ManyDogs 1: A Multi-Lab Replication Study of Dogs’ Pointing Comprehension

ManyDogs Project*, Julia Espinosa[1], Emily E. Bray[2,3], Daphna Buchsbaum[4], Sarah-Elizabeth Byosiere[5], Molly Byrne[6], Marianne S. Freeman[7], Gitanjali E. Gnanadesikan[2,8], C.-N. Alexandrina Guran[9,10], Daniel J. Horschler[12], Ludwig Huber[11], Angie M. Johnston[6], Evan L. MacLean[2,8], Madeline H. Pelgrim[4], Laurie Santos[12], Zachary A. Silver[12], Jeffrey R. Stevens[13], Christoph J. Völter[11], & Lena Zipperling[11]

[1] Department of Psychology, University of Toronto

[2] School of Anthropology, University of Arizona

[3] Canine Companions for Independence

[4] Cognitive, Linguistic, & Psychological Sciences Department, Brown University

[5] Thinking Dog Center, CUNY Hunter College

[6] Department of Psychology and Neuroscience, Boston College

[7] Animal Health and Welfare Research Centre, University Centre Sparsholt

[8] Cognitive Science Program, University of Arizona

[9] Vienna Cognitive Science Hub, University of Vienna

[10] Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna

[11] Messerli Research Institute, University of Veterinary Medicine Vienna

[12] Department of Psychology, Yale University

[13] Department of Psychology, Center for Brain, Biology & Behavior, University of Nebraska-Lincoln

Author note

This is the finalized preprint of our pre-registered report that has been accepted in principle at *Animal Behavior and Cognition.*

*The ManyDogs Project is a consortium of labs worldwide working to produce reproducible research on canine science. Contact the ManyDogs Project at manydogsproject@gmail.com. More information is available at https://manydogsproject.github.io/.

Correspondence concerning this article should be addressed to Julia Espinosa, Department of Psychology, University of Toronto, 100 St George Street, Toronto, ON M5S 3G3, Canada. E-mail: jhespino@umich.edu

Abstract

To promote collaboration across canine science, address reproducibility issues, and advance open science practices within animal cognition, we have launched the ManyDogs consortium, modeled on similar ManyX projects in other fields. We aimed to create a collaborative network that (a) uses large, diverse samples to investigate and replicate findings, (b) promotes open science practices of preregistering hypotheses, methods, and analysis plans, (c) investigates the influence of differences across populations and breeds, and (d) examines how different research methods and testing environments influence the robustness of results. Our first study combines a phenomenon that appears to be highly robust—dogs' ability to follow human pointing—with a question that remains controversial: do dogs interpret pointing as a social communicative gesture or as a simple associative cue? We collected preliminary data (N = 61) from a single laboratory on two conditions of a 2-alternative object choice task: (1) Ostensive (experimenter pointed to a baited cup after making eye-contact and saying the dog's name); (2) Non-ostensive (experimenter pointed to a baited cup without making eye-contact or saying the dog's name). Dogs followed the ostensive point, but not the non-ostensive point, significantly more often than expected by chance. Preliminary results also provided suggestive evidence for variability in point-following across dog breeds. The next phase is the global participation stage of the project. We propose to replicate this protocol in a large and diverse sample of research sites, simultaneously assessing replicability between labs and further investigating the question of dogs' point-following comprehension.

Keywords: Domestic dog; Reproducibility; Human pointing; Social cognition; Interspecific interaction; Object choice task

ManyDogs 1: A Multi-Lab Replication Study of Dogs’ Pointing Comprehension

The scientific literature within animal behavior is beset with contradictory claims and findings. Variability in results can arise due to methodological differences across studies, response measures that lack standardization, underpowered studies, and/or individual differences across animals (Rodriguez et al., 2021). Teasing apart the relative contributions of these factors can be challenging. Replication of results is essential to understand the variation between studies and to maintain external validity while maximizing the internal validity of experiments (Stevens, 2017; Voelkl et al., 2018; Farrar et al., 2020). Additionally, replication helps discern true effects from spurious findings, by strengthening evidence for the former and weakening evidence for the latter (McShane et al., 2019), thus improving knowledge and informing future research avenues. However, it can be challenging to independently replicate others’ methodologies: replication studies can be difficult to fund and publish, and there may be publication biases (Agnoli et al., 2021; Farrar et al., 2021). Thus, independent laboratory research on its own is not enough to stabilize effects in the literature— standardized replication remains essential. Despite this, relatively few empirical claims within psychology or animal behavior have been subject to direct replication attempts (Makel et al., 2012).

A number of consortium projects have begun to address replication issues in various psychological sciences, including social psychology (Klein et al., 2014), primate cognition (Many Primates et al., 2019) and developmental psychology (ManyBabies Consortium, 2020). These projects promote large-scale collaborations through open science platforms, with groups across multiple institutions working on a common project. Each ManyX project has a specific focus relevant to the concerns of its subfield; however, the overarching mission of each of these projects is the same—investigate the boundaries of reproducibility in the subfield and identify factors that influence reproducibility.

ManyDogs

Canine science is a relatively new subfield within animal behavior, with an explosion of studies over the past two decades (Aria et al., 2021). Similarly to other disciplines, canine science has struggled with underpowered studies and idiosyncratic methodologies, which make it difficult to assess and reconcile conflicting findings (Rodriguez et al., 2021). To address the issue of reproducibility within the field of canine science, we have developed a new consortium project: ManyDogs. Drawing inspiration from other ManyX projects (e.g., ManyBabies, ManyLabs, ManyPrimates), the primary goals of the first ManyDogs project are to (1) enhance replicability in the field of canine science, (2) provide a platform for testing questions that require large and/or diverse samples, (3) quantify differences across labs and investigate how these differences might influence study results, and (4) foster international collaborations moving forward. We aim to do this in a collaborative network that (a) uses large, diverse samples to investigate and replicate findings, (b) promotes open science practices of preregistering hypotheses, methods, and analysis plans, and (c) examines how different research methods and testing environments influence the robustness of the results. Thus, there is an exciting opportunity to initiate replication efforts in canine science, including explorations of the robustness of basic findings in the field.

As part of enhancing the replicability of results across the field of canine science, through the collaborative efforts of ManyDogs we aim to begin quantifying differences across labs (e.g., in testing environments, methodological approaches, and analysis techniques) to investigate how these differences influence study results. We hope a closer analysis of these inter-lab differences will provide useful information for developing a set of best practices (Byers-Heinlein et al., 2020), similar to what the field of infant cognition has achieved with the findings from ManyBabies, who in their first study replicated infants´ bias for infant-directed speech, but produced a more moderate effect size (The ManyBabies Consortium, 2020). By building large international datasets, we will also be able to investigate questions that none of us could address alone, such as questions about the impact of individual differences in training history, breed, or geographical location on cognition and behavior. Lastly, we hope this will be the first project of many, and that researchers in all areas of canine science will see this platform as a useful tool for generating additional collaborations.

Addressing questions in a large-scale collaboration will provide several valuable opportunities for the field of canine science. First, given the robust power associated with large datasets, our initial study will afford us the best opportunity to date to answer our theoretical question of interest—do dogs understand and act on human pointing gestures as social communicative cues? Second, we can more directly evaluate the boundaries of reproducibility in the still-emerging field of canine science by investigating how much variation in effect size there is in dogs’ overall tendencies to follow pointing gestures across labs. Moreover, with sufficient participation from different research units, we hope to understand the potential causes of variability in effect sizes by investigating the influence of specific differences across labs and/or populations. Third, this project will inform future estimates of statistical power for similar studies in canine science. Finally, we will be able to conduct exploratory analyses on a highly diverse dataset targeted at investigating (a) how other measured factors (e.g., breed) might influence the reproducibility of canine science research in general and (b) the tendency of dogs to follow pointing gestures specifically.

ManyDogs 1: Understanding Human Pointing Gestures

To achieve these goals, we will use a “single study” approach, in which we design one specific study for all participating labs to conduct in parallel. This approach was modeled after the ManyBabies project, and since many of the logistical concerns of infant research are similar to those found in canine research, this approach provided appropriate structure for our first study. First, as with any research with non-verbal individuals (e.g., infants, non-human animals), research with dogs is typically more time intensive than adult human psychology research, as all dogs have to be tested one-by-one with extensive training phases on longer behavioral measures. Second, it can be difficult to determine the cause of contradictory findings given vast individual, cultural, training-related, and breed-related differences among canine populations. Due to the intersections of these differences, it is very difficult to determine the reason behind failed replications across labs: do they reflect meaningful individual differences across different populations, or different methodological approaches across labs? Implementing a single, methodologically uniform study across labs will provide the opportunity for us to directly investigate some of these sources of variability.

For our first study, we have chosen to investigate dogs’ interpretation of human pointing gestures. Dogs’ ability to follow human pointing is a highly robust finding in canine science (e.g., Miklósi et al., 1998; Soproni et al., 2001; Hare et al., 2002; Kaminski & Nitzschner, 2013), though factors such as rearing environment and living conditions may influence point following behavior (Udell et al., 2010; D’Aniello et al., 2017). To study this ability further, and assess the feasibility of the ManyDogs approach, we have chosen a simple choice task that can be standardized across dog labs, addressing a question that is theoretically interesting to many researchers in the field: how do dogs understand and act on human pointing? Do they perceive it as a social communicative gesture—whether informative or imperative—or as a simple associative cue? Social communicative gestures, such as pointing, convey information from the signaler to the observer, and are frequently enhanced by ostensive cues (such as eye-contact, gaze alternation to a target, or vocal signals) that make the intentionally informative nature of the gesture understood (Csibra, 2010). Another way to interpret an intentional pointing gesture is that the signaler is providing an imperative that requires a particular response from the observer (e.g., Kirchhofer et al., 2012). While these two accounts lead to differences in how the cue is received and understood, both involve social signals. However, it has also been proposed that point following in dogs is based on associative learning mechanisms without any specific, 'infant-like' understanding of the human's communicative-referential intention (e.g., Wynne et al., 2008). Thus, point following in dogs could be the result of learning to associate a reward such as food with either the specific gesture, or human hands more generally. We outline our hypotheses for these various explanations below.

