Cotton-top tamarins (*Sanguinus oedipus oedipus*) were the first nonhuman species to be tested, and the results of this study are the basis of claims that monkeys show a comparable gravity bias to young children. Although seven out of nine individuals searched the gravity location in their first trial, the bias did not compellingly persist across multiple trials—in Trial 2, only two out of nine individuals searched the gravity location (Hood et al., 1999), and so whether this meets the "challenging to overcome" criterion of a naïve theory is debatable. Three subjects succeeded at the task across 16 trials, and the errors made by the other six subjects were distributed evenly between the gravity and middle locations. Therefore, although this study provides evidence that tamarins' initial search may have been influenced by gravity, their behavior across trials does not bear the hallmarks of a naïve theory, given that an initial bias was easily overcome by several individuals, and errors were as likely to be directed at the middle location as the gravity location. Hood et al. (1999) also noted that several tamarins developed a preference to search the middle location, which they suggested was due to a lack of differentiation between the gravity and middle locations, both of which were in closer proximity to where the reward was dropped from than the correct location.

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

Another callitrichid species—common marmosets (*Callithrix jacchus*)—did not exhibit a gravity bias even in Trial 1 when they were naïve to the diagonal tube task (four out of seven individuals searched the gravity location; Cacchione & Burkart, 2012), but it is difficult to draw any firm conclusions from this small sample of individuals. In their first block of 16 trials marmosets' searches were randomly distributed between the three locations, but when they erred they were significantly more likely to search the gravity location than the middle location (though this difference disappeared in their second block of trials). Interestingly, when presented with a looking-time version of the task, marmosets looked significantly longer when the reward was revealed to have ended up in the gravity container than when it was revealed to have ended up in the correct container. Thus, although they were not able to search correctly at above-chance level, it is possible that they were implicitly aware of the role of the tube in constraining the reward's movement (Cacchione & Burkart, 2012; see Lee & Kuhlmeier, 2013, for similar findings with 2-year-old children).

A study by Cacchione and Call (2010) presented all four species of nonhuman great ape with the diagonal tube task, and found that they did not exhibit a gravity bias (only eight out of 22 subjects searched the gravity location in Trial 1)—in fact, they were able to locate the food item at above-chance levels from their first trial (Cacchione & Call, 2010). However, analysis of the errors made by the apes showed that they were significantly more likely to search the gravity location than the middle location, suggesting that apes may indeed hold naïve beliefs about gravity, but unlike 2-year-old children, they are usually able to suppress acting on the basis of this belief when it is inappropriate (as in the case of the diagonal tube). The findings of an earlier study by Tomonaga, Imura, Mizuno, and Tanaka (2007) fit with the idea that great apes might have a dormant gravity bias. Their task used a different measure to the other studies described here (prediction before the dropping event rather than search afterward) and involved two crossed tubes rather than a single diagonal tube. In this context, both juvenile and adult chimpanzees selected the gravity option at above-chance levels, and a further experiment ruled out that their choices were based on proximity. Although the apparent presence of a gravity bias in this study versus the lack of a reliable gravity bias in Cacchione and Call's (2010) task might be explained by the different response measures used, it is also possible that apes are able to solve the single diagonal tube task, but reveal a gravity bias when the task is more complex because more tubes are intertwined, which is known to increase children's preference for the gravity location (Hood, 1995; Lee & Kuhlmeier, 2013).

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

Only one study to date has investigated whether a nonprimate species exhibits gravity-biased search in the diagonal tube task. When domestic dogs (*Canis familiaris*) were presented with the diagonal tube task by Osthaus, Slater, and Lea (2003), they searched the correct location significantly less often than in a control task with a straight up and down tube (where the gravity location and correct location were the same). In the diagonal tube task, eight out of 16 dogs searched the gravity location in their first trial. Although dogs chose the gravity location more frequently than the two alternatives in Trial 1, their performance did not differ significantly from random search (two-tailed binomial test: 0.33 chance of searching gravity location; $p = .19$; not reported in the original article). The number of dogs searching the gravity location decreased rapidly across trials, and in Trial 16 only two out of 16 dogs made a gravity error. Concurrently, the number of dogs searching in the correct location increased across trials: from three out of 16 in Trial 1 to 10 out of 16 in Trial 16. Across all trials several dogs searched in the middle location (five out of 16 in Trial 1, and four out of 16 in Trial 16); in fact, overall, more than 40% of searches were directed at the middle location. The authors suggested that searching the middle might represent a strategy that dogs adopt when they are uncertain about the reward's location. In a follow-up experiment in which the middle location was removed as a search option (Experiment 3, Osthaus et al., 2003), six out of eight dogs searched the gravity location in their first trial, but they learned to search the correct location even more quickly than they did in the experiment in which the middle location was available as a search option.

Taken together, these data provide no evidence for a group-level gravity bias in dogs that persists across trials (i.e., that could constitute a naïve theory of gravity), and suggestive but nonsignificant evidence for a possible initial gravity bias present in Trial 1. However, the authors concluded that "Dogs, like toddlers and non-human primates, display a gravity bias," though they acknowledged that "dogs can learn to overcome this [gravity bias]" (p. 497, Osthaus et al., 2003). On the basis of the findings of this single study, several authors have gone on to report that dogs exhibit a persistent gravity bias (Bascandziev & Harris, 2011; Cacchione & Call, 2010; Joh et al., 2011; Kundey, De Los Reyes, Taglang, Baruch, & German, 2010; Range, Möslinger, & Virányi, 2012; Tomonaga et al., 2007). However, having examined the data presented in Osthaus et al. (2003), we do not believe there are grounds for such a strong conclusion. It is therefore puzzling that the claim that dogs having a robust and persistent gravity bias comparable with that of human toddlers is so pervasive in the literature.

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

Seeing as several dogs in Experiments 1a–c exhibited a tendency to search the middle location, as was the case in Osthaus et al.'s (2003) previous study with dogs, in Experiment 2 we replicated our Experiment 1a but with the middle location removed, to see how dogs' search shifted when searching the middle location was no longer an option. This also provided a replication of Osthaus et al.'s (2003) Experiment 3, but with a larger sample of dogs (16 vs. 8). On the basis of the results of our Experiment 2, in Experiments 3a and 3b we probed whether dogs' search might indeed be influenced by a gravity bias in some situations, or whether their behavior might in fact be better explained by proximity between the reward's release point and the search locations. These experiments represent novel versions of the diagonal tube task, as although the role of proximity has been indirectly explored in comparisons of the vertical tubes task with the horizontal tubes task (Hauser et al., 2001; Hood et al., 2000) and the version involving upward motion (Hood, 1998), to our knowledge it is the first time that the gravity location and most proximal location to where the reward is dropped from have been deconfounded in a vertical version of the single diagonal tube task in any species (though see Experiment 2 of Tomonaga et al., 2007, for a test of the proximity bias in a two-tube version of the task).

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

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

General Method

Subjects

All test subjects were pet dogs whose owners volunteered to participate in the study. Dog owners were recruited via e-mail, local advertisements, and local dog training facilities, and subsequently completed a questionnaire. To participate, dogs could not have a previous history of aggression toward humans and had to be in generally good health (including no known issues with their vision or hearing). There were no breed or age restrictions, though all dogs but one were at least 6 months old (see Table S1 in the online supplemental materials for further subject details including breed). Dogs participated either in the Canine Cognition Lab at the University of Toronto or in a similar sized space at a dog training facility in the Toronto area. Each dog only participated in one of the experiments.

