Animal Behavior and Cognition
ABC 2023, 10(3): 232-286
DOI: https://doi.org/10.26451/abc.10.03.03.2023

.

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

ManyDogs Project†, Julia Espinosa[1,*], Jeffrey R. Stevens[2,3], Daniela Alberghina[4], Harley E. E. Alway[5], Jessica D. Barela[2,3], Michael Bogese[6], Emily E. Bray[7,8], Daphna Buchsbaum[9], Sarah-Elizabeth Byosiere[6], Molly Byrne[10], Camila M. Cavalli[11], Leah M. Chaudoir[12,13], Courtney Collins-Pisano[14], Hunter J. DeBoer[2], Laura E. L. C. Douglas[8], Shany Dror[15], Marina V. Dzik[11], Beverly Ferguson[16], Laura Fisher[16], Hannah C. Fitzpatrick[2], Marianne S. Freeman[17], Shayla N. Frinton[18], Maeve K. Glover[19], Gitanjali E. Gnanadesikan[12,13], Joshua E. P. Goacher[20], Marta Golańska[21], C.-N. Alexandrina Guran[22,23], Elizabeth Hare[24], Brian Hare[25], Mia Hickey[26], Daniel J. Horschler[27], Ludwig Huber[28], Hoi-Lam Jim[29,20], Angie M. Johnston[10], Juliane Kaminski[20], Debbie M. Kelly[30], Valerie A. Kuhlmeier[19], Lily Lassiter[27], Lucia Lazarowski[31], Jennifer Leighton-Birch[20], Evan L. MacLean[7,32], Kamila Maliszewska[21], Vito Marra[4], Lane I. Montgomery[14], Madison S. Murray[6], Emma K. Nelson[25], Ljerka Ostojić[33,34], Shennai G. Palermo[10], Anya E. Parks Russell[10], Madeline H. Pelgrim[9], Sarita D. Pellowe[35,36], Anna Reinholz[21], Laura A. Rial[11], Emily M. Richards[27], Miriam A. Ross[9], Liza G. Rothkoff[6], Hannah Salomons[25], Joelle K. Sanger[2], Laurie Santos[27], Angelina R. Schirle[8], Shania J. Shearer[17], Zachary A. Silver[27], Jessica M. Silverman[18], Andrea Sommese[15], Tiziana Srdoc[37,33], Hannah St. John-Mosse[20], Angelica C. Vega[6], Kata Vékony[15], Christoph J. Völter[28], Carolyn J. Walsh[5], Yasmin A. Worth[2], Lena M. I. Zipperling[28], Bianka Żołędziewska[21], and Sarah G. Zylberfuden[9]

[1]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA
[2]Department of Psychology, University of Nebraska-Lincoln, Lincoln, NE, USA
[3]Center for Brain, Biology & Behavior, University of Nebraska-Lincoln, Lincoln, NE, USA
[4]Department of Veterinary Sciences, Messina University, Messina, Italy
[5]Department of Psychology, Memorial University of Newfoundland, St. John's, NL, Canada
[6]Department of Psychology, Hunter College, City University of New York, New York City, NY, USA
[7]College of Veterinary Medicine, University of Arizona, Tucson, AZ, USA
[8]Canine Companions for Independence, National Headquarters, Santa Rosa, CA, USA
[9]Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI, USA
[10]Department of Psychology and Neuroscience, Boston College, Chestnut Hill, MA, USA
[11]Canid Behavior Research Group, Instituto de Investigaciones Médicas, University of Buenos Aires, National Scientific and Technical Research Council, Buenos Aires, Argentina
[12]School of Anthropology, University of Arizona, Tucson, AZ, USA
[13]Cognitive Science Program, University of Arizona, Tucson, AZ, USA
[14]Department of Psychological Sciences, Auburn University, Auburn, AL, USA
[15]Department of Ethology, Eötvös Loránd University, Budapest, Hungary
[16]Leader Dogs for the Blind, Rochester, MI, USA
[17]Animal Health and Welfare Research Centre, University Centre Sparsholt, Hampshire, UK
[18]Department of Psychology, University of Victoria, Victoria, BC, Canada
[19]Department of Psychology, Queen's University, Kingston, ON, Canada
[20]Centre for Comparative and Evolutionary Psychology, Department of Psychology, University of Portsmouth, Portsmouth, UK
[21]Faculty of Psychology, University of Warsaw, Warsaw, Poland
[22]Vienna Cognitive Science Hub, University of Vienna, Vienna, Austria
[23]Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, Vienna, Austria
[24]Dog Genetics LLC, Astoria, NY, USA
[25]Department of Evolutionary Anthropology, Duke University, Durham, NC, USA
[26]Psychology Department, University of Arizona, Tucson, AZ, USA

[27]Department of Psychology, Yale University, New Haven, CT, USA
[28]Comparative Cognition, Messerli Research Institute, University of Veterinary Medicine Vienna, Medical University of Vienna and University of Vienna, Vienna, Austria
[29]Institute for Advanced Study, Kyoto University, Kyoto, Japan
[30]Department of Psychology, University of Manitoba, Winnipeg, MB, Canada
[31]Canine Performance Sciences, Auburn University College of Veterinary Medicine, Auburn, AL, USA
[32]Department of Psychology, University of Arizona, Tucson, AZ, USA
[33]Department of Psychology and Division of Cognitive Sciences, Faculty of Humanities and Social Sciences, University of Rijeka, Rijeka, Croatia
[34]Centre for Mind and Behaviour, University of Rijeka, Rijeka, Croatia
[35]Department of Cognitive and Behavioural Ecology, Memorial University of Newfoundland, St. John's, NL, Canada
[36]East Coast Canine Dog Training, St John's, NL, Canada
[37]Faculty of Philosophy and Education, University of Vienna, Vienna, Austria

*Corresponding author (Email: manydogsproject@gmail.com; jespinosa@g.harvard.edu)