With a single experiment that can be carried out at most canine research sites and is intended for widespread global participation, we intend to explore dogs' responses in two different pointing conditions: an ostensive condition (pointing with eye-contact and dog-directed speech) and a non-ostensive condition (pointing without accompanying eye-contact or speech). By investigating dogs' responses to these two contrasting pointing styles with a large and diverse sample, we aim to shed light on dogs' understanding of human pointing gestures, but more importantly, also establish a foundation for multi-lab open science collaborations in canine science.

One of the earliest findings in canine science that catalyzed the growth of the field is that dogs follow pointing gestures more accurately, spontaneously, and flexibly than other species, such as great apes (Bräuer et al., 2006). It is now well-replicated that dogs follow human pointing (Miklósi et al., 1998; Soproni et al., 2001; e.g., Hare et al., 2002; Kaminski & Nitzschner, 2013), even from a very young age (Bray et al., 2021). However, researchers still disagree as to whether dogs interpret human pointing as a social, communicative gesture or whether they simply associate human hands or limbs with food, and if the former, whether they perceive the gesture as informative or imperative. Human children follow pointing from an early age, but only if it is prefaced by clear direct ostensive cues that signal the pointer's intent to provide information (i.e., eye contact, high-pitched infant-directed speech, and/or the child's name; Behne et al., 2005). Thus, for young children these intentional direct ostensive cues are necessary to interpret pointing as an informative gesture. Although a large body of previous research with dogs has demonstrated that dogs are capable of following pointing when it is prefaced by intentional direct ostensive cues (e.g., Miklósi et al., 1998; Soproni et al., 2001; Hare et al., 2002; Kaminski & Nitzschner, 2013; Tauzin et al., 2015a), it is less clear whether these ostensive cues are indeed necessary in the same way they are for human children (i.e., required to perceive the cue as informative).

Researchers have investigated dogs' point-following responses in several ways, from simple conditioning to understanding the cooperative intent and referential (informative) content of the gesture (Pongrácz et al., 2004; Range et al., 2009; Topál et al., 2009; Virányi & Range, 2009; Kupán et al., 2011; Kaminski et al., 2012; Marshall-Pescini et al., 2012; Téglás et al., 2012; Scheider et al., 2013; Moore et al., 2015; Tauzin et al., 2015a, b; Duranton et al., 2017), but to our knowledge only two studies have investigated how ostensive cues influence the way dogs understand and act on pointing (Kaminski et al., 2012; Tauzin et al., 2015a). In one study, an experimenter pointed while either making eye contact with the dog (i.e., an ostensive cue) or looking down at her arm (Kaminski et al., 2012). Dogs were more likely to follow the pointing gesture if the experimenter was making eye contact than if she was not. In fact, dogs in the condition without ostensive eye contact did not follow the pointing gesture above chance levels, while dogs in the condition with ostensive eye contact did. This suggests that ostensive cues may be necessary for dogs to follow pointing. Crucially, however, although eye contact is a sufficient ostensive cue, it is not a necessary cue, as dogs follow pointing gestures even when a person's back is turned, as long as they use high-pitched speech (Kaminski et al., 2012). In another study, an experimenter pointed with ostensive cues (i.e., eye contact and calling the dog's name) either preceding or following the gesture (Tauzin et al., 2015a). Dogs were more likely to follow pointing gestures if the ostensive cues preceded the pointing than if they came after, and only performed above chance levels when the ostensive cues preceded the gesture. Together, these two studies provide promising initial evidence that dogs may find ostensive cues necessary for following pointing gestures. However, in some instances neutral cues performed before the pointing gesture, such as hand clapping (e.g., clapping control condition, Tauzin et al., 2015a) have appeared to increase point following in dogs. It is possible that the facilitating effects of ostensive cues result only from low-level effects like attention-raising (e.g., Szufnarowska et al., 2014; Gredebäck et al., 2018) instead of being a means to identify the communicative intention, as higher-level theories such as Natural Pedagogy theory propose (Csibra, 2010). However, assessing this will require further experiments, with proper control conditions and clear, contrasting predictions. The latter is especially important given that higher-level theories incorporate attentional mechanisms in their explanations; however, this is beyond the scope of the current replication study.

In this study, we aim to test if ostensive cueing has a facilitating effect on dogs' ability to follow pointing gestures from humans. To this end we will compare each dog's performance in conditions with and without ostensive cues preceding the pointing gesture. In the ostensive version, the gesture will be preceded by two kinds of salient direct social cues: eye-contact and calling the dog’s name in a high-pitched voice. Recent studies suggest that dogs, like humans from a very early age, react to human-given gestures only when they are accompanied with such ostensive cues (Téglás et al., 2012; Bray et al., 2021). In the non-ostensive version, the experimenter will perform the hand gesture in a less communicative manner, clearing their throat—to ensure the subjects’ attention without speech—and without eye contact, instead looking down at the floor.

Hypotheses and Predictions

Our main hypothesis is that preceding ostensive cues have a facilitating effect on dogs’ following of human pointing gestures. We predict that if dogs perceive pointing gestures as socially informative cues, they will follow points significantly above chance level in the Ostensive condition, but not in the Non-ostensive condition. Under this hypothesis, pointing gestures alone are not sufficient for dogs to successfully interpret and follow social gestures given by human informants. If we find the dogs in our study perform better in the Ostensive condition than in the Non-ostensive condition, it would provide some evidence that the pointing gesture needs to be preceded by special, ostensive signals from the human demonstrator. If, on the other hand, no difference is observed between conditions, this could suggest that dogs understand pointing as the result of a learned gesture-reward association.

A second hypothesis regards the question of whether dogs interpret pointing gestures as imperative or informative. For humans, the pointing gesture is itself conveying information, namely about the location of an object (e.g., Tomasello et al., 2005). For dogs some researchers have assumed that the gesture is instead interpreted as an imperative directive ordering them where to go (Topál et al., 2009; Wobber & Kaminski, 2011; Kaminski et al., 2012; Kaminski &

Nitzschner, 2013). As argued by Topál et al. (2014), ostensively cued human behaviors can often act as imperatives for the dog, inducing a ‘ready-to-obey’ attitude that may result from the domestication of dogs and/or from their extensive experience with humans. This claim is supported by evidence that dogs prefer following a human’s gesture even if it is against their better knowledge (Scheider et al., 2013; Szetei et al., 2003), although this may also be analogous to human infants, as explained by the Natural Pedagogy account (Csibra & Gergely, 2009). Unlike the informative account, there is no clear prediction on dogs’ point-following behavior in the Non-ostensive condition if they view it as an imperative; it is possible they would follow pointing equally in both conditions, or it is possible that the ostensive cues would still signal intentionality and result in higher levels of point-following in the Ostensive condition. Thus, our planned experimental contrast will not definitively answer this question. However, we expect that if dogs view pointing cues as imperative, training history and trainability would be significant predictors of their performance in both conditions.

Our third and final prediction for the study is that, as has previously been demonstrated in similar paradigms (Bray et al., 2020a, 2021), dogs are not using olfactory cues to find hidden food in this task, and thus we will not see group level performance that is significantly above chance in the Odor Control condition.

In this registered report, we first present the results of preliminary data collection of ostensive versus non-ostensive point-following—validating our pre-registered protocol within a single lab—and then outline the proposed expansion of the study, which will follow identical procedures but include data from multiple labs. The labs will be recruited through an open call to encourage global participation.

Methods

Here, we present a proposed study design to address our research questions. In addition, we include preliminary data from an initial pre-registered study from a single laboratory. Additional methods and videos of pointing conditions are available as supplementary materials on the Open Science Framework (https://osf.io/9r5xf/). The task will be an object-choice paradigm, based on methods by Bray et al. (2020a, b), involving the choice between two cups, under one of which a piece of food is hidden. Most methodological details (e.g., distances, times, setup, phases) are closely based on Bray et al.'s methods, with modifications made to either (1) better accommodate the manipulation of ostension of the present study, or (2) relax and simplify abort criteria for easier implementation with diverse pet dogs in varied contexts. Cups will be opaque and false-baited with a treat taped to the inside of each cup to control for odor cues in warm-up and test trials (unbaited cups will be used for the Odor Control condition). Subjects will have up to 25 s to choose a cup on each trial. A choice is defined as the subject physically touching the cup with their snout or a front paw (not an ear, back leg, or tail). If the subject does not make a choice within 25 s, a "no-choice" will be recorded and the trial repeated. If the subject has two no-choice responses in a row, they will undergo refamiliarization prior to reattempting to complete the warm-up phase or test trials (see refamiliarization procedure below).

Throughout the study, the handler will sit in a chair behind the dog, holding the dog stationary and facing toward the experimenter while the baiting is carried out. The experimenter will be a trained researcher and will maintain a seated position during trials, looking at the floor during the entirety of each choice period to avoid cueing the subject (Figure 1). The handler may be either a trained researcher or the dog's guardian, as appropriate for a given lab. In cases where the guardian is not handling during the study, we recommend (but do not require) that they remain in the room, seated behind the handler. To minimize the potential for unintentional cueing, trained handlers will close their eyes during baiting and cueing (opening them only once the dog has been released), while guardian handlers will close their eyes for the entirety of the trial duration. We believe that this protocol will sufficiently ensure that dogs are not cued to choose a particular location by the handler, especially given that previous empirical work aimed at assessing the Clever Hans effect in point-following tasks in dogs suggests that the effects of any unintentional cueing may be less robust than is often suggested (Schmidjell et al., 2012; Hegedüs et al., 2013).