Materials

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

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

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

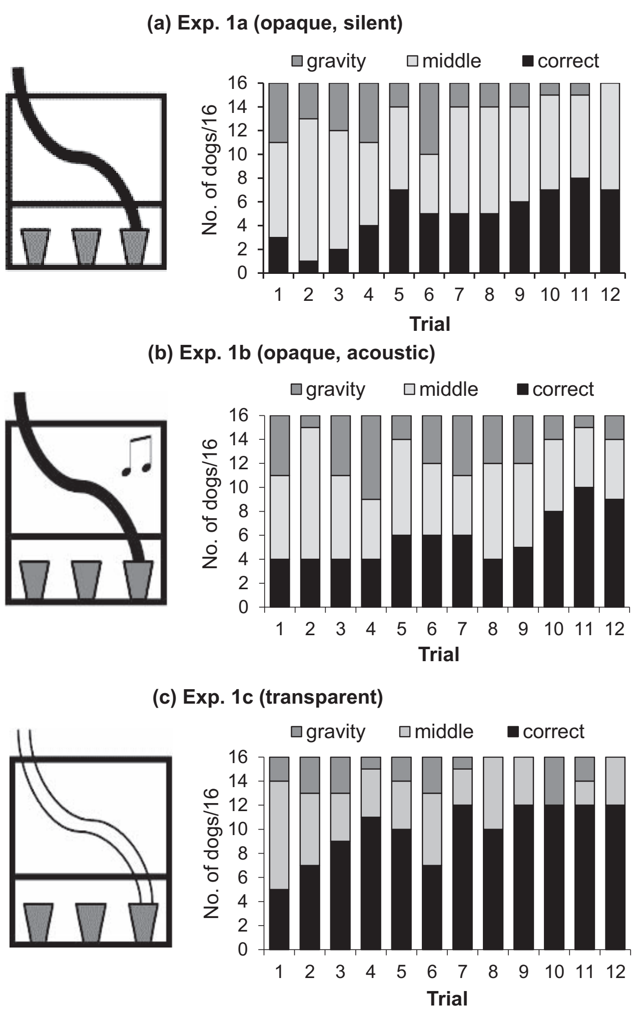

Figure 2. Schematic representation of the configuration of the apparatus and the number of dogs searching each location across Trials 1–12 in (a–c) Experiments 1a–c where there were always three search locations. Black indicates correct cup; light-gray indicates middle cup; midgray indicates gravity cup.

Procedure

Warm-up. The aim of the warm-up was to ensure that dogs felt comfortable in the testing space, and that they would interact with the cups to indicate their choice of search location during the experiment. Upon arrival in the testing area, dogs were given $\sim$5 min to explore the space off-leash while the owner completed an informed consent form. After this initial exploration period, dogs were introduced to the cups by the main experimenter, who placed one cup on the ground, then showed the dog a treat and dropped it into the cup and encouraged the dog to retrieve it by giving a release command ("OK!" unless the owner suggested an alternative). Some dogs spontaneously knocked the cup over and retrieved the treat; for dogs that touched the cup with their muzzle or paw but did not knock it over, the experimenter tipped the treat out for the dog to eat. This was repeated until the dog touched/knocked over the cup a total of three times. After this initial off-leash warm-up period, dogs were put on leash and handled by a second experimenter (handler). Owners were present during testing but were seated at the side of the room behind the dog's starting position (Fig. S1 in the online supplemental materials) and were asked not to interact with their dog during the experiment.

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

Cup pretraining trials. The aim of the cup pretraining trials was to introduce dogs to searching for treats in the cups while they were positioned in the apparatus, and to both measure and reduce the influence of any previous location biases during the test phase. The cups were in position in the bottom of the apparatus (the number and location of cups varied between experiments; see individual experiment sections and Figures 2 and 4 for details), and the tube was not present.

The main experimenter knelt behind the apparatus and the handler held the dog on leash in front of the apparatus at a distance of $\sim$160 cm. The experimenter showed the dog a treat, moved it back and forth above the frame midsection to ensure the dog was tracking it, then dropped it through one of the holes into the cup underneath (Fig. S2a in the online supplemental materials, Video S1 in the online supplemental materials). The experimenter then put her hands behind her back, stared at a fixed point on the wall behind the dog, then gave a release command (e.g., "OK!") and the dog was allowed to search exhaustively for the treat. A choice was defined as a dog making physical contact with a cup with their muzzle or paw (sniffing a cup, staring at a cup, or lying down in front of a cup did not constitute a choice). As in the warm-up, once the dog had touched the cup the experimenter tipped the treat out of the cup for the dog if necessary. Once a dog had chosen a cup it was removed by the experimenter. This dropping of treats into cups was repeated in a pseudorandom order (with the constraint that the treat was not dropped into the same location in more than two consecutive trials) until the dog successfully located the treat on their initial search on six consecutive trials (an equal number of times from each location). The individual cups were randomly interchanged between trials so the same cup did not always appear in the same location.

Following Osthaus et al. (2003), we set the maximum number of cup pretraining trials to 30; however, if a dog showed a persistent location/side bias (defined as 12 consecutive searches of the same location), we administered the following training: Treats were no longer dropped into the preferred cup, and pseudorandomly dropped into the other cups until dogs got six consecutive trials correct; then we reverted to all three cups and they had to get another six in a row correct. Therefore, in a few cases, the total number of trials including these training trials went above 30 (see individual experiment results for the range of pretraining trials required to reach criterion).

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

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

For each test trial, the experimenter knelt behind the apparatus and the handler held the dog on leash in front of the apparatus at a distance of ~160 cm. The experimenter showed the dog a treat, moved it back and forth across the top of the frame until the dog tracked it, then dropped it into the top of the tube, showed the dog her empty hand, placed her hands on her lap, stared straight ahead at a fixed point behind the dog, then gave the dog a release command (e.g., "OK!") to search for the treat (see Video S2 in the online supplemental materials). The dog was allowed to search exhaustively until they located the treat. We allowed exhaustive search to match previous studies with dogs (Osthaus et al., 2003), apes (Cacchione & Call, 2010), and monkeys (Hood et al., 1999). Pilot work also revealed that dogs quickly stopped participating (they refused to search) if they were only allowed to search one location and chose incorrectly, which meant they were not rewarded for that trial. Although one might imagine that allowing exhaustive search reduces the incentive for the subject to make an initial correct choice (because they ultimately get a reward anyway), previous work has demonstrated that this is not the case for monkeys at least: In two tubes task studies, performance did not differ according to whether subjects were allowed to search exhaustively, or only allowed to search a single location and therefore went unrewarded if they chose incorrectly (Hauser et al., 2001).

Data coding and analysis. In both the cup pretraining trials and the test trials, we scored the location that dogs searched first. We coded searches as correct or incorrect. For the three-cup versions of the task (Experiments 1a–1c), if dogs searched incorrectly then their search was further coded as directed at the middle or gravity location. To investigate performance in Trial 1 of each experiment, we used chi-square goodness of fit tests (three-cup versions) or binomial tests (two-cup versions) to see whether the distribution of dogs' search differed from random. We used mixed effects logistic regression models that assumed a fixed slope across subjects (including a random slope term did not significantly improve fit for any experiment) to examine successful performance across all trials in each individual experiment, and to look for change in performance over trials (lme4 package Version 1.1.13; Bates, Mächler, Bolker, & Walker, 2015) in the R environment (Version 3.3.3; R Development Core Team, 2017). We used the same approach to examine the nature of dogs' errors, and look for any change in the nature of errors made across trials. We also compared overall performance in each experiment to chance. For the three-cup versions of the task (Experiments 1a–1c), we adjusted the intercept to account for testing against 0.33 (as opposed to the standard 0.5 in the two-cup versions) and used that to calculate an adjusted z statistic and obtain the correct p value. We used binomial tests to examine the performance of individual dogs. For all three-cup versions (Exp. 1a–1c), where we were interested in seeing whether individual dogs either searched correctly, in the middle, or at the gravity location significantly more often than expected by chance, we used a Bonferroni correction for multiple comparisons, so for these tests α was 0.017 (0.05/3). Finally, mixed effects logistic regression models were also used to compare performance between-subjects across all three-cup versions (Exp. 1a–1c) and all two-cup versions (Exp. 2–4) of the task. As for the individual experiment analyses, we assumed a fixed slope across subjects because including a random slope term did not significantly improve fit for either comparison. All tests were two-tailed and α was 0.05.