†The ManyDogs Project is an international consortium of research teams working on reproducible research in the field of canine science. Contact the ManyDogs Project at manydogsproject@gmail.com. More information about the consortium is available at www.manydogs.org

Citation – ManyDogs Project, Espinosa, J., Stevens, J. R., Alberghina, D., Always, H. E. E., Barela, J. D., Bogese, M., Bray, E. E., Buchsbaum, D., Byosiere, S-E., Byrne, M., Cavalli, C. M., Chaudoir, L. M., Collins-Pisano, C., DeBoer, H. J., Douglas, L. E. L. C., Dror, S., Dzik, M. V., Ferguson, B., … Zylberfuden, S. G. (2023). ManyDogs 1: A Multi-Lab Replication Study of Dogs' Pointing Comprehension. *Animal Behavior and Cognition, 10*(3), 232-286. https://doi.org/10.26451/abc.10.03.03.2023

Abstract – To promote collaboration across canine science, address replicability 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 pre-registering 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 reliable—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 data (N = 455) from 20 research sites on two conditions of a 2-alternative object choice task: (1) Ostensive (pointing to a baited cup after making eye-contact and saying the dog's name); (2) Non-ostensive (pointing without eye-contact, after a throat-clearing auditory control cue). Comparing performance between conditions, while both were significantly above chance, there was no significant difference in dogs' responses. This result was consistent across sites. Further, we found that dogs followed contralateral, momentary pointing at lower rates than has been reported in prior research, suggesting that there are limits to the robustness of point-following behavior: not all pointing styles are equally likely to elicit a response. Together, these findings underscore the important role of procedural details in study design and the broader need for replication studies in canine science.

Keywords – Domestic dog; Replicability; Human pointing; Social cognition; Interspecific interaction; Object choice task

---

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, with successful replications 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 or editorial biases that deter scientists from attempting replication work (Neuliep, 1990; 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.

A number of consortium projects have begun to address replication issues in various psychological sciences, including social psychology (Klein et al., 2014), primate cognition (ManyPrimates et al., 2022) and developmental psychology (The 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 replicability in the subfield and identify factors that influence replicability.

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). Similar 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 replicability within the field of canine science, we have developed a new consortium project: ManyDogs (ManyDogs Project et al., 2022). Drawing inspiration from other ManyX projects (e.g., ManyBabies, ManyLabs, ManyPrimates), the primary goals of the 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 pre-registering 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 reliability 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, with a more moderate effect size than what was reported in primary research (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 that statistical tests performed on large data sets tend to have large statistical power, 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 replicability 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 replicability 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 used a “single study” approach, in which one specific protocol was implemented by all participating labs in parallel. This approach was modeled after the ManyBabies project, and since many of the logistical concerns of infant research are like those found in canine research, this approach provided the 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 must 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 difficult to pinpoint 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 provides the opportunity for us to directly investigate some of these sources of variability.

For our first study, we collectively chose to investigate dogs’ interpretation of human pointing gestures. Point-following behavior in dogs is of the earliest findings in canine science that catalyzed the growth of the field, particularly because they seem to respond 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; Hare et al., 2002; Kaminski & Nitzschner, 2013), even from a very young age (Bray, Gnanadesikan, et al., 2021), though factors such as rearing environment and living conditions may influence point-following behavior (Udell et al., 2010; D’Aniello et al., 2017). However, researchers still disagree as to whether dogs show this behavior because they interpret human pointing as a social, communicative gesture or whether they simply associate human hands or limbs with food. Under the association explanation, 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.

Under the social communicative explanation, pointing gestures convey information from the signaler to the observer. Pointing is 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 human pointing is received and understood, both involve social signals. 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 (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). Here, dogs were more likely to follow the pointing gesture when the experimenter was making eye contact than when she was not. In fact, dogs in the condition without ostensive eye contact did not follow the pointing gesture above what would be expected by chance, 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., 2015). Dogs were more likely to follow pointing gestures when the ostensive cues preceded the pointing than when 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 ostensive cues are necessary for dogs to follow 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.

To study point-following behavior further, and assess the feasibility of the ManyDogs approach, we chose 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? We designed an experiment that could be carried out at most canine research sites and was intended for widespread global participation, we explored dogs' responses in two different pointing conditions: an Ostensive condition (pointing with eye-contact and dog-directed speech) and a Non-ostensive condition (pointing with averted gaze and throat-clearing control cue). By investigating dogs' responses to these two contrasting pointing contexts 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.

Hypotheses and Predictions

Our main hypothesis was that ostensive cues preceding a human pointing gesture have a facilitating effect on dogs' following of human pointing gestures. We predicted 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. Prior to conducting this study, we proposed that if we found 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 was observed between conditions, this could suggest that dogs understand pointing as the result of a learned gesture-reward association.

A second hypothesis was related to 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 (Szetei et al., 2003; Scheider et al., 2013), although this may also be analogous to human infants, as explained by the Natural Pedagogy account (Csibra & Gergely, 2009). Unlike the informative account, we had 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 could not definitively answer this question. However, we expected 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 was that, as has previously been demonstrated in similar paradigms (Bray et al., 2020; Bray, Gnanadesikan, et al., 2021), dogs would not use olfactory cues to find hidden food in this task, and thus we would not see group level performance that is significantly above chance in the Odor Control condition.