Procedure

Warm-ups

To familiarize subjects with the testing space, the experimenter, and finding food under cups, a series of warm-up exercises will be conducted. These warm-ups are not intended to be predictive of test performance, simply to build an association between cups and rewards and gauge the subject's willingness to participate in the task and indicate a choice (in a similar paradigm, Bray et al. 2021 found that performance on warm-ups was not predictive of performance on a pointing task). Throughout the warm-up phases, dogs will be spoken to in a high-pitched voice using pet-directed speech; additionally, experimenters will attempt to make eye contact with subjects at the beginning of each trial when showing them the food reward. All cups used for warm-ups will be false-baited to ensure that the cups smell like food and to minimize dogs' ability to choose cups based on their odor. Subjects will proceed to test trials after completing all phases of the warm-ups, or after 15 minutes has elapsed from beginning Phase 1. If, during warm-ups, subjects do not respond on two consecutive trials they will undergo refamiliarization with the previous phase to encourage participation. Exclusion and abort criteria are detailed in the section below.

Phase 1: Visible Placement and Free-form Cup Association

First, there will be at least two repetitions of visible treat placement on the floor in front of the experimenter to ensure the subject is willing to approach the experimenter and eat off the floor in the testing area. Additional trials may be used as necessary. After the subject retrieves the treat successfully from each visible placement, the experimenter will play a free-form cup game to familiarize the subject with finding treats under cups and to encourage them to indicate a choice by touching the cup. In the free-form cup game, the experimenter will show a single treat before placing it on the floor and covering it with a cup. The experimenter will vocally encourage the subject to approach and touch the cup, rewarding them with the treat underneath. This hiding process will be repeated at least three times or until the subject readily touches the cup. On every trial (true of all trial types throughout the study), subjects are allowed to make only one choice and will be rewarded on trials where they touch the baited cup first. Upon choosing, the experimenter will lift the cup, exposing the treat for the subject to eat.

Phase 2: One-cup Alternating

The second phase familiarizes the subject with the setup and general trial procedure and ensures they are willing to approach the cup locations to the right and left of the experimenter (Figure 1).

In this phase, only one cup will be presented in each trial and placed at either the right or left of the experimenter, in one of the two designated cup positions, which are 1 m apart from each other, along a line 1.35 m in front of the dog’s starting box (see Figure 1 and Figure 2). At the start of each trial, the reward will be visibly placed under the cup; the experimenter will attempt to make eye contact with the dog as they bait the cup. The subject will then be required to indicate a choice by physically touching the cup on four trials within a maximum number of seven trials. After each successful trial, the cup will be presented on the opposite side to ensure the subject receives two rewards in each location. Subjects that do not complete four touches within seven trials will be excluded (see refamiliarization and abort criteria below).

Phase 3: Two-cup Alternating

The third phase ensures that the subject attends to the experimenter's actions, is willing to approach both cup locations when a cup is present at each location simultaneously (i.e., not side-biased), and is not choosing randomly. These trials will be identical to the previous phase, except that two identical cups will be used, such that the subject must attend while one cup is baited by the experimenter in order to choose correctly. The experimenter will attempt to make eye contact with the dog as they visibly bait the cup. Several predetermined sequences of baiting locations (four pseudo-random orders, with no more than two trials in a row on the same side) will be counterbalanced across the conditions within a lab (each sequence used four times within the minimum sample of subjects). Subjects will be required to choose correctly on the first presentation of four of the most recent six trials (sliding window) to advance to the test trials; trials in which the dog does not choose correctly will be immediately repeated to minimize side biases. Subjects that do not meet this criterion within 20 total trials (including repeated trials) will be excluded. The experimental setup is shown in Figure 1.

Test Trials

The test trials will include two blocks of eight trials each—one block for each of the two conditions (ostensive vs. non-ostensive)—with the order counterbalanced across individuals. The two blocks will be separated by a one min play break and a re-familiarization (two trials of the two-cup alternating procedure from the warm-up Phase 3).

In both conditions, occluded baiting will be used and each trial begins with the occluder placed in front of the experimenter, by the experimenter, hiding the two cups from the subject's view. Both cups will be false-baited to minimize the dogs’ ability to use odor cues. The experimenter will first visually show the subject the food reward, and then place the reward underneath one of two cups, both of which will be hidden behind the occluder (standardized size across labs: 30 centimeters (cm) tall x 58 cm wide). The experimenter will then remove the occluder and place it behind them, then simultaneously slide the two cups outward from their central position until they are 1 m apart, and then provide one of the pointing cues (described below). Across conditions, experimenters will use a contralateral momentary point, holding the point stationary for 2 s before returning to resting position, maintaining a downward gaze. Although there will be variation across labs and experimenters, the experimenter’s finger will be approximately 30 cm from the cup during the pointing cue. Once in resting position, and after waiting for 1 s, experimenters will cue the handler to release the subject using a neutral word (“now”) and neutral tone to avoid additional social cueing from the experimenter. The handler will release the subject by dropping the leash and saying “okay!” or any similar release command usually used with the subject on which the subject was previously trained. The dog may only choose one cup per trial and will be prevented from making a second choice by removal of the cups or blocking the dog’s access. If they choose the baited cup, they are allowed to eat the food; if they choose the unbaited cup, they are shown the empty space under the cup and no reward is given. On test trials, no praise is given for choosing the baited cup. Except for the gesturing components, detailed below, all other aspects of the test trials will be identical in both conditions.

The primary dependent measure for all test trials will be the proportion of trials in which the subject chooses the baited cup. Subjects have 25 s to make a choice on each trial, and they must complete all test trials of both pointing conditions to be included in registered analyses. Individual exclusion criteria are detailed below.

Ostensive Condition

At the start of each ostensive trial, the experimenter will make eye contact with the subject and say “[dog name], look!” in high-pitched pet-directed speech, while visibly presenting the treat. After treat placement, cup movement, and occluder removal, the experimenter again repeats “[dog name], look!” in pet-directed speech and makes eye contact before presenting the pointing gesture (see Figure 3). While giving the neutral release signal and while the subject approaches, the experimenter will look down at the floor directly in front of them.

Non-ostensive Condition

At the start of each non-ostensive trial, the experimenter will look down and clear their throat to get the subject’s attention while presenting the treat. Before pointing, the experimenter will clear their throat again to attract the subject’s attention and continue to avert their gaze by looking at the ground in front of them while they present the momentary pointing gesture, and while the subject approaches and indicates a choice. Throat clearing was chosen as an easy to produce cue that is familiar to dogs, and not generally associated with ostensive cues or intentional communication, but that would still attract the dog’s attention thus balancing auditory cues across pointing conditions. The experimenter will not speak to the dog during the non-ostensive trials, only speaking the neutral “now” as a cue for the handler to release the dog.

Odor Control Condition

After both blocks of test trials, another one m play break will take place. Finally, in the four odor control trials, the cups will be baited identically to the test trials, except: (1) clean, un-baited cups will be used, without a treat taped into the cup (thus making it easier for subjects to potentially use scent cues if they are using an olfactory search strategy), (2) only one verbal cue will be given when presenting the treat, “[dog name], look,” and (3) no pointing gesture will be provided before the subject is released to search. Based on previous results with similar paradigms Bray et al. (2020a), we expect most subjects to perform at chance levels on these trials. We will therefore use a reduced number of odor control trials to avoid dogs getting discouraged and refusing to participate. This data will not be used on an individual level to exclude subjects, but rather used in post-hoc analyses to investigate dogs’ ability to use olfactory information, or other unintentional cues, at the level of lab, breed, or training background.

Refamiliarization and Abort Criteria

If subjects stop participating during any phases of the task (i.e., refusing treats or making two consecutive no-choice responses in warm-ups or test trials, where no-choice is failing to touch a cup within 25 s), a re-familiarization process will be used. This involves returning to the immediately previous warm-up phase if this behavior occurs during One Cup or Two Cup warm-ups, or if during test trials, then returning to the Two Cup warm-ups in an attempt to re-engage the subject (see supplementary materials for details of this procedure). If refamiliarization with a previous phase does not successfully re-engage the subject in the task, or if the subject makes a total of four no-choice responses in any single phase of the warm-ups or test trials, or if the subject exhibits signs of distress, testing will be aborted. One exception to the abort rule may be made if the subject participates in the Non-ostensive pointing condition first and reaches the limit of no-choice responses. In the absence of signs of distress, the Non-ostensive condition may be aborted and the subject moved on to the Ostensive pointing condition. This exception allows for subjects to try participating in the pointing condition with comparatively greater attention-raising effects, which may be more likely to elicit a response due to the ostensive cues accompanying the gesture. While subjects that do not complete all test trials of both conditions are ineligible to be included in primary analyses, a frequency of the subjects that only respond when points are preceded by ostensive cues is nevertheless informative for determining point following behavior at the group level.

Coding and Reliability

Choices will be coded live by the experimenter. Additionally, videos will be recorded when possible to enable reliability coding, as well as coding of additional exploratory measures. For each participating lab a subset of the data (at least 8 subjects for data submissions with ≤ 40 subjects, see sampling plan below, or 20% of subjects of data submission containing > 40 subjects) will be re-coded for reliability. Recoding should contain equal numbers of subjects from each pointing condition. When possible, reliability coding will be done from video by a coder who is blind to the hypothesis of the project; otherwise, a secondary live coder will be used (only in the event that video data collection is impossible). Labs whose data does not meet the inter-rater reliability threshold of $\kappa \geq 0.9$ will be excluded.