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

Experiment 1: Replicating Previous Versions of Diagonal Tube Task

Experiment 1a: The Classic Diagonal Tube Task

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

Subjects. Sixteen dogs (four male, 12 female; mean age = 66 ± 10 months) participated in Experiment 1a (Table S1 in the online supplemental materials). Three additional dogs were tested but excluded because they did not reach criterion in the cup pretraining trials (one), or because they failed to complete the test trials (two).

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

Results. On average, dogs required 12.2 ± 2.3 trials (mean ± SEM; range = 6–44; median = 9.5) to reach criterion in the cup pretraining trials. In their first test trial dogs did not show a bias to search any particular location (chance: 16/3 = 5.33 dogs searching correctly; χ^2 = 2.38, df = 2, p = .304); rather, they searched randomly for a treat that traveled invisibly and inaudibly down a diagonal tube. Half of the dogs (eight out of 16) searched the middle location, five searched the gravity location, and three searched the correct location (Figure 2a). Therefore, we found no evidence that dogs' search behavior was guided by gravity when they were naïve to the task.

Across the 12 test trials, 52% of searches were directed to the middle location ($\text{mean}_{\text{middle}}$ = 6.2 ± 0.7 trials), 32% of searches were directed to the correct location ($\text{mean}_{\text{correct}}$ = 3.8 ± 0.7 trials), and just 16% of searches were directed to the gravity location ($\text{mean}_{\text{gravity}}$ = 2.1 ± 0.6 trials; Figure 3a). A mixed effects logistic regression model revealed that, as a group, dogs' tendency to search the correct location did not differ significantly from the 33% expected by chance (z = 0.61, p = .54).

The mixed effects logistic regression model revealed a significant improvement in performance across trials (Figures 2a and 3b; trial log odds = 0.17, z = 3.192, p = .001, Table S2 in the online supplemental materials). We ran a separate mixed effects logistic regression model to examine the search errors that dogs made, which revealed that, across the 12 test trials, on trials where dogs erred they were significantly more likely to search the middle location than the gravity location (z = 3.71, $p < .001$; Figure 3a). Across trials, the number of gravity searches decreased significantly relative to the number of middle searches (trial log odds = −0.16, z = −2.27, p = .022; Figure 2a).

We also examined individual performance, and whether individual dogs had a preference for any of the search locations. After correcting for multiple comparisons (Bonferroni correction; α = .017; a dog had to search the same location in at least nine out of 12 trials to produce a p value <0.017 in a binomial test), only one out of 16 dogs searched correctly significantly more often than expected by chance across the 12 test trials (10 out of 12 trials correct; $p < .001$). Two out of 16 dogs had a significant middle-location preference (10 out of 12 and 11 out of 12 trials; $p < .001$), and no dogs exhibited a significant preference for searching the gravity location (maximum number of gravity searches = eight out of 12; Table S3 in the online supplemental materials).

Discussion. When no perceptual cues were available, like children (Hood, 1995), monkeys (Hood et al., 1999; Hauser et al., 2001), and dogs (Osthaus et al., 2003) tested previously, dogs as a group failed to locate a reward dropped down a diagonal tube, either in Trial 1 or across 12 trials. However, there was no evidence that they searched on the basis of a gravity bias—when dogs erred, they were significantly more likely to search the middle location than the gravity location. In this respect their performance differed from that of tamarins, who exhibited a Trial 1 gravity bias (Hood et al., 1999), and children, who seemed to show a gravity bias that was difficult to overcome, even after several repeated trials (Hood, 1995). The performance of dogs in the current experiment also differed from that of great apes, who were able successfully locate the reward at above chance levels within nine trials (Cacchione & Call, 2010). Apes have previously demonstrated superior physical–causal reasoning skills compared with dogs (Bräuer, Kaminski, Riedel, Call, & Tomasello, 2006), so it is possible that they were better able to understand the role of the tube in constraining the path of the reward. In the following two experiments, we explored whether highlighting the tube's physical–causal mechanism by making the reward's passage through the tube audible (Experiment 1b) or visible (Experiment 1c) would improve dogs' performance in the diagonal tube task.

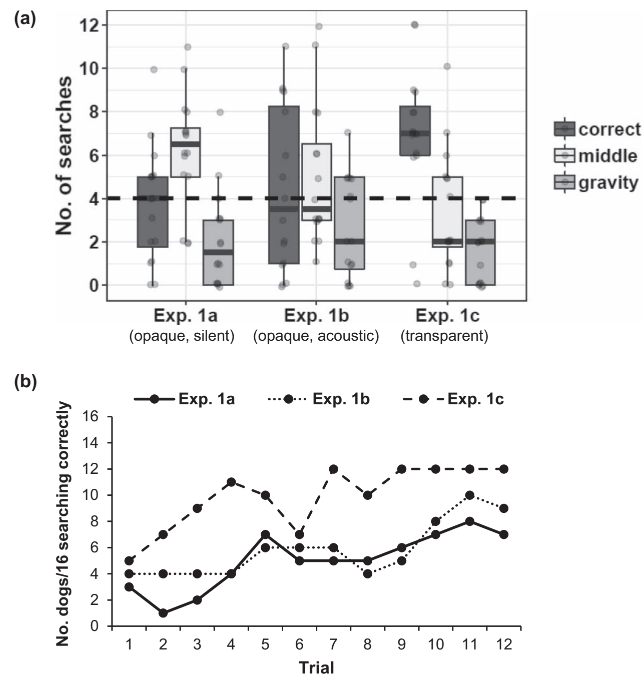

Figure 3. (a) Box plot showing the distribution of dogs' searches of the correct, middle, and gravity locations in Experiments 1a (opaque, silent), 1b (opaque, acoustic), and 1c (transparent, silent). The dashed horizontal line represents the expected number of searches per location if search was random.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

Experiment 1b: Auditory Cues Available

In Experiment 1b, we investigated whether being able to hear the reward traveling through the tube (but still not hear it landing in the cup) would enable dogs to perform better in the task, either because they could acoustically track the reward traveling through the tube, or because the sound provided some information regard-

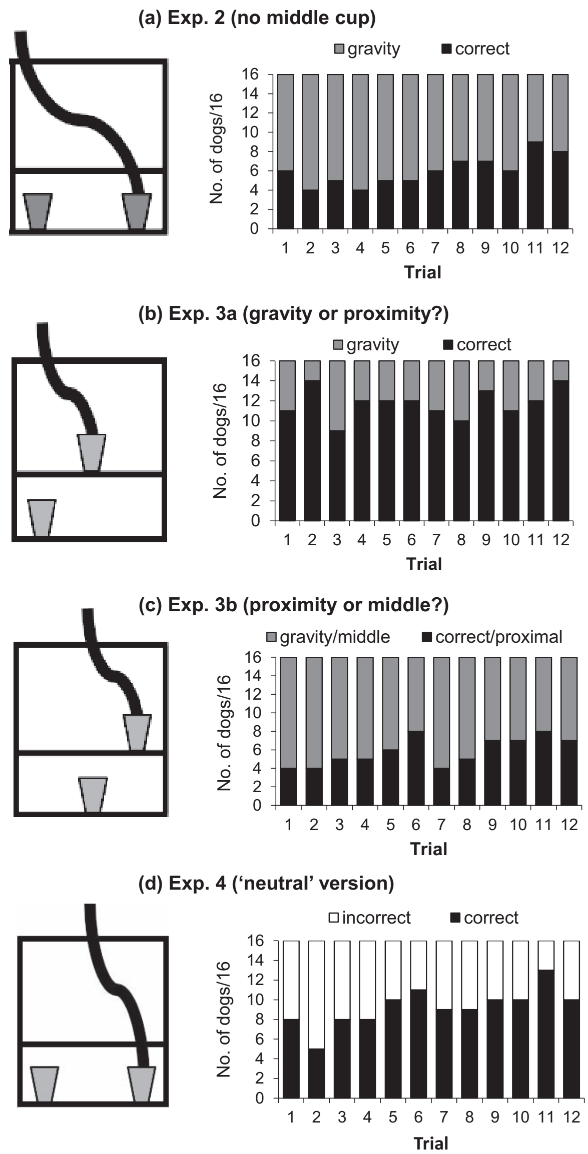

Figure 4. Schematic representation of the configuration of the apparatus and the number of dogs searching each location across Trials 1–12 in (a–d) Experiments 2–4 where there were always two search locations. Black indicates correct cup; light-gray indicates middle cup; white indicates incorrect cup, which in Exp. 4 was neither in the gravity nor middle location.

ing the tube's causal mechanism. Great apes tested with a comparable version searched randomly in their first trial, but were able to successfully locate the reward at above chance levels within their first block of nine trials (Cacchione & Call, 2010). Dogs have not previously been presented with this version of the diagonal tube task.