Methods

Ethics Statement

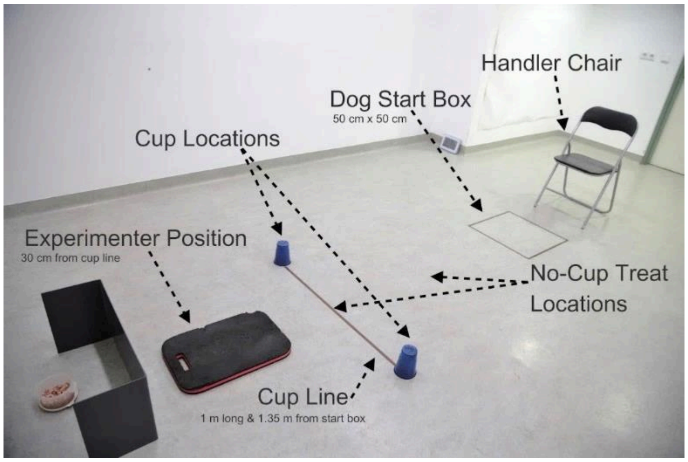
Each participating research site obtained explicit ethical approval or a letter of exemption to carry out the study from their respective institution research ethics committee prior to implementing the protocol (Table S1). Sites at which the protocol was deemed exempt from institution oversight were provided with letters from the institution research ethics committee stating the experimental protocol complied with the country's national animal protection legislation governing non-invasive animal research.

Protocol

Here, we present the study design implemented across multiple sites to address our research questions. Videos of pointing conditions and the pre-registration of the methods are available as on the Open Science Framework (https://osf.io/9r5xf/) and the detailed experimental procedure is included in the Supplemental Materials. We selected an object-choice paradigm, based on methods by Bray et al. (2020) and Bray, Gruen, et al. (2021), 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) were closely based on Bray et al.'s (2020) 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 (Figure 1). The cups were opaque and false-baited with a treat taped to the inside of each cup to control for odor cues in all trials except the Odor Control condition, which used clean un-baited cups. Subjects were allowed up to 25 s to choose a cup on each trial. A choice was 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 did not make a choice within 25 s, a "no-choice" was recorded and the trial repeated. If the subject made two no-choice responses in a row, they underwent refamiliarization prior to reattempting to complete the warm-up phase or test trials (see refamiliarization procedure below).

Figure 1

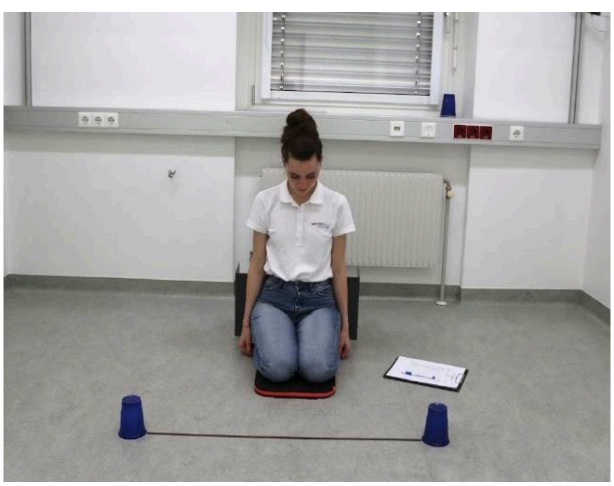
Experimental Set-Up and Stimuli for ManyDogs 1



The protocol recommended that the handler be seated in a chair behind the dog during the study, holding the dog stationary and facing toward the experimenter while the baiting was carried out. This was allowed to vary by site when necessary. For example, if the size of the testing room did not allow space for a chair, then the handler would stand behind the dog. The experimenter was always a trained researcher; they maintained a seated position during trials and looked at the floor during the entirety of each choice period to avoid cueing the subject (Figure 2). The handler was either a trained researcher or the dog's guardian, in accordance with a given site's typical research practices. In cases where the guardian was not handling during the study, we recommended (but did not require) that they remain in the room, seated behind the handler. To minimize the potential for unintentional cueing, the protocol stipulated that trained handlers should bow their head and close their eyes during baiting and cueing (opening them only once the dog has been released), while guardian handlers were asked to close their eyes for the entirety of the trial duration. We believed that this measure would sufficiently ensure that dogs were 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 reliable than is often suggested (Schmidjell et al., 2012; Hegedüs et al., 2013).

Figure 2

Experimenter in Resting Position, Sitting on their Legs with their Hands by the Sides of the Thighs, and Looking Down



Warm-up

Procedure

To familiarize subjects with the experimenter, the testing space, and the task of finding food under cups, several phases of warm-ups were conducted. These warm-ups were 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, Gnanadesikan, et al., 2021 found that performance on warm-ups was not predictive of performance on a pointing task). Throughout the warm-up phases, dogs were spoken to in a high-pitched voice using pet-directed speech to facilitate attention to the stimuli (Ben-Aderet et al., 2017; Jeannin et al., 2017); additionally, experimenters attempted to make eye contact with subjects at the beginning of each trial when showing them the food reward and both handler and experimenter praised the dog for making a choice. All cups used for warm-ups were false-baited to ensure that the cups smelled like food and to minimize dogs' ability to choose cups based only on odor. Subjects proceeded to test trials after completing all phases of the warm-ups, or after 15 minutes had elapsed from beginning the first phase of warm-ups. If, during warm-ups, subjects did not respond on two consecutive trials they underwent refamiliarization with the previous phase to encourage participation. Exclusion and abort criteria are detailed in the section below.

Warm-up Phase 1: Visible Placement and Free-form Cup Association

Warm-ups began with at least two repetitions of visible treat placement on the floor in front of the experimenter to ensure the subject was willing to approach the experimenter and eat off the floor in the testing area. Additional trials were used as necessary if a dog did not immediately eat the food. After the subject retrieved the treat successfully from each visible placement, the experimenter initiated 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 showed a single treat before placing it on the floor and covering it with a cup. The experimenter vocally encouraged the subject to approach and touch the cup, then rewarded them with the treat underneath. This hiding process was repeated at least three times or until the subject readily touched the cup.

Warm-up Phase 2: One-cup Alternating

The second warm-up phase familiarized the subject with the setup and general trial procedure to ensure they were willing to approach the cup locations to the right and left of the experimenter (Figure 2). In this phase, only one cup was presented in each trial and placed at either the right or left of the experimenter in one of the two designated cup positions, which were 1 m apart from each other along a line 1.35 m in front of the dog's starting box (Figures 1, 2). At the start of each trial, the reward was visibly placed under the cup; the experimenter attempted to make eye contact with the dog as they baited the cup. The subject was 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 was presented on the opposite side to ensure the subject received two rewards in each location. Subjects who did not complete four touches within seven trials were excluded (see refamiliarization and abort criteria below). On every trial (true of all trial types throughout the study), subjects were allowed to make only one choice (i.e., touch) and were rewarded on trials where they touched the baited cup first. Upon choosing, the experimenter lifted the cup, exposing the treat for the subject to eat.