Survey Data

Prior to participation in the behavior study, dog owners and guardians will complete a survey on their dog’s background including breed, training history, and other demographics. Dog owners and guardians will also complete the Canine Behavioral Assessment and Research Questionnaire (C-BARQ©, www.cbarq.org) (Serpell & Hsu, 2001; Hsu & Serpell, 2003). See supplementary materials on OSF for the complete text of our in-house surveys. We included the C-BARQ trainability score as a covariate in our confirmatory analysis to account for the potential impact of varying individual training histories on the dogs’ task performance.

Sampling Plan

This experiment will be conducted at labs around the world. In addition to current consortium labs committed to collecting data (Table 1), we will recruit additional canine science labs and research centers through relevant listservs, conferences and social media channels. Labs will self-select into the project with the only criteria being that they (1) follow the protocol for setting up and running the study, (2) obtain ethics approval from their institution, and (3) collect data from at least 16 subjects that meet submission requirements. Because of the nature of this project, the exact number of participating labs/collaborators cannot be specified ahead of time. Our plan is to fix a data collection end date, and any labs/collaborators who collect data from the minimum of 16 dogs by the end date will be included in the analysis. A minimum number of dogs per lab is set to allow for an assessment of between-lab variation in performance.

For similar reasons, the number of subjects cannot be specified ahead of time. Each lab/collaborator that submits data for this project is required to collect behavioral data, and strongly encouraged to submit video data, from a minimum of 16 dogs in order to be included in final analyses.

Collaborator Onboarding Process

We are using an online survey (hosted on Qualtrics) to recruit research sites to contribute data for ManyDogs 1. Upon completion of this survey, the onboarding process is initiated during which one of the ManyDogs administrative team corresponds closely with the new collaborator to assist them with obtaining ethics approval and with submitting information about their research site to our database. The information that we collect about each site includes a detailed floor plan of the area in which the collaborators plan to collect data, along with details about sound attenuation, room ventilation, if they are using personal protective equipment (PPE) and if so what type, their research assistant training process, and general information about the population from which they will be recruiting individual participants.

To preserve the highest possible level of similarity in how different sites implement the protocol, we have a mandatory experimenter training that must be completed by each research site prior to collecting data. In the training, sites are required to submit a video of their trained experimenter performing each phase of the study protocol and then receive detailed feedback from ManyDogs administrators on how to improve their execution of pointing, etc. The video submission-feedback cycle may be repeated as necessary to achieve consistency and uniformity in the protocol. The second part of the training instructs collaborators in our data entry process, which is completed using online surveys (Qualtrics). Using prepared practice coding sheets, researchers go through the steps of entering behavioral data and receive feedback on any areas for improvement. Upon completing both sections of training, research sites are given an explicit recommendation to begin collecting data and invited to stay in close contact with the ManyDogs admin team throughout their implementation of the protocol. To facilitate frequent and efficient communication between contributors (as well as the ManyDogs admin), we maintain an active Slack workspace that promotes open discussion and troubleshooting in all aspects of participation in the study.

Data Analysis

Data will be analyzed in R Statistical Software (R Core Team, 2021). As an inference criterion, we will use *p*-values below .05. Where possible, we will supplement the frequentist statistics with Bayes factors.

Performance Relative to Chance

We will conduct one-sample (two-tailed) t-tests to compare the subjects’ aggregated performance across trials to the chance level (0.5) separately for each condition (Ostensive, Non-ostensive, and Odor Control). We will also conduct these analyses separately for each lab.

In addition to the frequentist analysis, we will calculate Bayes factors for the t-tests using the `ttestBF()` function (with default, non-informative priors) from the *BayesFactor* package in R (Morey & Rouder, 2018).

Condition Comparison

For our main analysis, we plan to fit a Generalized Linear Mixed Model (GLMM) with binomial error distribution and logit link function using the `glmer()` function from the *lme4* package (Bates et al., 2015). This model will include condition (Ostensive and Non-ostensive only), order of condition (Ostensive first, Non-ostensive first), trial number within condition, dog sex, dog neuter status, dog age (in years), and dogs' trainability score based on the C-BARQ questionnaire (Hsu & Serpell, 2003) as fixed effects and subject and lab as random intercepts. The full model, including fixed effects, random intercepts, and random slopes is defined by: `Correct choice ~ condition + order_condition + trial_within_condition + sex*desexed + age + C-BARQ_trainability_score+ (condition + trial_within_condition +  | Subject ID) + (condition+ order_condition + trial_within_condition + sex*desexed + age + C-BARQ_trainability_score | Lab ID)`. In a second model, we will repeat this analysis with only purebred and known crossbred dogs, excluding mixes of unknown breeds, or of more than two breeds (only breeds/crossbreeds with at least 8 individuals will be included) and include the random effect of breed in this model: `Correct choice ~ condition + order_condition + trial_within_condition + sex*desexed + age + C-BARQ_trainability_score + (condition + trial_within_condition + | Subject ID) + (condition+ order_condition + trial_within_condition + sex*desexed + age + C-BARQ_trainability_score| Lab ID) + (condition+ order_condition + trial_within_condition + sex*desexed + age + C-BARQ_trainability_score | Breed ID)`. We will only include random slopes if the corresponding predictor variable varies in at least 50% of the levels of the random intercept. We will only include the random slope of the interaction if there is sufficient variation in both of its terms in at least 50% of the levels of the random intercept. We will only include the correlations between random intercepts and random slopes if including them results in a model with better fit (i.e., smaller log-likelihood).

All covariates will be centered and scaled to a standard deviation of 1. The random slope components of the factors will be centered to ensure that the results are not conditional on the choice of the reference category.

If the models do not converge, we will follow the steps reported by Bolker (2014). If these procedures do not fix convergence issues, we will remove correlations between random effects then remove random slopes, if needed, in the following order: Lab ID, Subject ID, Breed ID.

For the GLMM, we will calculate likelihood ratio tests using the `drop1()` function from *lme4* (using a chi-square test, Barr et al., 2013) with p-values below .05 as the criterion to make inferences about fixed effects.

In addition to the frequentist GLMM, we will calculate Bayes factors for the models from Bayesian models using the `brm()` function from the *brms* package (Bürkner, 2017, 2018) with default, non-informative priors. We will then use the `bayes_factor()` function to compare models, using bridge sampling for repetitions (Gronau et al., 2020). The Bayes factors will represent the evidence for the full model relative to the full model without the fixed effect under investigation. The Bayesian analysis will be supplemental, and inferences will be drawn from the frequentist statistics.

Genetic Analysis of Among-breed Heritability

To assess among-breed heritability (MacLean et al., 2019), we will fit an animal model (Wilson et al., 2010) which incorporates a genetic effect with a known covariance structure to estimate the proportion of phenotypic variance attributable to additive genetic effects. Genetic analyses will take a breed-average approach, integrating publicly available genetic data on the breeds in our dataset, rather than genotyping the individuals in the cognitive experiment.

Breed average genetic similarity will be represented by an identity-by-state (IBS) matrix calculated from publicly available genetic data collected using the Illumina CanineHD bead array (Parker et al., 2017). The proportion of single-nucleotide polymorphisms (SNPs) identical by state between pairs of individual dogs will be calculated using PLINK (Chang et al., 2015). These values will then be averaged for every pair of breeds in order to generate a breed-average IBS matrix. This breed-average IBS matrix will be extrapolated to an individual-level IBS matrix for the purposes of our analysis. For individuals of different breeds, the IBS value will be set to the average similarity between those breeds in the genetic dataset. For individuals of the same breed, the IBS value will be set to the average IBS value among members of that breed in the genetic dataset. The purpose of this approach is to simultaneously incorporate a measure of between- and within-breed genetic similarity, retaining the ability to model phenotypes at the individual, rather than breed-average level. Only breeds represented by N ≥ 8 individuals will be included in these analyses.

Heritability models will be fit using the `brm()` function from the *brms* package (Bürkner, 2017, 2018) with weakly informative priors. We will use 12,000 iterations per chain, with the first 2,000 iterations being used as a warm-up, and a subsequent thinning interval of 10 iterations for retention of samples for the posterior distributions. We will report the mean and 90% credible interval for the posterior distribution of heritability estimates for this analysis.

Heritability models will include breed-mean body mass, sex, and age as covariates. We will fit three separate models using the following dependent measures: (1) proportion correct in the Ostensive condition, (2) proportion correct in the Non-ostensive control condition, and (3) a difference score between these conditions, in which performance in the Non-ostensive condition is subtracted from performance in the Ostensive condition.

Model performance will be assessed by visualizing fitted values vs residuals and quantile-quantile plots. If problems are detected at this stage, models will be refit using an appropriate statistical transformation of the dependent measure.

Preliminary Data

In order to validate our study design and analysis plan, we collected preliminary data from a pilot experiment at the Clever Dog Lab at the University of Veterinary Medicine in Vienna, Austria. We pre-registered the study design, procedure, predictions, and confirmatory analysis prior to data collection at the Open Science Framework (https://osf.io/gz5pj/). The data and analysis script are available online at ManyDogs OSF.

Methods

Ninety-one dogs (Males = 38, M_{Age} = 5.13 years, *SD* = 3.31) across a variety of breeds participated in the pilot experiment. Of these, a subset of 61 dogs (Males = 26, M_{Age} = 4.74 years, *SD* = 3.25) were tested after our pre-registration was submitted; all statistical models are limited to these individuals. An additional 12 dogs started but did not complete the experiment due to lack of motivation (n = 10) or fear/anxiety (n = 2). The study was discussed and approved by the institutional ethics and animal welfare committee in accordance with Good Scientific Practice guidelines and national legislation (ETK-081/05/2020).