Subjects. Sixteen dogs (10 male, six female; mean age = 42 ± 9 months) participated in Experiment 1b (Table S1 in the online supplemental materials). Three additional dogs were tested but excluded because they did not complete the cup pretraining trials (one), or because their choice of search location was unclear in the test trials (two).

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

Results. On average, dogs required 12.9 ± 2.6 trials (mean ± SEM; range = 6–35; median = 6) to reach criterion in the cup pretraining trials. In the first test trial, dogs did not show a bias for searching any particular location ($\chi^2 = 0.88$, $df = 2$, $p = .644$; middle: seven out of 16; gravity: five out of 16; correct: four out of 16; Figure 2b).

Across the 12 test trials, 42% of searches were directed to the middle location ($\text{mean}_{\text{middle}} = 5.0 \pm 0.8$ trials), 36% of searches were directed to the correct location ($\text{mean}_{\text{correct}} = 4.4 \pm 0.1$ trials), and just 22% of searches to the gravity location ($\text{mean}_{\text{gravity}} = 2.6 \pm 0.6$ trials; Figure 3b). The mixed effects logistic regression model revealed that, as a group, dogs' tendency to search the correct location did not differ from the 33% expected by chance ($z = -0.41$, $p = .68$; Figure 3a). However, their performance improved significantly across trials (trial log odds = 0.21, $z = 3.60$, $p < .001$; Figures 2b and 3b; Table S2 in the online supplemental materials). Examining the search errors that dogs made revealed that, as in Experiment 1a, dogs were significantly more likely to search the middle location than the gravity location ($z = 2.42$, $p = .015$; Figure 3a). Across trials, the number of gravity searches relative to the number of middle searches did not change significantly (trial log odds = −0.03, $z = -0.48$, $p = .632$; Figure 2b).

We also examined individual performance, and whether individual dogs had a preference for any of the search locations. After correcting for multiple comparisons (Bonferroni correction; $\alpha = .017$; a dog had to search the same location in at least nine out of 12 trials to produce a p value <0.017 in a binomial test), four dogs searched correctly significantly more often than expected by chance across 12 trials (nine out of 12 to 11 out of 12 trials correct; $p < .012$, see Table S4 in the online supplemental materials for apparatus configuration information for these dogs). Two dogs showed a significant middle-location preference (11 out of 12 and 12 out of 12 trials; $p < .001$), and as in Experiment 1a, no individual dogs exhibited a significant gravity location preference (maximum number of gravity searches = seven out of 12; Table S3 in the online supplemental materials).

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

Experiment 1c: Transparent Tube

Hood (1995) found that if the tube was translucent so that it was possible to observe the movement of the object dropped down it, then 2.5-year-old children were able to successfully locate the item. In Experiment 1c, we investigated whether dogs—who have not previously been tested with a transparent version of the diagonal tubes task—would be able to solve the diagonal tube task if they were able to *see* the reward moving through the tube, that is, when the reward was *visibly* displaced by gravity and constrained by the tube. This is important because if dogs do not succeed in this version of the task, this might suggest that there are other task demands limiting their performance. For example, an inability to search correctly in a transparent version could be due to lack of motivation, some physical constraint of the apparatus (e.g., dogs avoid searching the cup with the tube attached because it is harder to access), or an object permanence/working memory failure, such that once the object is out of sight dogs are completely unable to reason about its location (though success in the cup pretraining trials makes this unlikely).

Subjects. Sixteen dogs (nine male, seven female; mean age = 67 ± 9 months) participated in Experiment 1c (Table S1 in the online supplemental materials). Five additional dogs were tested but excluded because they did not reach criterion in the cup pretraining (two), they did not complete the cup pretraining trials (one), they did not complete the test trials (one), or because their choice of search location was unclear in the test trials (one).

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

Results. On average, dogs required 10.4 ± 1.1 trials (mean ± SEM; range = 6–19; median = 9.5) to reach criterion in the cup pretraining trials. As in Experiments 1a and 1b, dogs did not show a bias for searching any particular location in their first trial ($\chi^2 = 4.63$, $df = 2$, $p = .099$), though again, more dogs searched the middle location (nine out of 16) than the gravity location (two out of 16) or the correct location (five out of 16).

Across the 12 test trials, 58% of searches were directed to the correct location ($\text{mean}_{\text{correct}} = 7.0 \pm 0.8$ trials), 28% of searches were directed to the middle location ($\text{mean}_{\text{middle}} = 3.4 \pm 0.7$ trials), and just 14% of searches to the gravity location ($\text{mean}_{\text{gravity}} = 1.6 \pm 0.4$ trials; Figure 2a). A mixed effects logistic regression model revealed that, as a group, dogs searched the correct location more often than the 33% expected by chance ($z = 3.14$, $p = .002$; Figure 2a). The mixed effects logistic regression model revealed a significant effect of trial on performance (log odds = 0.18; $z = 3.59$, $p < .001$), so, as in Experiments 1a and 1b, dogs were more likely to search correctly across trials (Figures 2c and 3b; Table S2 in the online supplemental materials). When dogs made search errors, as in Experiments 1a and 1b, they were more likely to be directed to the middle location than the gravity location ($z = 2.245$, $p = .025$; Figure 3a). Across trials, the number of gravity searches decreased relative to the number of middle searches, but not significantly so (trial log odds = −0.11, $z = -1.73$, $p = .08$; Figure 2c).

We also examined individual performance, and whether individual dogs had a preference for any of the search locations. After correcting for multiple comparisons (Bonferroni correction; $\alpha = .017$; a dog had to search the same location in at least nine trials to produce a p value <0.017 in a binomial test), four out of 16 dogs searched correctly significantly more often than expected by chance across 12 trials (nine out of 12 to 12 out of 12 trials correct; $p < .012$, see Table S4 in the online supplemental materials for apparatus configuration information for these dogs), one out of 16 dogs had a significant middle-location preference (10 out of 12 trials; $p < .001$), and no dogs exhibited a significant gravity location preference (maximum number of gravity searches = four out of 12; Table S3 in the online supplemental materials).

Discussion. When the displacement of the reward through the tube was visible, dogs, like 2-year-old children (Hood, 1995) tended to succeed at searching correctly for it across the 12 test trials (though dogs did still make errors, especially in early trials). Like dogs, cotton-top tamarins that participated in a transparent tube version of the task searched randomly in Trial 1, but only two out of five tamarins performed above chance across 10 trials (Hauser et al., 2001). Importantly, Exp. 1c shows that solving the diagonal tube task is within the capabilities of dogs if they have sufficient perceptual information, that is, poor performance in opaque versions is not due to a lack of motivation, physical constraints imposed by the apparatus, or a working memory/object permanence failure.

Comparison of Performance in Experiments 1a–1c and Interim Discussion

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

Mixed effects logistic regression that assumed a fixed slope across subjects was used to compare dogs' ability to search correctly between Experiments 1a, 1b, and 1c (Table S5 in the online supplemental materials). Dogs were significantly more likely to search the correct location when the tube was transparent (Exp. 1c), compared with when no perceptual cues were available (log odds = 1.47, $z = 2.52$, $p = .012$), as well as when only acoustic cues were available (log odds = 1.31, $z = 2.23$, $p = .026$; Figure 3b). Performance did not differ between Exp. 1a and Exp. 1b (log odds = 0.16, $z = 0.27$, $p = .790$; Figure 3b). Thus, being able to see the reward's trajectory helps dogs to identify its end location, but being able to hear it traveling through the tube does not.