Warm-up Phase 3: Two-cup Alternating

The third warm-up phase ensured that the subject attended to the experimenter's actions, was willing to approach both cup locations when a cup was present at each location simultaneously (i.e., not side-biased), and was not choosing randomly. These trials were identical to the previous phase, except that two indistinguishable cups were used (Figure 1), such that the subject needed to attend while one cup was baited by the experimenter in order to choose correctly. The experimenter attempted to make eye contact with the dog as they visibly baited 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) were counterbalanced across the conditions and within a site (i.e., each sequence used four times within the minimum sample of subjects). Subjects were 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 did not choose correctly were immediately repeated to minimize side biases. Subjects that did not meet the criterion within 20 total trials (including repeated trials) were excluded. The experimental setup is shown in Figure 1.

Test Trials

The test trials included two blocks of eight trials each—one block for each of the two conditions (Ostensive vs. Non-ostensive)—with the order counterbalanced across individuals tested at each site. The two blocks were separated by a one-minute play break and a re-familiarization (two trials of the two-cup alternating procedure from the warm-up Phase 3).

In both test conditions, occluded baiting was used, and each trial began with the experimenter placing an occluder in front of themself, hiding the two cups from the subject's view. As in the warm-ups, both cups were false-baited to minimize the dogs' ability to use odor cues in the task. The experimenter showed the subject the food reward by holding it out centrally in front of their torso, then placed it underneath one of two cups hidden behind the occluder (standardized occluder size across sites: 30 cm tall x 58 cm wide). Then the experimenter removed the occluder and placed it behind them, after which they simultaneously slid the two cups outward from their central position until they were 1 m apart. With the cups in position, the experimenter provided one of the pointing cues (described below). Across conditions, experimenters used a contralateral momentary point, holding the point stationary for 2 s before returning to the resting position and maintaining a downward gaze while the dog made a choice. Although there was some necessary variation across experimenters due to individual anatomy, the experimenter's finger was positioned approximately 30 cm from the cup during the pointing cue. Once in resting position, and after waiting for 1 s, experimenters cued the handler to release the subject using a neutral word ("now") and neutral tone to avoid additional social cueing from the experimenter. The handler released the subject by dropping the leash and saying "okay!" or any similar release command that the subject was familiar with. The dog was able to choose one cup per trial and was prevented from making a second choice by the experimenter removing the cups or blocking the dog's access. If the dog chose the baited cup, they were rewarded with the hidden food; if they chose the unbaited cup, they were shown the empty space under the cup and no reward was given. On test trials, the experimenter did not praise the dog for choosing the baited cup. The handler did not praise the dog for their cup choice, but did praise the dog upon recall to the starting position. Except for the gesturing components, detailed below, all other aspects of the test trials were identical in both conditions.

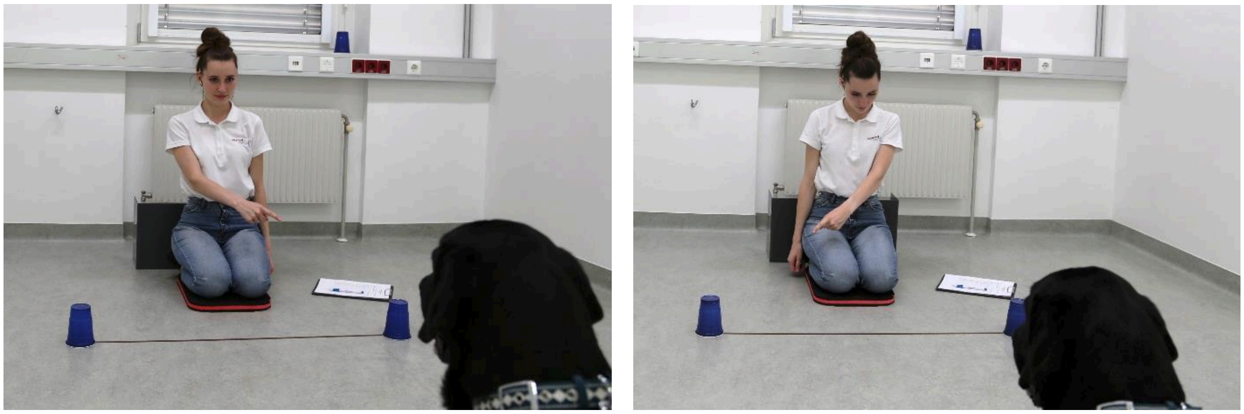
The primary dependent measure for each test condition was the proportion of trials in which the subject chose the baited cup. Subjects had 25 s to make a choice on each trial, and they were required to complete all test trials of both pointing conditions to be included in pre-registered analyses. Individual exclusion criteria are detailed below.

Ostensive Condition

At the start of each Ostensive trial, the experimenter made eye contact with the subject and said "[dog name], look!" in high-pitched pet-directed speech, while visibly presenting the treat. After the treat placement, cup movement, and occluder removal, the experimenter again repeated "[dog name], look!" in pet-directed speech and made eye contact before presenting the pointing gesture (Figure 3). While giving the neutral release signal and while the subject approached, the experimenter looked down at the floor directly in front of them.

Figure 3

Experimenter Demonstrating the Ostensive and Non-Ostensive Pointing Gestures Used in the Test Conditions



Note. Ostensive pointing gestures shown left, and non-ostensive pointing gestures shown right.