This study used the methods specified above and the analytic plan specified in the OSF pre-registration. A meat-based sausage treat was used, and odor cues were controlled by rubbing the interior of the cups with sausage prior to warm-ups and test trials. With the exception of four subjects (who were handled by a female research assistant), subjects were handled throughout the study by their guardians. While data were live-coded by the experimenter, a second rater naive to the hypotheses and theoretical background of the study scored the video data of 18 randomly selected dogs (ca. 30% of the pre-registered sample). We used Cohen’s kappa to assess the interobserver reliability of the binary response variable “correct choice.” The two raters were in complete agreement ($\kappa = 1$, N = 360).

Data Analysis

To evaluate whether dogs’ performance in correctly choosing the cup with the treat deviated significantly from the chance level of 0.5 in the Ostensive, Non-ostensive, and Odor Control conditions, we first aggregated the data across trials for each individual and condition. We then conducted one-sample t-tests to compare the performance against chance.

To compare the performance between the test conditions, we fitted a GLMM with binomial error distribution and logit link function. We included the predictor variables condition, order of condition, trial number within condition, sex, age, and dogs’ trainability score based on the C-BARQ questionnaire. Additionally, we included the random intercept of subject ID and the random slopes of condition and trial number within subject ID. Note that, unlike the proposed study, this analysis did not include dog neuter status or lab ID in the model.

Confidence intervals for the predictors were derived based on 1,000 parametric bootstraps using a function kindly provided by Roger Mundry (based on the `bootMer()` function of the package *lme4*). To check for collinearity, we determined variance inflation factors (VIF) using the function `vif()` (R package *car*, Fox & Weisberg, 2019). Collinearity was not an issue, with a maximum VIF of 1.02 (VIF > 10 suggests strong collinearity, Quinn & Keough, 2002). To evaluate model stability, we dropped one level of the subject ID random effect at a time and compared the model estimates of the resulting models. This procedure revealed the model to be stable with respect to the fixed effects. Bayesian models used 4 chains with 12,000 iterations per chain (including 2,000 warm-up iterations).

Results

Performance Relative to Chance

The dogs (N = 61) performed significantly better than expected by chance in the Ostensive condition ($M = 0.60$, 95% CI $[0.55, 0.65)$, $t(60) = 4.41$, $p < .001$, $BF_{10} = 459.91$) but not in the Non-ostensive condition ($M = 0.53$, 95% CI $[0.49, 0.57)$, $t(60) = 1.47$, $p = .146$, $BF_{10} = 0.39$) or the Odor Control condition ($M = 0.46$, 95% CI $[0.41, 0.51)$, $t(60) = -1.45$, $p = .151$, $BF_{10} = 0.38$) (Figure 4).

Condition Comparison

The dogs were significantly more likely to choose the baited cup in the Ostensive condition compared to the Non-ostensive condition ($\chi^2(1) = 5.11$, p = .024, $BF_{10} = 3.88$) (Figure 4A). None of the control predictors (order of condition, trial number within condition, sex, age, C-BARQ trainability score) had any effect on dogs’ choices (Table 2).

Discussion

Our results from the preliminary data suggest that ostensive cueing plays an enhancing role in dogs’ ability to follow pointing gestures from humans: dogs successfully followed pointing gestures at above chance levels in the Ostensive condition but not in the Non-ostensive or Odor Control conditions, and they followed ostensively cued points significantly more often than non-ostensively cued ones. These results suggest that ostensive cues may be sufficient for dogs to successfully interpret and follow social gestures given by human informants. Conversely, dogs did not successfully follow all pointing gestures, suggesting that the mere presence of a human point is not interpreted by dogs as an imperative command, or as a sufficient associative cue for point following. This preliminary result is in line with previous research suggesting that dogs' performance on point following tasks improves when ostensive cues are present (Kaminski et al., 2012; Tauzin et al., 2015a; Tauzin et al., 2015b; but see also Scheider et al., 2013).

This finding supports the hypothesis that dogs view human pointing as a social communicative signal, not a simple association. Whether they view this communication as informative or imperative is less clear at the moment. The first theoretical account proposed in the introduction implies that dogs, like humans, perceive pointing as a *cooperative signal* (Hare & Tomasello, 2005) informing them where to find hidden food. The second account proposes that following communicative pointing results from the perception of the pointing gesture as an *imperative signal* ordering or, a somewhat weaker signal, suggesting where to go (Kaminski & Nitzschner, 2013; Scheider et al., 2013). Still, both accounts suggest that dogs understand the human gesture as a social signal, which contrasts with the simpler, asocial account of following the human hand (or finger) as a result of conditioning (Wynne et al., 2008; e.g., Dorey et al., 2010). Both the imperative hypothesis and the hand-food association hypothesis are supported by Delay (2016). By using an eye-tracker to determine dogs' looking behavior during the human pointing gesture, she found that dogs readily followed the movement of the pointing arm, but very rarely extended the signals further to the cups. In general, dogs looked at the experimenter's head-area the most. These results are therefore in line with Tauzin et al. (2015) suggesting that dogs perceive pointing as a spatial signal (where to go) rather than as a signal that refers to an object (e.g., Kaminski et al., 2012). More work will be needed to distinguish between the informative and imperative accounts. Some of our proposed and further exploratory analyses may begin to address these questions by looking at individual-level variation and the importance of training and trainability.

It is worth noting, however, that while the difference in dogs' performance across conditions in the preliminary data was statistically significant, this difference is subtle. This slight difference should also warn us against overestimating the role of ostensive signals. Contrary to the assumptions of the theory of (Human) Natural Pedagogy (Csibra & Gergely, 2006, 2009), the ostensive-communicative signals might simply capture the dog's attention slightly more of changing the perception and interpretation of the pointing gesture (the so-called genericity bias). Even in child studies, it has been shown that non-ostensive signals can have similar effects to ostensive signals, if only they are salient enough (Gredebäck et al., 2018). The so-called ostensive cues of direct gaze and dog-directed speech are perhaps particularly attention-grabbing stimuli and therefore we would need adequate controls for differences in covert attention before drawing firm conclusions (de Bordes et al., 2013; Szufnarowska et al., 2014). In general it is difficult to equate the salience of eye-gazing and dog-directed speech with a nonsocial stimulus. But staring at the eyes of another is a strong attention-getter for adult individuals in almost all social species (Emery, 2000). Our final, larger sample will allow for greater statistical power, and as a result, increased confidence in our conclusions. Greater confidence will not only be achieved through the increase in sample size, but also through increased variance in the sample, with different experimenters, and dog populations across a multitude of labs.

An additional benefit of the multi-lab approach proposed by ManyDogs is the potential to explore the role of individual differences such as training history, testing environment, breed, and age on dogs' ability to follow pointing gestures. Such analyses are difficult (if not impossible) to conduct in single-lab studies due to the lack of statistical power as well as the potential homogeneity of training history amongst dogs recruited from the same geographic area. The multi-lab approach allows for the sampling of dogs with a variety of training backgrounds and breeds. In addition to enabling these analyses of individual differences, sampling a more diverse population of dogs will likely result in more generalizable data, and thus, more externally valid conclusions.

Establishing Large-Scale Collaboration in Canine Science

Beyond contributions to theory via a multi-lab empirical replication of a key finding in the field, the present initiative provides an opportunity to establish an infrastructure to support future large-scale collaborations in canine science. In the process of designing and implementing the global participation stage of ManyDogs Study 1, we are bringing together diverse groups of scientists across multiple locations and research backgrounds. Our collectively built platform will facilitate open science practices such as pre-registration and creation of registered reports, both crucial features for promoting reproducibility. A key component to the success of the ManyDogs initiative will be opening formal and informal channels of communication (e.g., listserv, Slack) between labs that encourage involvement from researchers at all levels of their careers. To provide information to researchers across the globe, as well as the public, we have already established a website as a comprehensive information source and a social media presence on Twitter (@ManyDogsProject) to facilitate information dissemination.

Large-scale collaborations are necessary for answering questions that are out of the reach of a single laboratory. Such questions include individual differences in behavior on diagnostic cognitive tests as well as the role of culture and training norms on behavior – questions that have long been the subject of speculation without concrete means for adequately-powered assessment. With the creation of ManyDogs, we aim to address these foundational questions in the emerging field of canine science.

Author Contributions

Conceptualization: Emily E. Bray, Daphna Buchsbaum, Sarah-Elizabeth Byosiere, Molly Byrne, Julia Espinosa, Gitanjali E. Gnanadesikan, Daniel J. Horschler, Angie M. Johnston, Evan L. MacLean, Madeline H. Pelgrim, Laurie Santos, and Zachary A. Silver. **Data Curation:** Julia Espinosa, Jeffrey R. Stevens, and Christoph J. Völter. **Formal Analysis:** Daphna Buchsbaum, Evan L. MacLean, Jeffrey R. Stevens, and Christoph J. Völter. **Investigation:** Lena Zipperling. **Methodology:** Emily E. Bray, Daphna Buchsbaum, Sarah-Elizabeth Byosiere, Molly Byrne, Julia Espinosa, Gitanjali E. Gnanadesikan, Daniel J. Horschler, Ludwig Huber, Angie M. Johnston, Madeline H. Pelgrim, and Christoph J. Völter. **Project Administration:** Daphna Buchsbaum, Sarah-Elizabeth Byosiere, Julia Espinosa, Gitanjali E. Gnanadesikan, Madeline H. Pelgrim, and Jeffrey R. Stevens. **Resources:** Laurie Santos. **Software:** Julia Espinosa and Jeffrey R. Stevens. **Supervision:** Ludwig Huber and Christoph J. Völter. **Visualization:** Evan L. MacLean, Jeffrey R. Stevens, and Christoph J. Völter. **Writing - Original Draft Preparation:** Julia Espinosa, Marianne S. Freeman, Gitanjali E. Gnanadesikan, Angie M. Johnston, Evan L. MacLean, Madeline H. Pelgrim, Zachary A. Silver, Jeffrey R. Stevens, and Christoph J. Völter. **Writing - Review & Editing:** Emily E. Bray, Daphna Buchsbaum, Sarah-Elizabeth Byosiere, Julia Espinosa, Marianne S. Freeman, Gitanjali E. Gnanadesikan, C.-N. Alexandrina Guran, Daniel J. Horschler, and Ludwig Huber.