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

In Experiments 1a and 1b, the middle cup was the most searched location and several individual dogs showed a significant middle-bias. Even in Experiment 1c where dogs succeeded at locating the reward overall, the middle was the second most common choice across trials. This preference for commencing searching in the middle was noted of the dogs in the diagonal tube condition of Osthaus et al.'s (2003) Experiment 1, and has also been recorded for some cotton-top tamarins (Hauser et al., 2001; Hood et al., 1999) and common marmosets (Cacchione & Burkart, 2012). It has previously been suggested that this tendency might be due to subjects confusing the middle location with the gravity location due to their adjacent spatial proximity (Hood et al., 1999). This seems plausible where there is no clear separation between adjacent search locations, as has been the case in many nonhuman animal versions of the diagonal tube task, including Osthaus et al. (2003). However, we deliberately modified our apparatus from Osthaus et al.'s (2003) to provide clear separation between the three search locations (and make the setup more similar to previous child studies), yet dogs still showed a tendency to search the middle location. Another possible explanation for searching the middle location is that, if dogs have some notion of the correct search location, and also a (weak?) gravity bias, then their tendency to search the middle location might reflect a kind of naïve averaging of competing biases. We explore this option in Experiment 3, but first, in Experiment 2, we replicate Osthaus et al.'s (2003) Experiment 3 with a larger sample, to revisit how dogs' search shifts when searching the middle location is not an option. Will they be more successful at locating the reward, or will they be more likely to search the gravity location when the middle option is removed?

Experiment 2: No Middle Search Location

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

Subjects

Sixteen dogs (10 male, six female; mean age = 45 ± 9 months) participated in Experiment 2 (Table S1 in the online supplemental materials). Two additional dogs were tested but excluded because they did not complete the test trials (one), or because the session was disrupted by outside noise (one).

Materials

The apparatus used in Experiment 2 was identical to Experiment 1a, except for that the middle cup was not present during cup pretraining trials or test trials, so there were only two possible search locations, both in the cup pretraining trials and the test trials (gravity and correct; Figure 4a). The light freeze-dried liver treats were used and the electric fan was switched on to mask any residual sound.

Results

On average, dogs required 8.1 ± 0.9 trials (mean ± SEM; range = 6–18; median = 6) to reach criterion in the cup pretraining trials. In the first test trial, six out of 16 dogs searched the correct location, which did not differ from chance (chance: 16/2 = 8 dogs searching correctly; binomial test: $p = .454$).

Across the 12 test trials, 63% of searches were directed to the gravity location ($\text{mean}_{\text{gravity}} = 7.5 \pm 0.8$ trials), and 37% of searches were directed to the correct location ($\text{mean}_{\text{correct}} = 4.5 \pm 0.8$ trials; Figure 5a). A mixed effects logistic regression model revealed that, although dogs as a group were more likely to search the gravity location than the correct location, this did not quite reach significance ($z = -1.86$, $p = .063$; Figure 5a). Dogs'

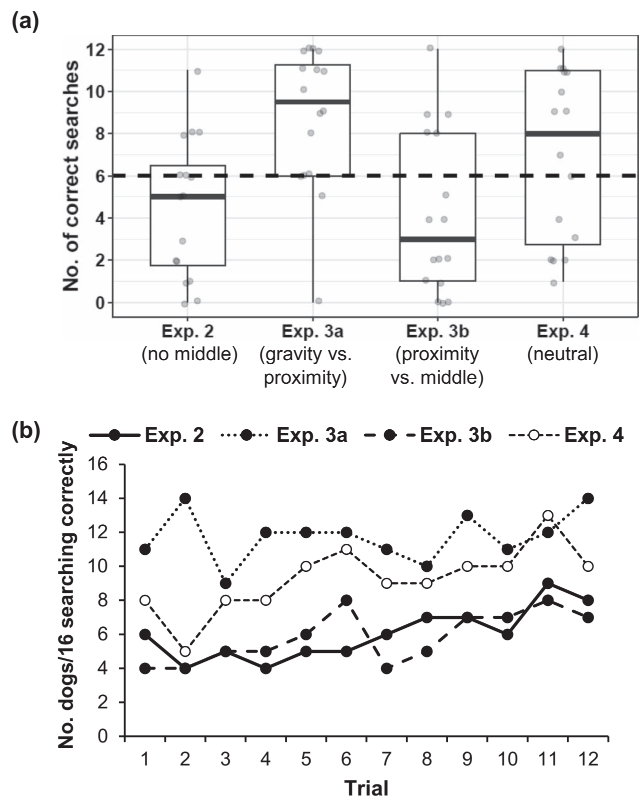

Figure 5. (a) Box plot showing the number of searches directed at the correct location in Experiments 2 (no middle location), 3a (gravity vs. proximity), 3b (gravity vs. middle), and 4 ("neutral" version). The dashed horizontal line represents the expected number of searches per location if search was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

performance improved significantly across the session (trial log odds = 0.12, $z = 2.39$, $p = .017$; Figures 4a and 5b; Table S2 in the online supplemental materials).

We also examined individual performance, and whether individual dogs had a preference for either of the search locations. One dog searched correctly significantly more often than expected by chance across 12 trials (11 out of 12 trials correct, $p = .006$), and six out of 16 dogs exhibited a significant preference to search the gravity location (10 out of 12 to 12 out of 12 trials; binomial test: $p < .039$; see Table S4 in the online supplemental materials for apparatus configuration information for these dogs).

Discussion

These results suggest that, when there is no middle location to search, dogs' tendency to search the middle location gets shifted to the gravity location. In contrast to our findings in this experiment, dogs tested with a comparable setup in Osthaus et al.'s (2003) Experiment 3 seemed to shift to searching the correct location, though only eight dogs were tested, so direct comparison of findings is challenging. Common marmosets on the other hand shifted to searching the gravity location when the middle cup was removed (Cacchione & Burkart, 2012). On the basis of our results in the present experiment, should we therefore conclude that dogs have a weak gravity bias that is masked by a stronger preference to search the middle location?

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

Experiment 3: Teasing Apart the Influence of Gravity, Proximity, and Middle Biases

To attempt to tease apart whether dogs' search in Experiment 2 was influenced by gravity or proximity, in Experiment 3 we pit gravity *against* proximity, by configuring the apparatus so that the gravity location is a greater distance from the top of the tube (where the reward is last seen by the dog) than the correct location (see Figure 4b and 4c). To our knowledge these versions of the diagonal tube task have not previously been presented to any species.

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

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

Experiment 3a: Gravity Versus Proximity/Middle

Subjects. Sixteen dogs (six male, 10 female; mean age = 51 ± 9 months) participated in Experiment 3a (Table S1 in the online supplemental materials). Four additional dogs were tested but excluded because they did not reach criterion in the cup pretraining trials (one), or because they did not complete the cup pretraining trials (one) or the test trials (two).

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

Results. On average, dogs required 9.4 ± 0.8 trials (mean ± SEM; range = 6–14; median = 9) to reach criterion in the cup pretraining trials. In the first test trial, 11 out of 16 dogs searched the correct location, which did not differ from chance (chance: eight dogs searching correctly; exact binomial test: $p = .21$).

Across the 12 test trials, 73% of searches were directed to the correct location ($\text{mean}_{\text{correct}} = 8.8 \pm 0.8$ trials; Figure 5a), and just 27% of searches to the gravity location ($\text{mean}_{\text{gravity}} = 3.3 \pm 0.8$ trials). Dogs as a group searched the correct location significantly more often than expected by chance ($z = 2.867$, $p = .004$; Figure 5a); that is, they were more likely to search the correct, proximal location than the gravity location. The mixed effects logistic regression model revealed no change in performance across trials (trial log odds = 0.06, $z = 1.00$, $p = .316$; Figures 4b and 5b; Table S2 in the online supplemental materials).