Non-Ostensive Condition

At the start of each Non-ostensive trial, the experimenter looked down and cleared their throat to get the subject's attention while presenting the treat. Before pointing, the experimenter cleared their throat again to attract the subject's attention and kept their gaze on the ground in front of them while they presented the momentary pointing gesture, and while the subject approached and made a choice (Figure 3). 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 did not speak to the dog during the Non-ostensive trials, only uttering the neutral "now" as a cue for the handler to release the dog.

Odor Control Condition

After both blocks of test trials, another one-minute play break took place. Finally, in the four odor control trials, the cups were baited identically to the test trials, except: (1) clean cups were used, without a treat taped into the cup (thus making it easier for subjects to potentially use scent cues if they were using an olfactory search strategy), (2) only one verbal cue was given when presenting the treat, "[dog name], look," and (3) no pointing gesture was provided before the subject was released to search. Based on previous results with similar paradigms by Bray, Gruen et al. (2021), we expected most subjects to perform at chance levels on these trials. We therefore used a reduced number of odor control trials to avoid dogs becoming discouraged and refusing to participate. This data was not intended for use on an individual level to exclude subjects, but rather for post-hoc analyses to investigate dogs' ability to use olfactory information, or other unintentional cues, at the level of site, breed, or training background.

Refamiliarization and Abort Criteria

If subjects stopped 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 was used. This involved 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 Supplemental Methods for details of this procedure). If refamiliarization with a previous phase did not successfully re-engage the subject in the task, or if the subject made a total of four no-choice responses in any single phase of the warm-ups or test trials, *or* if the subject exhibited signs of distress, testing was aborted. One exception to the abort rule was allowed if the subject participated in the Non-ostensive pointing condition first and reached 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 allowed 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 did 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

Cup choices were coded live by the experimenter. Additionally, videos were recorded to enable inter-rater reliability coding (re-coding) after the fact. For each participating site, upon completing data collection a subset of 8 subjects was randomly selected for re-coding from their sample that met inclusion criteria and all test trials of each subject were coded to check against the original data record. All re-coding was done from video by a research assistant who was blind to the hypothesis of the project. We set an inter-rater reliability threshold of $\kappa \geq 0.9$ for individual sites. The raters who re-coded a subset of the trials had very high reliability with the original coding for choice ($\kappa = 0.98$, 95% CI [0.97, 0.98], $N = 2486$). Individual site reliability ranged from $\kappa = 0.92$-1.00.

Survey Data

Prior to participation in the behavior study, dog owners and guardians completed a survey on their dog's background, including breed, training history, and other demographics. Dog owners and guardians also completed the trainability scale of the Canine Behavioral Assessment and Research Questionnaire (C-BARQ©, www.cbarq.org) (Serpell & Hsu, 2001; Hsu & Serpell, 2003). All C-BARQ items were made available to the guardian, though only the trainability scale was required for our pre-registered analyses. See pre-registered 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.

Research Sites

This experiment was conducted at 21 different research sites across nine countries on three continents (Table S1). Sites were able to 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 or official exemption from their institution, and (3) collect data from at least 16 subjects that completed the study. After piloting the protocol in 2020 at one site, and COVID-19 related delays, we initiated data collection for the study in late October, 2021. We put out an open call on social media and on email lists to advertise the project, and we ultimately accepted data submissions for the main experiment from 20 research sites, closing the data collection period at the end of January, 2023. We set a minimum number of dogs per site at $n = 16$ to allow for an assessment of between-lab variation in performance; there was no upper limit on the number of dogs that a site could test.

Collaborator Onboarding Process

We used an online survey (hosted on Qualtrics) to recruit research sites to contribute data for ManyDogs 1. Upon completion of this survey, the onboarding process was initiated; one of the ManyDogs administrative team corresponded closely with the new collaborator to assist them with obtaining ethics approval and with registering their research site in our database. The information that we collected about each site included a detailed floor plan of the area in which the collaborators collected data, along with details about sound attenuation, room ventilation, if they are using personal protective equipment (PPE) and if so what type (e.g., face masks), their research assistant training process, and general information about the population from which they recruited individual participants. Sites were able to pursue a variety of data collection strategies and were free to recruit from different populations, including family pets or working dogs, and also to run the test in the environment that best suited their team, e.g., outside, in a dog facility, at the guardian's home, or at their private research site.

To strive for a high level of similarity in how different sites implemented the protocol, we designed an experimenter training process to be completed by each site prior to data collection. In the training, sites were required to submit two rounds of videos. The initial video focused on the trained experimenter performing each phase of the study protocol. The second video was of the first participant that had completed all phases of the study at the site. ManyDogs administrators provided detailed written feedback on how to improve protocol execution, and virtual coaching meetings were scheduled as necessary. The video submission-feedback cycle could be repeated as necessary to achieve consistency and uniformity in the protocol. The second type of training instructed collaborators on how to carry out the data entry process, which we designed to be completed through surveys hosted on Qualtrics. Researchers practiced the data entry process with pre-prepared practice coding sheets, receiving feedback and repeating the steps as necessary. Upon completing both types of training, research sites were given an explicit recommendation to begin collecting data and encouraged 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 project administrators), we maintained an active Slack workspace with designated channels for open discussion of progress and troubleshooting in all aspects of participation in the study.

Pilot Experiment

In order to validate our study design and analysis plan, in the summer of 2020 we collected preliminary data from a pilot experiment as part of a master's thesis 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.

Ethics Statement

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).

Pilot Sample

Ninety-one dogs (M:F = 38:53, mean ± SD age = 5.1±3.3) across a variety of breeds participated in the pilot experiment. Of these, a subset of 61 dogs (M:F = 26:35, mean ± SD age = 4.7±3.3 years [range = 0-12]) were tested after our pre-registration was submitted; all statistical models using pilot data are limited to these individuals. None of the pilot data was included in the main experiment analyses below. An additional 12 dogs started but did not complete the experiment due to lack of motivation (n = 10) or fear/anxiety (n = 2).

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 ($\kappa = 1$, $N = 360$).

Pilot 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 Generalized Linear Mixed Model (GLMM) with a 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).