References

Agnoli, F., Fraser, H., Singleton Thorn, F., & Fidler, F. (2021). Australian and Italian psychologists' view of replication. *Advances in Methods and Practices in Psychological Science*, *4*(3), 1–15. doi:10.1177/25152459211039218

Aria, M., Alterisio, A., Scandurra, A., Pinelli, C., & D'Aniello, B. (2021). The scholar's best friend: Research trends in dog cognitive and behavioral studies. *Animal Cognition*, *24*(3), 541–553. doi:10.1007/s10071-020-01448-2

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001

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

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, *8*(6), 492–499. doi:10.1111/j.1467-7687.2005.00440.x

Bolker, B. (2014). Lme4 convergence warnings: troubleshooting. In *RPubs*. https://rpubs.com/bbolker/lme4trouble1.

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

Bray, E. E., Gnanadesikan, G. E., Horschler, D. J., Levy, K. M., Kennedy, B. S., Famula, T. R., & MacLean, E. L. (2021). Early-emerging and highly heritable sensitivity to human communication in dogs. *Current Biology*, *31*. doi:10.1016/j.cub.2021.04.055

Bray, E. E., Gruen, M. E., Gnanadesikan, G. E., Horschler, D. J., Levy, K. M., Kennedy, B. S., Hare, B. A., & MacLean, E. L. (2020a). Cognitive characteristics of 8- to 10-week-old assistance dog puppies. *Animal Behaviour*, *166*, 193–206. doi:10.1016/j.anbehav.2020.05.019

Bray, E. E., Gruen, M. E., Gnanadesikan, G. E., Horschler, D. J., Levy, K. M., Kennedy, B. S., Hare, B. A., & MacLean, E. L. (2020b). Dog cognitive development: A longitudinal study across the first 2 years of life. *Animal Cognition*, *24*(2), 311–328. doi:10.1007/s10071-020-01443-7

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi:10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. doi:10.32614/RJ-2018-017

Byers-Heinlein, K., Bergmann, C., Davies, C., Frank, M. C., Hamlin, J. K., Kline, M., Kominsky, J. F., Kosie, J. E., Lew-Williams, C., Liu, L., Mastroberardino, M., Singh, L., Waddell, C. P. G., Zettersten, M., & Soderstrom, M. (2020). Building a collaborative psychological science: Lessons learned from ManyBabies 1. *Canadian Psychology/Psychologie Canadienne*, *61*(4), 349–363. doi:10.1037/cap0000216

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), s13742-015-0047-8. doi:10.1186/s13742-015-0047-8

Consortium, M. (2020). Quantifying sources of variability in infancy research using the infant-directed-speech preference. *Advances in Methods and Practices in Psychological Science*, *3*(1), 24–52. doi:10.1177/2515245919900809

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141–168. doi:10.1111/j.1468-0017.2009.01384.x

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. doi:10.1016/j.tics.2009.01.005

Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In Y. Munakata & M. H. Johnson (Eds.), *Processes of Change in Brain and Cognitive Development. Attention and Performance, XXI* (pp. 249–274). Oxford University Press.

D’Aniello, B., Alterisio, A., Scandurra, A., Petremolo, E., Iommelli, M. R., & Aria, M. (2017). What’s the point? Golden and Labrador retrievers living in kennels do not understand human pointing gestures. *Animal Cognition*, *20*(4), 777–787. doi:10.1007/s10071-017-1098-2

de Bordes, P. F., Cox, R. F. A., Hasselman, F., & Cillessen, A. H. N. (2013). Toddlers’ gaze following through attention modulation: Intention is in the eye of the beholder. *Journal of Experimental Child Psychology*, *116*(2), 443–452. doi:10.1016/j.jecp.2012.09.008

Dorey, N. R., Udell, M. A. R., & Wynne, C. D. L. (2010). When do domestic dogs, *Canis familiaris*, start to understand human pointing? The role of ontogeny in the development of interspecies communication. *Animal Behaviour*, *79*(1), 37–41. doi:10.1016/j.anbehav.2009.09.032

Duranton, C., Range, F., & Virányi, Z. (2017). Do pet dogs (*Canis familiaris*) follow ostensive and non-ostensive human gaze to distant space and to objects? *Royal Society Open Science*, *4*(7), 170349. doi:10.1098/rsos.170349

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, *24*(6), 581–604. doi:10.1016/S0149-7634(00)00025-7

Farrar, B. G., Ostojić, L., & Clayton, N. S. (2021). The hidden side of animal cognition research: Scientists’ attitudes toward bias, replicability and scientific practice. *PLOS ONE*, *16*(8), e0256607. doi:10.1371/journal.pone.0256607

Farrar, B., Boeckle, M., & Clayton, N. (2020). Replications in comparative cognition: What should we expect and how can we improve? *Animal Behavior and Cognition*, *7*(1), 1–22. doi:10.26451/abc.07.01.02.2020

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third Edition). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Gredebäck, G., Astor, K., & Fawcett, C. (2018). Gaze following is not dependent on ostensive cues: A critical test of natural pedagogy. *Child Development*, *89*(6), 2091–2098. doi:10.1111/cdev.13026

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2020). bridgesampling: An R package for estimating normalizing constants. *Journal of Statistical Software*, *92*(10), 1–29. doi:10.18637/jss.v092.i10

Hare, B., Brown, M., Williamson, C., & Tomasello, M. (2002). The domestication of social cognition in dogs. *Science*, *298*(5598), 1634–1636. doi:10.1126/science.1072702

Hare, B., & Tomasello, M. (2005). Human-like social skills in dogs? *Trends in Cognitive Sciences*, *9*(9), 439–444. doi:10.1016/j.tics.2005.07.003

Hegedüs, D., Bálint, A., Miklósi, Á., & Pongrácz, P. (2013). Owners fail to influence the choices of dogs in a two-choice, visual pointing task. *Behaviour*, *150*(3/4), 427–443. doi:10.1163/1568539X-00003060

Hsu, Y., & Serpell, J. A. (2003). Development and validation of a questionnaire for measuring behavior and temperament traits in pet dogs. *Journal of the American Veterinary Medical Association*, *223*(9), 1293–1300. doi:10.2460/javma.2003.223.1293

Kaminski, J., & Nitzschner, M. (2013). Do dogs get the point? A review of dog-human communication ability. *Learning and Motivation*, *44*(4), 294–302. doi:10.1016/j.lmot.2013.05.001

Kaminski, J., Schulz, L., & Tomasello, M. (2012). How dogs know when communication is intended for them. *Developmental Science*, *15*(2), 222–232. doi:10.1111/j.1467-7687.2011.01120.x

Kirchhofer, K. C., Zimmermann, F., Kaminski, J., & Tomasello, M. (2012). Dogs (*Canis familiaris*), but not chimpanzees (*Pan troglodytes*), understand imperative pointing. *PLOS ONE*, *7*(2), e30913. doi:10.1371/journal.pone.0030913

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr., R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong, W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., … Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology*, *45*(3), 142–152. doi:10.1027/1864-9335/a000178

Kupán, K., Miklósi, Á., Gergely, G., & Topál, J. (2011). Why do dogs (*Canis familiaris*) select the empty container in an observational learning task? *Animal Cognition*, *14*(2), 259–268. doi:10.1007/s10071-010-0359-0

MacLean, E. L., Snyder-Mackler, N., vonHoldt, B. M., & Serpell, J. A. (2019). Highly heritable and functionally relevant breed differences in dog behaviour. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1912), 20190716. doi:10.1098/rspb.2019.0716

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*(6), 537–542. doi:10.1177/1745691612460688

Many Primates, Altschul, D. M., Beran, M. J., Bohn, M., Call, J., DeTroy, S., Duguid, S. J., Egelkamp, C. L., Fichtel, C., Fischer, J., Flessert, M., Hanus, D., Haun, D. B. M., Haux, L. M., Hernandez-Aguilar, R. A., Herrmann, E., Hopper, L. M., Joly, M., Kano, F., … Watzek, J. (2019). Establishing an infrastructure for collaboration in primate cognition research. *PLOS ONE*, *14*(10), e0223675. doi:10.1371/journal.pone.0223675

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

McShane, B. B., Tackett, J. L., Böckenholt, U., & Gelman, A. (2019). Large-scale replication projects in contemporary psychological research. *The American Statistician*, *73*(S1), 99–105. doi:10.1080/00031305.2018.1505655

Miklósi, Á., Polgárdi, R., Topál, J., & Csányi, V. (1998). Use of experimenter-given cues in dogs. *Animal Cognition*, *1*(2), 113–121. doi:10.1007/s100710050016

Moore, R., Mueller, B., Kaminski, J., & Tomasello, M. (2015). Two-year-old children but not domestic dogs understand communicative intentions without language, gestures, or gaze. *Developmental Science*, *18*(2), 232–242. doi:10.1111/desc.12206

Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of bayes factors for common designs*. https://CRAN.R-project.org/package=BayesFactor

Parker, H. G., Dreger, D. L., Rimbault, M., Davis, B. W., Mullen, A. B., Carpintero-Ramirez, G., & Ostrander, E. A. (2017). Genomic analyses reveal the influence of geographic origin, migration, and hybridization on modern dog breed development. *Cell Reports*, *19*(4), 697–708. doi:10.1016/j.celrep.2017.03.079

Pongrácz, P., Miklósi, Á., Timár-Geng, K., & Csányi, V. (2004). Verbal attention getting as a key factor in social learning between dog (*Canis familiaris*) and human. *Journal of Comparative Psychology*, *118*(4), 375–383. doi:10.1037/0735-7036.118.4.375