We also examined individual performance, and whether individual dogs had a preference for either of the search locations. Half of the dogs (eight out of 16) searched the correct/middle location significantly more often than expected by chance across 12 trials (10 out of 12 to 12 out of 12 trials correct; binomial test: $p <$ 0.039, see Table S4 in the online supplemental materials for apparatus configuration information for these dogs). Only one dog exhibited a significant preference to search the gravity location (12 out of 12 trials, $p < .001$; Table S3 in the online supplemental materials).

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

Experiment 3b: Gravity/Middle Versus Proximity

Subjects. Sixteen dogs (seven male, nine female; mean age = 57 ± 11 months) participated in Experiment 3b (Table S1 in the online supplemental materials). Four additional dogs were tested but excluded because they didn't reach criterion in the cup pretraining trials (two), or because they did not complete the test trials (two).

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

Results. On average, dogs required 10.75 ± 1.24 trials (mean ± SEM; range = 6–21; median = 9) to reach criterion in the cup pretraining trials. In the first test trial, four out of 16 dogs searched the correct location, which did not differ significantly from chance (chance: eight dogs searching correctly; binomial test: $p = .077$).

Across the 12 test trials, 35% of searches were directed to the correct location ($\text{mean}_{\text{correct}} = 4.2 \pm 1.0$ trials; Figure 5b), and 65% of searches to the gravity location ($\text{mean}_{\text{gravity}} = 7.8 \pm 1.0$ trials). Thus, although dogs tended to search incorrectly, overall performance did not quite reach significance ($z = -1.895$, $p = .058$; Figure 5b). As in Experiment 3a, the mixed effects logistic regression model revealed no change in performance across trials (trial log odds = 0.08, $z = 1.50$, $p = .132$; Figures 4c and 5b; Table S2 in the online supplemental materials).

We also examined individual performance, and whether individual dogs had a preference for either of the search locations. One dog searched correctly significantly more often than expected by chance across 12 trials (12 out of 12 trials, $p < .001$), and eight out of 16 dogs exhibited a significant preference to search the gravity/middle location (10 out of 12 to 12 out of 12 trials; binomial test: $p < .039$; Table S3 in the online supplemental materials; see Table S4 in the online supplemental materials for apparatus configuration information for these dogs).

Discussion. This shift from above-chance performance in Experiment 3a to close-to-below-chance performance in Experiment 3b, despite the fact that the spatial relationship between the tube and the two search locations was the same in both cases, demonstrates that above all else, dogs' search is directed to the center of the apparatus. This result is in line with Osthaus et al.'s (2003) Experiment 4, which showed that dogs searched the "gravity location" more often when it was located bottom-middle, and also searched correctly more often when the correct location was bottom-middle. This finding also enables us to rule out several previously posited explanations for why individuals tend to search the middle location in three-cup versions of the diagonal tube task. First, it eliminates the possibility that dogs search the middle location because they confuse it spatially with the gravity location, as suggested by Hood et al. (1999), as in our Experiments 3a and 3b the middle and gravity locations are clearly physically separated, both horizontally and vertically. Therefore, it seems infeasible that dogs could confuse the two locations spatially. Second, it also rules out the possibility that dogs are performing some sort of naïve averaging that leads them to search in the center, because here there are only two available search options. Finally, it also excludes the suggestion that individuals search the middle because the middle and gravity locations are both closer to reward's dropping point than the correct location (Hood et al., 1999), as this was not true in our Experiment 3b, where the correct location was in closer proximity to the reward's dropping point than the middle location.

Experiment 4: Does Removing Sources of Bias Reveal Successful Performance?

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

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

Subjects

Sixteen dogs (10 male, six female; mean age = 41 ± 7 months) participated in Experiment 4 (Table S1 in the online supplemental materials). No dogs had to be excluded from this experiment.

Materials

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

Results

On average, dogs required 7.1 ± 0.4 trials (mean ± SEM; range = 6–10; median = 6) to reach criterion in the cup pretraining trials. Dogs did not show a bias for searching any particular location in their first trial; eight out of 16 dogs searched the correct location and eight out of 16 dogs searched the incorrect location (chance: eight dogs searching correctly; binomial test: $p = 1.00$).

Across the 12 test trials, 58% of searches were directed to the correct location ($\text{mean}_{\text{correct}} = 6.9 \pm 1.0$ trials; Figure 5a), and 42% of searches to the incorrect location ($\text{mean}_{\text{incorrect}} = 5.1 \pm 1.0$ trials). Dogs as a group failed to search the correct location significantly more often than expected by chance ($z = 1.08$, $p = .28$; Figure 5a), though according to the mixed effects logistic regression model, their performance improved significantly across the session (trial log odds = 0.17, $z = 3.00$, $p = .003$; Figures 4d and 5b; Table S2 in the online supplemental materials).

We also examined individual performance, and whether individual dogs had a preference for either of the search locations. Six dogs searched correctly significantly more often than expected by chance across 12 trials (10 out of 12 to 12 out of 12 trials, binomial test: $p < .039$). Four dogs exhibited a significant preference for the incorrect location (10 out of 12 to 11 out of 12 trials, $p < .039$; Table S3 in the online supplemental materials; see Table S4 in the online supplemental materials for apparatus configuration information for these dogs).

Discussion

By removing the gravity and central locations and making both search options equally proximal to the reward's dropping point, we eliminated potential cues that could be influencing dogs' search behavior. If search biases were masking dogs' actual knowledge of the physical–causal structure of the task in previous experiments, then we would have expected them to succeed here. This was not the case—although the majority of searches were directed to the correct location, overall performance was not better than chance. However, performance was improved relative to some of our other experiments (Figure 5, see next section for model comparing these experiments), providing some evidence that eliminating sources of bias may have helped dogs to some extent, potentially by enabling them to focus on the relevant cue of the tube.

Comparison of Performance in Experiments 2–4 and Interim Discussion

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

We used mixed effects logistic regression that assumed a fixed slope across subjects to compare dogs' ability to search correctly between Experiments 2, 3a, 3b, and 4 (Table S6 in the online supplemental materials). Dogs were significantly more likely to search the correct location when it was positioned in the middle of the apparatus and most proximal to the point where it was last seen (Exp. 3a), compared with in Experiment 2 where there was no middle cup (log odds = 2.23, $z = 3.49$, $p < .001$) and compared with Experiment 3b, when the incorrect/gravity location was in the middle (log odds = 2.42, $z = 3.58$, $p < .001$; Figure 5a and 5b). Dogs also performed better in Experiment 4 where potential sources of bias were eliminated than in Experiment 3b (log odds = 1.46, $z = 2.21$, $p = .027$). There were no other significant differences between experiments in terms of dogs' ability to search correctly, though there was a pattern of greater success in Experiment 4 compared with Experiment 2 (log odds = 1.27, $z = 1.94$, $p = .052$; Figure 5b and 5c).

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

In Experiment 4, dogs' performance was significantly improved relative to Experiment 3b but not better than chance. This suggests that eliminating potential sources of bias may go some way to improving dogs' search for a reward invisibly displaced down a diagonal tube, but does not reveal successful performance, that is, it is not the case that search biases are masking dogs' true knowledge of the physical–causal structure of the task (Gómez, 2005).

General Discussion

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

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

A lack of ability to reason about the constraints imposed by the tube to locate hidden food fits with dogs' performance in other physical–causal reasoning tasks, where they have been outperformed by great apes (Bräuer et al., 2006) and wolves (Lampe et al., 2017). Solving the diagonal tube task by reasoning about its physical–causal structure requires knowledge of object permanence, invisible displacement, object solidity, and gravity, as well as the ability to elicit an appropriate search response (Tecwyn & Buchsbaum, 2018). Although there is some evidence that dogs may possess an implicit understanding of object solidity based on looking-time experiments (Pattison, Miller, Rayburn-Reeves, & Zentall, 2010), studies that have investigated whether they can accurately search for invisibly displaced objects have proven inconclusive (Collier-Baker, Davis, & Suddendorf, 2004; Fiset & Leblanc, 2007; Miller, Rayburn-Reeves, & Zentall, 2009), with dogs only compellingly passing specific simplified versions of invisible displacement tasks (Miller et al., 2009; Zentall & Pattison, 2016).