Pilot Results

In the pilot experiment, we tested 61 dogs (M:F = 26:35, mean ± SD age = 4.7±3.3 years [range = 0-12]). Approximately 41% of the dogs were spayed or neutered, 98.4% were purebred, and all lived in private homes.

Performance Relative to Chance

The dogs ($N = 61$) performed better than expected by chance in the Ostensive condition ($M = 0.60$, 95% CI [0.55, 0.65], $t(60) = 4.41$, $p < .001$, $BF_{10} = 459.9$) but not in the Non-ostensive condition (Mean = 0.53, 95% CI [0.49, 0.57], $t(60) = 1.47$, $p = .146$, $BF_{10} = 0.39$) or the Odor Control condition (Mean = 0.46, 95% CI [0.41, 0.51], $t(60) = -1.45$, $p = .151$, $BF_{10} = 0.38$) (Figure S6).

Condition Comparison

The dogs chose the baited cup more in the Ostensive condition compared to the Non-ostensive condition ($X^2(1) = 5.11$, $p = 0.02$, $BF_{10} = 3.9$) (Figure S6A). 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 S2).

Main Experiment

Data Analysis

We analyzed data using R (Version 4.2.3; R Core Team, 2021) and the R packages *BayesFactor* (Version 0.9.12.4.4; Morey & Rouder, 2018), *bayestestR* (Version 0.13.0; Makowski et al., 2019), *brms* (Version 2.19.0; Bürkner, 2017a, 2017b), *car* (Version 3.1.1; Fox & Weisberg, 2019), *flextable* (Version 0.9.0; Gohel & Skintzos, 2023), *ggdist* (Version 3.2.1; Kay, 2023), *gghalves* (Version 0.1.4; Tiedemann, 2020), *ggpubr* (Version 0.6.0; Kassambara, 2023), *here* (Version 1.0.1; Müller, 2020), *kableExtra* (Version 1.3.4; Zhu, 2021), *knitr* (Version 1.42; Xie, 2015), *lme4* (Version 1.1.32; Bates et al., 2015), *papaja* (Version 0.1.1; Aust & Barth, 2020), *patchwork* (Version 1.1.2; Pedersen, 2020), *performance* (Version 0.10.2; Lüdecke et al., 2021), *psych* (Version 2.3.3; Revelle, 2021), *rmarkdown* (Version 2.20; Xie et al., 2018, 2020), *rstan* (Version 2.21.8; Stan Development Team, 2020), and *tidyverse* (Version 2.0.0; Wickham et al., 2019). Data, analysis scripts, and pre-registered methods (videos) are available at the Open Science Framework (https://osf.io/9r5xf/), as is pre-registration of our design and analysis plan (https://doi.org/10.17605/OSF.IO/GZ5PJ).

As an inference criterion, we used *p*-values below .05. Where possible, we supplemented the frequentist statistics with Bayes factors.

Performance Relative to Chance

We conducted 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 also conducted these analyses separately for each lab.

In addition to the frequentist analysis, we calculated 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 fitted a GLMM with binomial error distribution and logit link function using the glmer() function from the *lme4* package (Bates et al., 2015). This model included condition (Ostensive and Non-ostensive), 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 was 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 planned to repeat the above 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 were going to 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 were not able to conduct the pre-registered breed analysis due to too few breeds with at least eight individuals (see *Departures from the pre-registration*).

Across both models, we only included random slopes if the corresponding predictor variable varied in at least 50% of the levels of the random intercept. We included the random slope of the interaction if there was sufficient variation in both of its terms in at least 50% of the levels of the random intercept. We only included 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 were centered and scaled to a standard deviation of 1. The random slope components of the factors were centered to ensure that the results were not conditional on the choice of the reference category.

For the GLMM, we calculated 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 calculated Bayes factors for the models from Bayesian models using the brm() function from the *brms* package (Bürkner, 2017a, 2017b) with weakly informative priors (Student t-distribution with mean 0, degrees of freedom 6, and scale 1.5: student_t(6, 0, 1.5)). We used 4 chains with 15,000 iterations per chain (including 5,000 warm-up iterations). We then used the hypothesis() function to test hypotheses that the estimates for each predictor were 0. The Bayes factors represented the evidence for the full model relative to the full model without the fixed effect under investigation. The Bayesian analysis was supplemental, and inferences were drawn from the frequentist statistics.

Genetic Analysis of Among-Breed Heritability

To assess among-breed heritability (MacLean et al., 2019), we used an animal model (Wilson et al., 2010) that incorporates a genetic effect with a known covariance structure to estimate the proportion of phenotypic variance attributable to additive genetic effects. Genetic analyses took 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 was 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 (IBS) between pairs of individual dogs was calculated using PLINK (Chang et al., 2015). These values were then averaged for every pair of breeds to generate a breed-average IBS matrix. Because we did not differentiate between poodles of different sizes (e.g., standard, miniature, toy) when recording breed information for our participants, we averaged genomic data across poodles of all sizes when calculating our breed-average IBS matrix. The breed-average IBS matrix was then extrapolated to an individual-level IBS matrix. For individuals of different breeds, the IBS value was set to the average similarity between those breeds in the genetic dataset. For individuals of the same breed, the IBS value was set to the average IBS value among members of that breed in the genetic dataset. The purpose of this approach was to 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 $N \geq 3$ individuals were included in these analyses.

Heritability models were fit using the brm() function from the *brms* package (Bürkner, 2017a, 2017b) with weakly informative priors for the beta coefficients (normal(0,1)). We used 24,000 iterations per chain, with the first 2,000 iterations used as a warm-up, and a subsequent thinning interval of 10 iterations for retention of samples for the posterior distributions.

Heritability models included sex and age as covariates. Age was z scored prior to analysis. We fit three separate models using the following dependent measures: (1) proportion of correct choices in the Ostensive condition, (2) proportion of correct choices in the Non-ostensive condition, and (3) a difference score between these conditions, in which performance in the Non-ostensive condition was subtracted from performance in the Ostensive condition.