Quinn, G. P., & Keough, M. J. (2002). *Experimental design and data analysis for biologists*. Cambridge University Press. doi:10.1017/CBO9780511806384

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Range, F., Heucke, S. L., Gruber, C., Konz, A., Huber, L., & Virányi, Z. (2009). The effect of ostensive cues on dogs' performance in a manipulative social learning task. *Applied Animal Behaviour Science*, *120*(3-4), 170–178. doi:10.1016/j.applanim.2009.05.012

Rodriguez, K. E., Herzog, H., & Gee, N. R. (2021). Variability in human-animal interaction research. *Frontiers in Veterinary Science*, *7*, 619600. doi:10.3389/fvets.2020.619600

Scheider, L., Kaminski, J., Call, J., & Tomasello, M. (2013). Do domestic dogs interpret pointing as a command? *Animal Cognition*, *16*(3), 361–372. doi:10.1007/s10071-012-0577-8

Schmidjell, T., Range, F., Huber, L., & Virányi, Z. (2012). Do owners have a Clever Hans effect on dogs? Results of a pointing study. *Frontiers in Psychology*, *3*, 558. doi:10.3389/fpsyg.2012.00558

Serpell, J., & Hsu, Y. (2001). Development and validation of a novel method for evaluating behavior and temperament in guide dogs. *Applied Animal Behaviour Science*, *72*(4), 347–364. doi:10.1016/S0168-1591(00)00210-0

Soproni, K., Miklósi, Á., Topál, J., & Csányi, V. (2001). Comprehension of human communicative signs in pet dogs (*Canis familiaris*). *Journal of Comparative Psychology*, *115*(2), 122126. doi:10.1037/0735-7036.115.2.122

Stevens, J. R. (2017). Replicability and reproducibility in comparative psychology. *Frontiers in Psychology*, *8*, 862. doi:10.3389/fpsyg.2017.00862

Szetei, V., Miklósi, Á., Topál, J., & Csányi, V. (2003). When dogs seem to lose their nose: An investigation on the use of visual and olfactory cues in communicative context between dog and owner. *Applied Animal Behaviour Science*, *83*(2), 141–152. doi:10.1016/S0168-1591(03)00114-X

Szufnarowska, J., Rohlfing, K. J., Fawcett, C., & Gredebäck, G. (2014). Is ostension any more than attention? *Scientific Reports*, *4*(1), 5304. doi:10.1038/srep05304

Tauzin, T., Csík, A., Kis, A., Kovács, K., & Topál, J. (2015a). The order of ostensive and referential signals affects dogs' responsiveness when interacting with a human. *Animal Cognition*, *18*(4), 975–979. doi:10.1007/s10071-015-0857-1

Tauzin, T., Csík, A., Kis, A., & Topál, J. (2015b). What or where? The meaning of referential human pointing for dogs (Canis familiaris). *Journal of Comparative Psychology*, *129*(4), 334–338. doi:10.1037/a0039462

Téglás, E., Gergely, A., Kupán, K., Miklósi, Á., & Topál, J. (2012). Dogs' gaze following is tuned to human communicative signals. *Current Biology*, *22*(3), 209–212. doi:10.1016/j.cub.2011.12.018

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. doi:10.1017/S0140525X05000129

Topál, J., Gergely, G., Erdőhegyi, Á., Csibra, G., & Miklósi, Á. (2009). Differential sensitivity to human communication in dogs, wolves, and human infants. *Science*, *325*(5945), 1269–1272. doi:10.1126/science.1176960

Topál, J., Kis, A., & Oláh, K. (2014). Dogs’ sensitivity to human ostensive cues: A unique adaptation? In J. Kaminski & S. Marshall-Pescini (Eds.), *The Social Dog* (pp. 319–346). Academic Press. doi:10.1016/B978-0-12-407818-5.00011-5

Udell, M. A. R., Dorey, N. R., & Wynne, C. D. L. (2010). The performance of stray dogs (*Canis familiaris*) living in a shelter on human-guided object-choice tasks. *Animal Behaviour*, *79*(3), 717–725. doi:10.1016/j.anbehav.2009.12.027

Virányi, Z., & Range, F. (2009). How does ostensive communication influence social learning in dogs? *Journal of Veterinary Behavior*, *4*(2), 47. doi:10.1016/j.jveb.2008.10.023

Voelkl, B., Vogt, L., Sena, E. S., & Würbel, H. (2018). Reproducibility of preclinical animal research improves with heterogeneity of study samples. *PLOS Biology*, *16*(2), e2003693. doi:10.1371/journal.pbio.2003693

Wilson, A. J., Réale, D., Clements, M. N., Morrissey, M. M., Postma, E., Walling, C. A., Kruuk, L. E. B., & Nussey, D. H. (2010). An ecologist’s guide to the animal model. *Journal of Animal Ecology*, *79*(1), 13–26. doi:10.1111/j.1365-2656.2009.01639.x

Wobber, V., & Kaminski, J. (2011). What do dogs understand about human communicative gestures? A novel synthesis. In V. DeGiovine (Ed.), *Dogs: Biology, behavior and health disorders* (pp. 93–109). Nova Science Publishers.

Wynne, C. D. L., Udell, M. A. R., & Lord, K. A. (2008). Ontogeny’s impacts on human-dog communication. *Animal Behaviour*, *76*(4), e1–e4. doi:10.1016/j.anbehav.2008.03.010

Table 1. Consortium labs currently signed up to contribute data at the time of registered report submission (in alphabetical order by institutional affiliation).

	Lab Title	**Institutional Affiliation**
1	Canine Cognition Center	Boston College, Department of Psychology and Neuroscience
2	Brown Dog Lab	Brown University, Cognitive, Linguistic, & Psychological Sciences Department
3	Duke Canine Cognition Center	Duke University, Department of Evolutionary Anthropology
4	Thinking Dog Center	Hunter College, CUNY, Department of Psychology
5	IWU Dog Scientists	Illinois Wesleyan University, Department of Psychology
6	Human-Animal Interaction Lab	Oregon State University, Department of Animal and Rangeland Sciences
7	Pet Behaviour Consulting	Università degli Studi di Messina, Department of Veterinary Sciences
8	DogUP	Università degli Studi di Padova, Department of Comparative Biomedicine and Food Science
9	Arizona Canine Cognition Center	University of Arizona, School of Anthropology
10	Canine Cognition and Human Interaction Lab	University of Nebraska-Lincoln, Department of Psychology
11	Dog Cognition Centre	University of Portsmouth, Department of Psychology
12	The Clever Dog Lab	University of Veterinary Medicine Vienna, Messerli Research Institute
13	Canine Cognition Center	Yale University, Department of Psychology

Table 2:

Results of GLMM of the dogs' choice performance

	Estimate	SE	Lower CI	Upper CI	χ^2	df	p	BF_{10}
(Intercept)	0.12	0.13	-0.14	0.39				
Condition	0.30	0.13	0.04	0.57	5.11	1	0.02	3.88
Order of condition	0.06	0.13	-0.22	0.29	0.18	1	0.67	0.38
Trial number	-0.10	0.07	-0.23	0.03	2.32	1	0.13	0.51
Sex	-0.08	0.14	-0.36	0.18	0.37	1	0.54	0.43
Age	-0.01	0.07	-0.15	0.13	0.03	1	0.86	0.18
C-BARQ trainability score	0.09	0.07	-0.03	0.23	1.96	1	0.16	0.45

Note. Reference categories—condition: Non-ostensive condition; order of condition: Non-ostensive condition first; sex: female; covariates trial number, age, and training experienced were centered and scaled to a standard deviation of 1. The standard deviations for the contribution of the random effects were 0.099 for the random intercept of subject, 0.159 for the random slope of condition within subject, and 0.063 for the random slope of trial number within subject.