Although dogs generally failed to search correctly in the diagonal tubes task, their errors were not of the same nature as those observed in children. Specifically, we found no evidence that dogs exhibit a gravity bias in the diagonal tube task, either across trials or in Trial 1. In fact, in all of the three-cup versions of the task (Experiments 1a–c), when dogs searched incorrectly they were significantly more likely to search the middle location than the gravity location. Even in experiments where dogs did mainly search the gravity location (Exp. 2 where there was no middle cup, and Exp. 3b where the gravity cup was in the middle), the distribution of their searches did not differ from chance. It is possible that the incorporation of a tube familiarization phase in the present study could have diminished dogs' gravity bias relative to that reported by Osthaus and colleagues (e.g., our Experiment 1a: six out of 16 Trial 1 gravity searches; Osthaus et al.'s (2003) comparable Experiment 1 diagonal condition: eight out of 16 gravity searches). However, given that overall our results generally replicated those of Osthaus et al., the tube familiarization appears not to have had a great impact on performance.

Why would dogs not exhibit a gravity bias? After all, they are subject to the same laws of physics as young children, and both species exist in a world where objects do typically fall straight down. Further, it seems likely that dogs have much experience of seeing objects (e.g., food, balls) being dropped onto the ground. One possibility is that even if dogs are able to predict that a dropped object will fall straight down, their cognition is fundamentally different to that of humans and they do not form a naïve theory on the basis of this information. This could also explain the qualitative difference in the gravity bias seen in children versus some other primates—perhaps only humans form and reason on the basis of a naïve theory of gravity, which results in perseverative searching of the gravity location. Other species (e.g., cotton-top tamarins) might predict that an unsupported item will fall straight down, but because they have not formed a robust theory about this, searching of the gravity location rapidly decreases after Trial 1. Relatedly, this prediction may not transfer to a situation where the object immediately moves out of sight (as is the case when it is dropped into an opaque tube). An alternative possibility is that human infants learn about the properties and behavior of objects, including the effect of gravity on objects, through their own actions—we are all familiar with toddlers in high chairs repeatedly throwing things onto the floor. Dogs' anatomy does not afford the same opportunity to act on objects and therefore limits the extent to which they are able to learn from observing the effects of their own actions on these objects. Presenting human infants who have not yet started manually interacting with objects with either an eye-tracking or looking time version of the diagonal tube task could enable investigation of this; if repetitive experience of acting on objects is critical for the development of a gravity bias, then these infants should not expect the object to end up in the gravity location. Work by Spelke and colleagues (1992) suggests showing 4-month-olds that an item dropped behind an occluder has remained suspended in midair does not appear to violate their expectations, thus lending support to the idea that young infants might not have an expectation that dropped objects will fall straight down to the ground.

Adapting looking-based measures with dogs would also enable the investigation of one further possibility: that dogs in fact *do* have a gravity bias (or, indeed, they are able to correctly predict where the reward will end up, as has been found for marmosets; Cacchione & Burkart, 2012), but this is not revealed by their search behavior. Dissociations between looking-based and action-based measures have been found for the tubes task and other physical reasoning tasks in nonhuman primates (Cacchione &

Burkart, 2012; Santos & Hauser, 2002) as well as young children (Lee & Kuhlmeier, 2013). Action-based versions of the tubes task pose executive demands, as well as requiring individuals to use "feedforward logic-causal inferences" (Cacchione & Rakoczy, 2017), so it is feasible that dogs might predictively look to the gravity location (or the correct location), but then proceed to search elsewhere.

In all of the experiments where a cup was positioned in the center of the apparatus (Exp. 1a–c; Exp. 3a–b), the majority of dogs' searches were directed to that location. Why might dogs have a preference to search initially in the middle? A tendency to search the middle has been observed previously in dogs (Osthaus et al., 2003), as well as in two different monkey species (Cacchione & Burkart, 2012; Hauser et al., 2001; Hood et al., 1999). However, although the authors of these studies speculated about potential reasons for a tendency to search the middle location (e.g., spatial confusion between the gravity and middle locations [Hood et al., 1999], approximation of the reward's position [Hauser et al., 2001], search the middle when uncertain [Osthaus et al., 2003]), previous work did not explore these possibilities experimentally. We took this on in our Experiment 3, the results of which suggested that dogs have a preference to commence their search at the center of the apparatus, as opposed to spatially confusing the gravity and middle locations, or engaging in naïve averaging of competing search preferences. The explanation offered by Osthaus et al. (2003) remains plausible—that when dogs are uncertain of the reward's location, they commence searching at the center of the apparatus. Future work could explore whether this strategy is specific to the diagonal tube task (e.g., related to the constraints of the frame) or a more general strategy under conditions of uncertainty, by, for example, hiding a reward in one of an array of cups and recording dogs' search behavior. If searching in the middle reflects a general strategy, dogs should also commence searching centrally in this context. Experiment 3 also allowed us to rule out the possibility that dogs' search might be influenced by proximity to the last place the reward was seen, which has never previously been explored in any species in the vertical version of the tubes task.

The diagonal tube task has been used to study the gravity bias and physical reasoning abilities in human children and a range of animal species, and so we chose to use this task here to replicate and extend this previous work. However, given that the tube is a very specific causal mechanism that is likely unfamiliar to dogs (and to animals more generally), future work should explore dogs' physical reasoning abilities using more ecologically plausible paradigms. Although what is known about domestic dog physical cognition suggests that the species might have relatively poor skills in this domain, physical and causal reasoning abilities have not been studied in dogs to the same extent as in other taxa (e.g., primates, corvids, parrots), and some of the more basic tasks that have been used to investigate intuitions about fundamental object properties such as solidity and support in other species have been bypassed in favor of more complex designs (Müller, Riemer, Range, & Huber, 2014). For example, a search-based version of the table or shelf task (Cacchione, Call, & Zingg, 2009; Hood et al., 2000; Spelke et al., 1992), suitably adapted for dogs, could be an appropriate means to investigate dogs' knowledge of solidity, as well as providing an additional paradigm with which to examine whether dogs' search might be guided by gravity, as seems to be the case for macaques presented with this task (Hauser et al., 2001).

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

In conclusion, across seven experiments we found no evidence that dogs spontaneously grasp that the tube constrains the path of the reward and guides it to the cup attached to its bottom end. However, our data also suggest that this failure is not primarily explained by a gravity bias. Based on current evidence, it is possible that a gravity bias might be unique to some primate species, or potentially (given the mixed evidence from nonhuman primate studies) unique to young human children. To better understand the origins of the gravity bias and the mechanisms underpinning it, additional groups should be tested with the diagonal tube task, ideally using a developmental comparative approach in which evidence for a gravity bias is examined in immature and mature individuals, across species that differ with respect to their causal knowledge and inhibitory control skills.