Model performance was assessed by visualizing posterior predictive checks and quantile-quantile plots, as well as review of summary statistics to ensure that chains converged (all Rhat values < 1.01).

Departures from the Pre-Registration

We set out to implement our procedure and analysis plan exactly as described above; however, we found at the end of the data collection period that the methods and data analysis plan involved, or required, deviations from our registration. These adjustments are:

1) In January, 2023, near the end of the data collection period, we discovered that for 12/20 sites, the handler had kept their eyes open during the baiting and cuing process of the experimental procedure. Of these 12, five used a trained handler, three used guardian handlers, and four used a combination of guardians and trained handlers. Eight of the sites using a trained handler or combination reported that their handler had seen where the treat was hidden on at least one trial. One site that used guardian handlers reported that they had not instructed the guardians to close their eyes or look down at their shoes during the cuing and baiting process.

2) Our pre-registered analysis of breed as a random effect required at least 8 individuals per breed. However, only six breeds had enough individuals to meet this criterion. Therefore, we conducted a comparable analysis in which we grouped breeds into 10 groups based on the Fédération Cynologique Internationale (FCI) breed categories. We included in our analysis purebred dogs from breed groups with at least 8 individuals in the sample (*N* = 243 of 7 FCI groups: Companion and Toy Dogs; Pinschers and Schnauzers; Pointing Dogs; Retrievers, Flushing Dogs, and Water Dogs; Sheepdogs and Cattledogs; Spitz and primitive types; Terriers).

3) Our original aim to explore the heritability of traits depended on reaching a threshold of eight individuals per breed. Because only six breeds met this criterion, we were not able to perform the originally planned pre-registered genetic analysis of among-breed heritability with our dataset. Instead, all breeds were included as long as at least three individuals had completed the study. Please see the exploratory analysis section of the paper for our post-hoc adjustment to this set of analyses.

Main Experiment Results

Across 20 sites, we tested 704 dogs and received demographic information for 701 of them (M:F = 331:370, mean ± SD age = 4.4 ± 3.1 years [range = 0.3-20.8]). Approximately 76.9% of the dogs were spayed or neutered, 53.8% were purebred, and 90.2% lived in private homes, 9.6% lived in group/kennel housing, and 0.3% lived in other housing (Table S1). However, 249 dogs of the 704 tested (35.37%) were excluded from the analysis because they failed to meet the inclusion criteria (235 failed to complete all trials and 14 experienced experimental errors during their sessions). This left 455 dogs for our analysis (M:F = 211:244, mean ± SD age = 4.5±3.1 years [range = 0.3-20.8]).

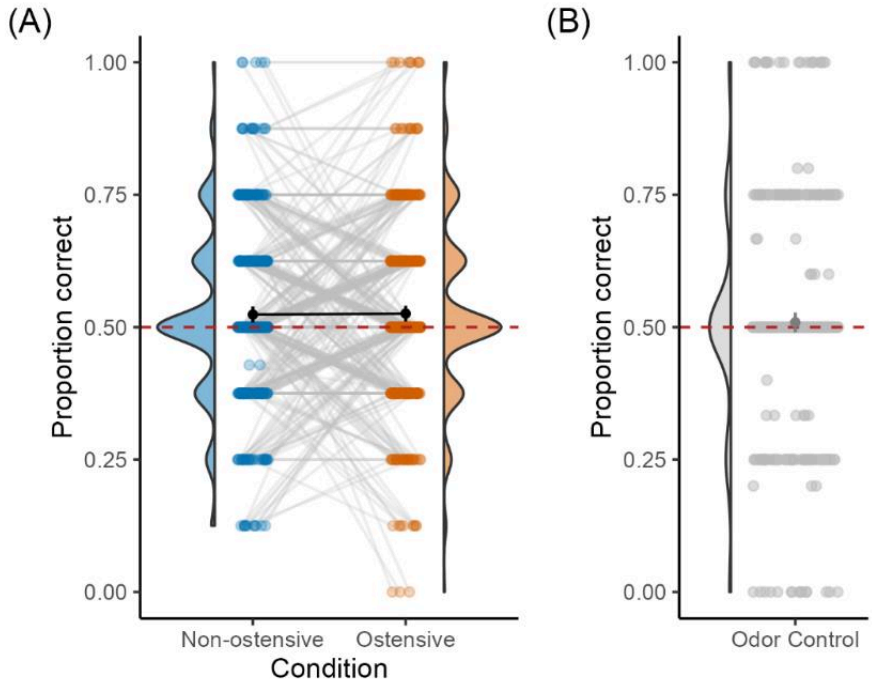
Confirmatory Analyses

Performance Relative to Chance

The dogs (N = 455) performed better than expected by chance in the Ostensive condition (Mean = 0.53, 95% CI [0.51, 0.54], $t(454) = 3.47$, $p < .001$, $BF_{10} = 19.4$) and in the Non-ostensive condition (Mean = 0.52, 95% CI [0.51, 0.54], $t(454) = 2.95$, $p = .003$, $BF_{10} = 3.8$) but not in the Odor Control condition (Mean = 0.51, 95% CI [0.49, 0.53], $t(413) = 0.92$, $p = .357$, $BF_{10} = 0.08$) (Figure 4). Mean performance in all conditions at individual sites typically did not differ from chance with a few exceptions: three sites had Ostensive performance greater than chance, three sites had Non-ostensive performance greater than chance (Table S3).

Figure 4

Violin and Dot Plot of Dogs' Performance (N = 455) across the (A) Non-Ostensive and Ostensive Conditions and the (B) Odor Control Condition



Note. The red dashed lines show the chance level of 0.5. Dots represent the mean proportion correct for each individual. The gray 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 means per condition.

Condition Comparison

The dogs did not choose the baited cup at different rates in the Ostensive condition compared to the Non-ostensive condition ($X^2(1) = 0.15$, $p = .70$) (Figure 4A). This pattern was consistent across almost all sites (Figure S7). 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 1).