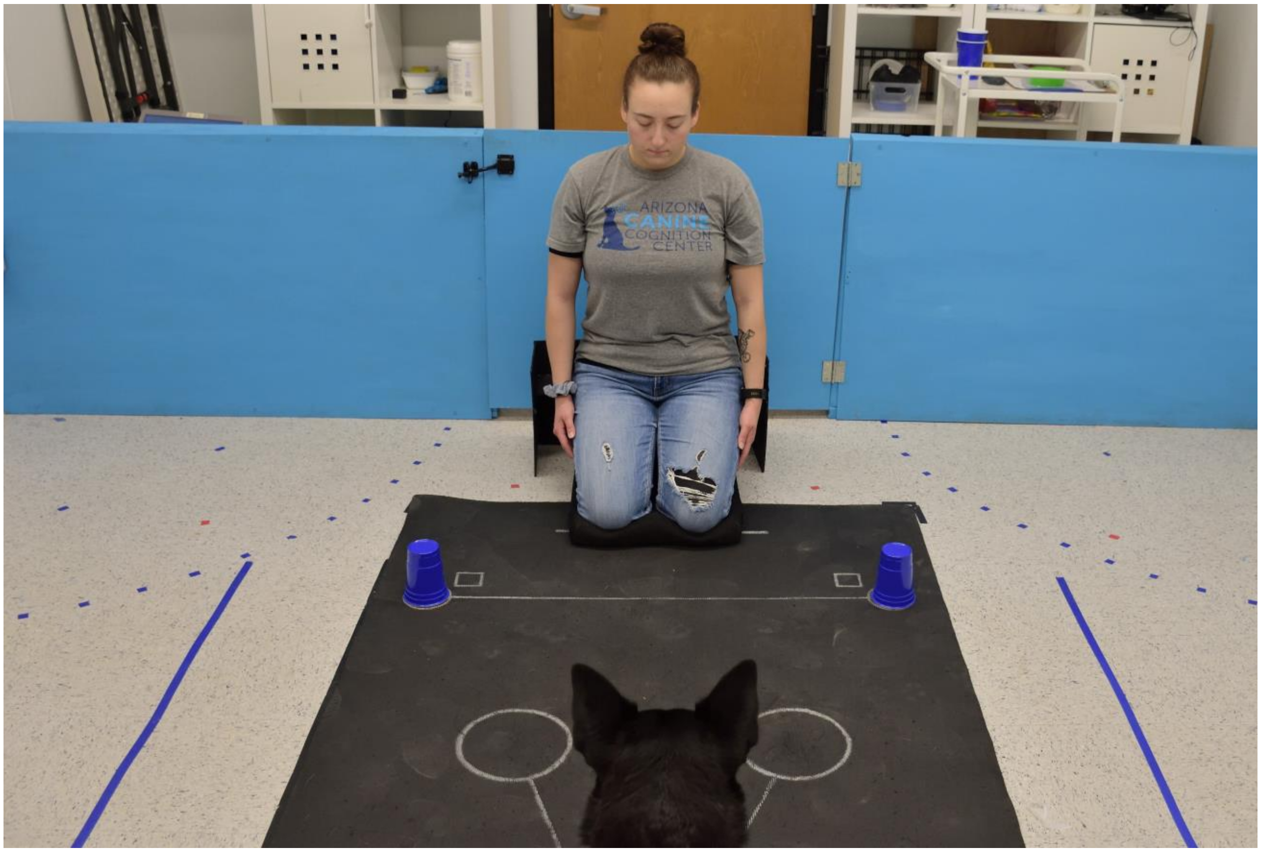


Figure 1. Dogs will be approximately 1.35 m away from the experimenter, centered between the two cup locations for all test trials and for one- and two-cup warm-ups.

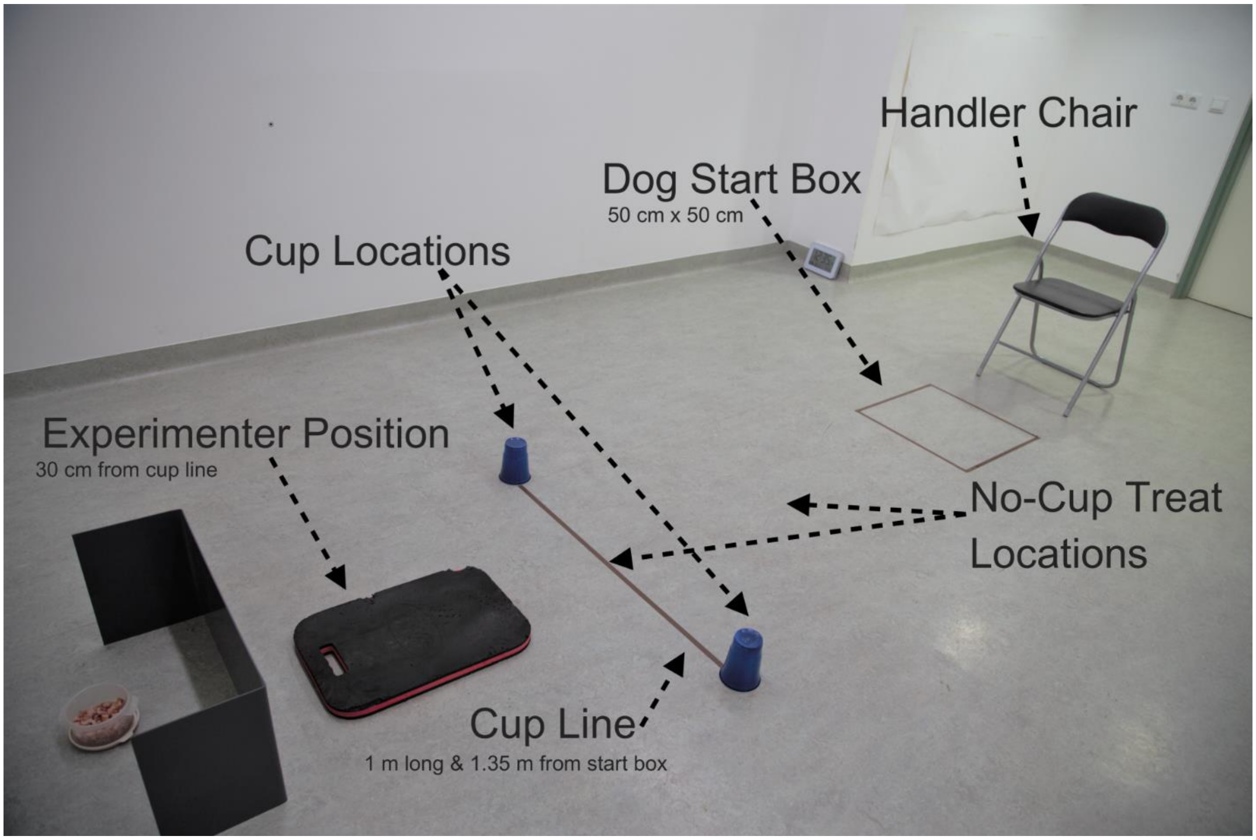
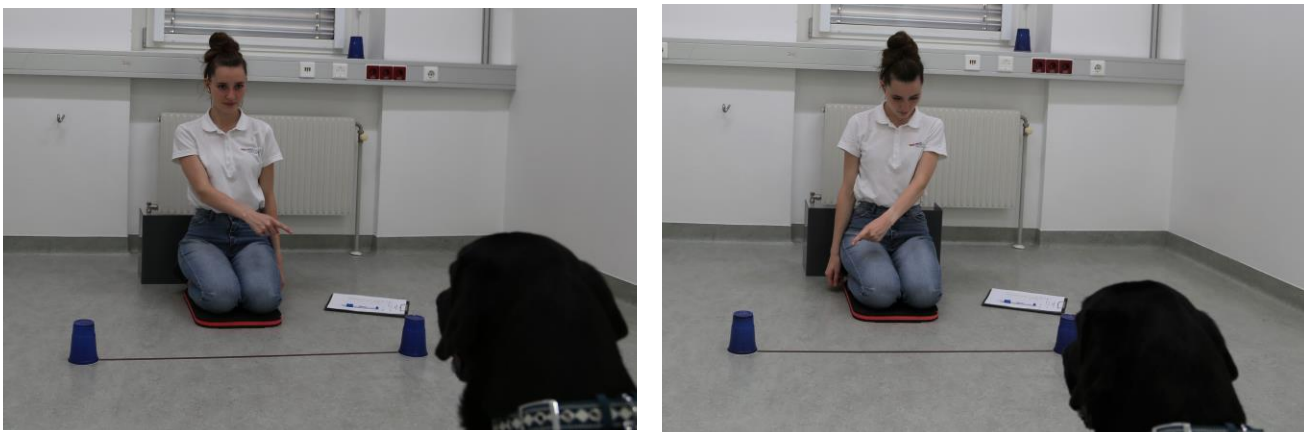


Figure 2. Photograph of the testing environment with measurements between stimuli, and experimenter, handler, and dog positions.

(a)



(b)

Figure 3. Ostensive and non-ostensive cues. (a) In the Ostensive condition, experimenters make eye contact with the dog and say the dog’s name. (b) In the Non-ostensive condition, experimenters look to the ground and clear their throat to get the dog’s attention.

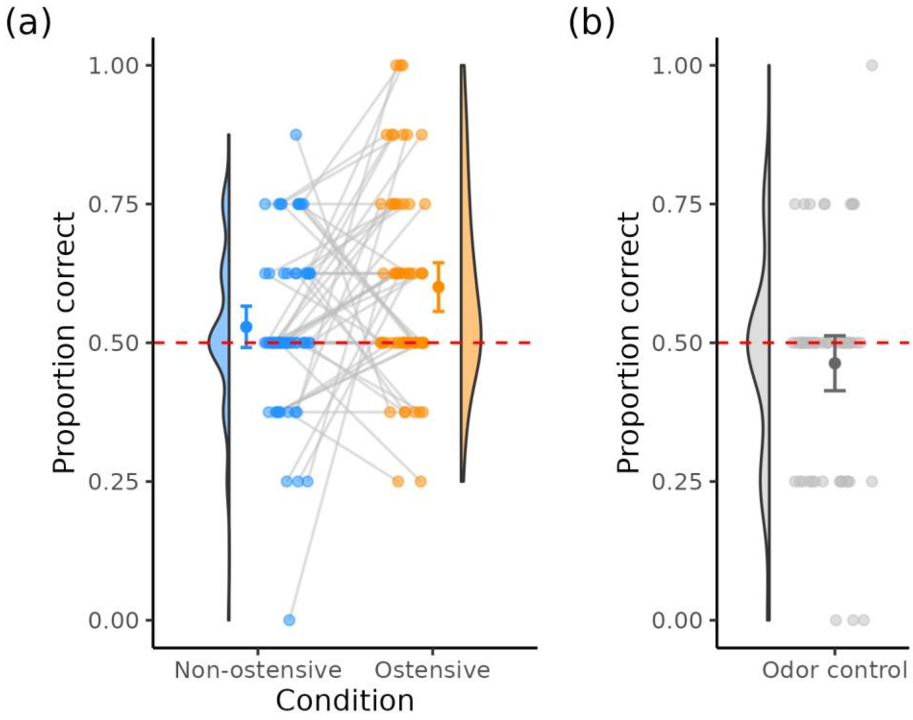


Figure 4. Violin and dot plot of dogs' performance (N=61) across the Non-ostensive and Ostensive conditions (a) and the Odor Control condition (b) of preliminary data. The red dashed lines show the chance level of 0.5. Dots represent the mean proportion correct for each individual. The grey lines connect dots representing the same individuals. The error bars represent 95% within-subjects confidence intervals; the filled circles on top of the error bars show the fitted model.