References

Baillargeon, R. (2002). The acquisition of physical knowledge in infancy: A summary in eight lessons. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 47–83). Malden, MA: Blackwell Publishers Ltd. http://dx.doi.org/10.1002/9780470996652.ch3

Baker, S. T., Gjersoe, N. L., Sibielska-Woch, K., Leslie, A. M., & Hood, B. M. (2011). Inhibitory control interacts with core knowledge in toddlers' manual search for an occluded object. *Developmental Science, 14,* 270–279. http://dx.doi.org/10.1111/j.1467-7687.2010.00972.x

Bascandziev, I., & Harris, P. L. (2010). The role of testimony in young children's solution of a gravity-driven invisible displacement task. *Cognitive Development, 25,* 233–246. http://dx.doi.org/10.1016/j.cogdev.2010.06.002

Bascandziev, I., & Harris, P. L. (2011). Gravity is not the only ruler for falling events: Young children stop making the gravity error after receiving additional perceptual information about the tubes mechanism. *Journal of Experimental Child Psychology, 109,* 468–477. http://dx.doi.org/10.1016/j.jecp.2011.03.010

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. http://dx.doi.org/10.18637/jss.v067.i01

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

Cacchione, T., & Burkart, J. M. (2012). Dissociation between seeing and acting: Insights from common marmosets (*Callithrix jacchus*). *Behavioural Processes, 89,* 52–60. http://dx.doi.org/10.1016/j.beproc.2011.10.010

Cacchione, T., & Call, J. (2010). Intuitions about gravity and solidity in great apes: The tubes task. *Developmental Science, 13,* 320–330. http://dx.doi.org/10.1111/j.1467-7687.2009.00881.x

Cacchione, T., Call, J., & Zingg, R. (2009). Gravity and solidity in four great ape species (*Gorilla gorilla, Pongo pygmaeus, Pan troglodytes, Pan paniscus*): Vertical and horizontal variations of the table task. *Journal of Comparative Psychology, 123,* 168–180. http://dx.doi.org/10.1037/a0013580

Cacchione, T., & Rakoczy, H. (2017). Comparative metaphysics: Thinking about objects in space and time. In J. Call (Ed.), *Handbook of comparative psychology* (pp. 579–599). Washington, D.C.: American Psychological Association. http://dx.doi.org/10.1037/0000012-026

Collier-Baker, E., Davis, J. M., & Suddendorf, T. (2004). Do dogs (*Canis familiaris*) understand invisible displacement? *Journal of Comparative Psychology, 118,* 421–433. http://dx.doi.org/10.1037/0735-7036.118.4.421

Fiset, S., & Leblanc, V. (2007). Invisible displacement understanding in domestic dogs (*Canis familiaris*): The role of visual cues in search behavior. *Animal Cognition, 10,* 211–224. http://dx.doi.org/10.1007/s10071-006-0060-5

Freeman, N. H., Hood, B. M., & Meehan, C. (2004). Young children who abandon error behaviourally still have to free themselves mentally: A retrospective test for inhibition in intuitive physics. *Developmental Science, 7,* 277–282. http://dx.doi.org/10.1111/j.1467-7687.2004.00346.x

Gómez, J.-C. (2005). Species comparative studies and cognitive development. *Trends in Cognitive Sciences, 9,* 118–125. http://dx.doi.org/10.1016/j.tics.2005.01.004

Hauser, M. D., Williams, T., Kralik, J. D., & Moskovitz, D. (2001). What guides a search for food that has disappeared? Experiments on cotton-top tamarins (*Saguinus oedipus*). *Journal of Comparative Psychology, 115,* 140–151. http://dx.doi.org/10.1037/0735-7036.115.2.140

Hood, B. M. (1995). Gravity rules for 2- to 4-year olds? *Cognitive Development, 10,* 577–598. http://dx.doi.org/10.1016/0885-2014(95)90027-6

Hood, B. M. (1998). Gravity does rule for falling events. *Developmental Science, 1,* 59–63. http://dx.doi.org/10.1111/1467-7687.00013

Hood, B. M., Hauser, M. D., Anderson, L., & Santos, L. (1999). Gravity biases in a non-human primate? *Developmental Science, 2,* 35–41. http://dx.doi.org/10.1111/1467-7687.00051

Hood, B. M., Santos, L., & Fieselman, S. (2000). Two-year-olds' naïve predictions for horizontal trajectories. *Developmental Science, 3,* 328–332. http://dx.doi.org/10.1111/1467-7687.00127

Hood, B. M., Wilson, A., & Dyson, S. (2006). The effect of divided attention on inhibiting the gravity error. *Developmental Science, 9,* 303–308. http://dx.doi.org/10.1111/j.1467-7687.2006.00493.x

Jaswal, V. K. (2010). Believing what you're told: Young children's trust in unexpected testimony about the physical world. *Cognitive Psychology, 61,* 248–272. http://dx.doi.org/10.1016/j.cogpsych.2010.06.002

Joh, A. S., Jaswal, V. K., & Keen, R. (2011). Imagining a way out of the gravity bias: Preschoolers can visualize the solution to a spatial problem. *Child Development, 82,* 744–750. http://dx.doi.org/10.1111/j.1467-8624.2011.01584.x

Joh, A. S., & Spivey, L. A. (2012). Colorful success: Preschoolers' use of perceptual color cues to solve a spatial reasoning problem. *Journal of Experimental Child Psychology, 113,* 523–534. http://dx.doi.org/10.1016/j.jecp.2012.06.012

Kundey, S. M. A., De Los Reyes, A., Taglang, C., Baruch, A., & German, R. (2010). Domesticated dogs' (*Canis familiaris*) use of the solidity principle. *Animal Cognition, 13,* 497–505. http://dx.doi.org/10.1007/s10071-009-0300-6

Lampe, M., Bräuer, J., Kaminski, J., & Virányi, Z. (2017). The effects of domestication and ontogeny on cognition in dogs and wolves. *Scientific Reports, 7,* 11690. http://dx.doi.org/10.1038/s41598-017-12055-6

Lee, V., & Kuhlmeier, V. A. (2013). Young children show a dissociation in looking and pointing behavior in falling events. *Cognitive Development, 28,* 21–30. http://dx.doi.org/10.1016/j.cogdev.2012.06.001

Miller, H. C., Rayburn-Reeves, R., & Zentall, T. R. (2009). What do dogs know about hidden objects? *Behavioural Processes, 81,* 439–446. http://dx.doi.org/10.1016/j.beproc.2009.03.018

Müller, C. A., Riemer, S., Range, F., & Huber, L. (2014). Dogs' use of the solidity principle: Revisited. *Animal Cognition, 17,* 821–825. http://dx.doi.org/10.1007/s10071-013-0709-9

Osthaus, B., Lea, S. E. G., & Slater, A. M. (2005). Dogs (*Canis lupus familiaris*) fail to show understanding of means-end connections in a string-pulling task. *Animal Cognition, 8,* 37–47. http://dx.doi.org/10.1007/s10071-004-0230-2

Osthaus, B., Slater, A. M., & Lea, S. E. G. (2003). Can dogs defy gravity? A comparison with the human infant and a non-human primate. *Developmental Science, 6,* 489–497. http://dx.doi.org/10.1111/1467-7687.00306

Pattison, K. F., Miller, H. C., Rayburn-Reeves, R., & Zentall, T. (2010). The case of the disappearing bone: Dogs' understanding of the physical properties of objects. *Behavioural Processes, 85,* 278–282. http://dx.doi.org/10.1016/j.beproc.2010.06.016

Range, F., Möslinger, H., & Virányi, Z. (2012). Domestication has not affected the understanding of means-end connections in dogs. *Animal Cognition, 15,* 597–607. http://dx.doi.org/10.1007/s10071-012-0488-8

R Development Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Santos, L. R., & Hauser, M. D. (2002). A non-human primate's understanding of solidity: Dissociations between seeing and acting. *Developmental Science, 5,* F1–F7. http://dx.doi.org/10.1111/1467-7687.t01-1-00216

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99,* 605–632. http://dx.doi.org/10.1037/0033-295X.99.4.605

Tecwyn, E. C., & Buchsbaum, D. (2018). Hood's gravity rules. In J. Vonk & T. Shackleford (Eds.), *Encyclopedia of animal cognition and behavior*. Cham, Switzerland: Springer International Publishing. Advance online publication. http://dx.doi.org/10.1007/978-3-319-47829-6_1535-1

Tomonaga, M., Imura, T., Mizuno, Y., & Tanaka, M. (2007). Gravity bias in young and adult chimpanzees (*Pan troglodytes*): Tests with a modified opaque-tubes task. *Developmental Science, 10,* 411–421. http://dx.doi.org/10.1111/j.1467-7687.2007.00594.x

Zentall, T. R., & Pattison, K. F. (2016). Now you see it, now you don't: Object permanence in dogs. *Current Directions in Psychological Science, 25,* 357–362. http://dx.doi.org/10.1177/0963721416664861

Received April 15, 2018
Revision received July 5, 2018
Accepted July 5, 2018 ■