Exploratory Analyses

Handler Bias

One of our departures from the pre-registered protocol involved the unintentional introduction of handler visual bias during the experiment. Eight of our twenty sites reported that some of the handlers (both researchers and guardians, depending on the method used at respective sites) had been able to view the cuing and baiting process of at least one test trial. The confirmatory analyses presented previously included all 20 sites, but here we conducted an exploratory analysis testing whether the potential of handler viewing influenced dog responses. To test this, we dummy coded all sites as either having the potential ("yes") or no potential ("no") for observing the baiting and cuing. Then, we added this variable as a fixed effect to the GLMM investigating condition effects on responses. Dogs did not choose the baited cup differently at sites with the potential for handler visual bias compared to sites without the potential for handler visual bias ($X^2(1) = 0.01$, $p = .92$).

Table 1

Results of GLMM of the dogs' choice performance

effect	Estimate	SE	Lower CI	Upper CI	Chi-square	df	*p*	BF*
(Intercept)	0.13	0.09	-0.04	0.32				
Condition	0.02	0.05	-0.08	0.11	0.15	1	.70	0.03
Condition order	-0.04	0.05	-0.13	0.05	0.56	1	.45	0.04
Trial number	-0.03	0.02	-0.07	0.02	1.35	1	.25	0.03
Age	0.01	0.03	-0.04	0.07	0.26	1	.61	0.02
Trainability score	-0.06	0.03	-0.12	0.01	2.61	1	.11	0.09
Sex:desexed	-0.01	0.12	-0.24	0.24	0.00	1	.95	0.08

Note. *Bayes factors for hypothesis that the predictor estimate is not 0. Thus, Bayes factors < 0.1 represent strong evidence that predictor estimates = 0.

Breed Group Effects

We were not able to conduct the pre-registered breed analysis due to too few breeds with at least 8 individuals. Therefore, we conducted a comparable analysis in which we grouped breeds into 10 groups based on the Fédération Cynologique Internationale (FCI) breed categories. We included in our analysis purebred dogs from breed groups with at least 8 individuals in the sample (N = 243 of 7 FCI groups: Companion and Toy Dogs; Pinscher and Schnauzer; Pointing Dogs; Retrievers, Flushing Dog and Water Dogs; Sheepdogs and Cattledogs; Spitz and primitive types; Terriers). For this subset of data, we fitted a binomial GLMM identical to our main model but including breed group as a random intercept (along with subject and site ID) and all possible random slope components. Condition had no effect on the dogs' choice performance ($X^2(1) = 0.52$, $p = .47$, $BF = 0.07$) (Figure 5). None of the control predictor variables (order of condition, trial number within condition, sex, neuter status, age, C-BARQ trainability score) had an effect on the dogs' choice performance (Table S4). The only trend was that dogs that started with the Ostensive condition tended to choose the baited cup less often than dogs that started with the Non-ostensive condition.

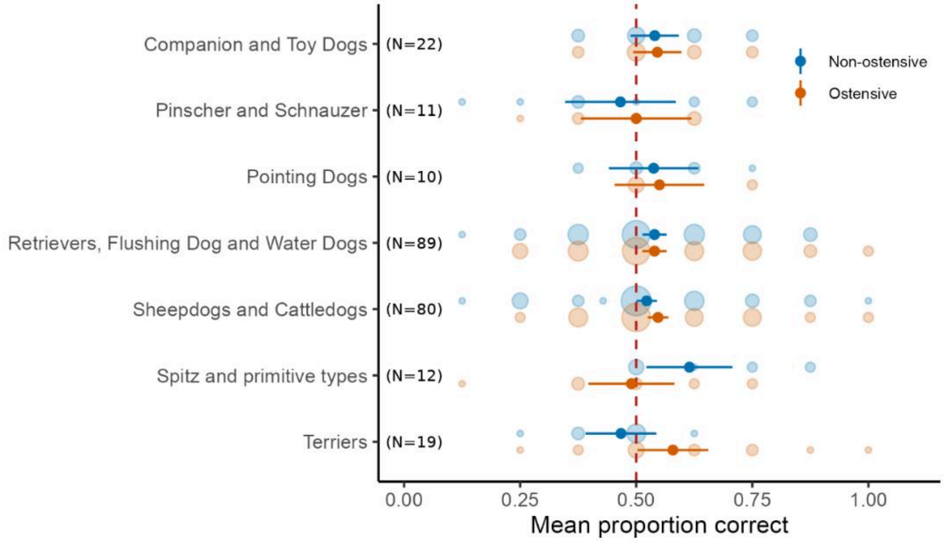
Among-Breed Heritability Effects

Because our final sample included only 6 breeds with 8 or more individuals, we could not meaningfully implement the pre-registered heritability analyses. Therefore, we conducted exploratory heritability models using a relaxed threshold for breed inclusion. We implemented these models for all purebred dogs with three or more individuals per breed (27 breeds; 208 individuals) that were also represented in the genetic data. Because we did not differentiate between poodles of different sizes (e.g., standard, miniature, toy) when recording breed information for our participants, we averaged genomic data across poodles of all sizes when calculating our breed average identity-by-state matrix. Additionally, because this resulted in a breed category characterized by substantial variation in body mass, we eliminated body mass as a covariate in the heritability models, retaining only covariates for dog sex and age.

We present the posterior distributions of heritability estimates in Figure 6. Posterior distributions tended to be asymmetrical with long tails and thus we summarize these results with the posterior mode and 90% highest density continuous intervals. In all cases the posterior mode was near 0 (Non-ostensive: 0.03, 90% highest-density continuous interval [0, 0.7]; Ostensive: 0.02, 90% highest-density continuous interval [0, 0.49]), indicating minimal genetic influence on the cognitive measures in this sample. The generally diffuse posterior distributions suggest that we cannot make confident inferences about genetic contributions to variance in the current sample.

Figure 5

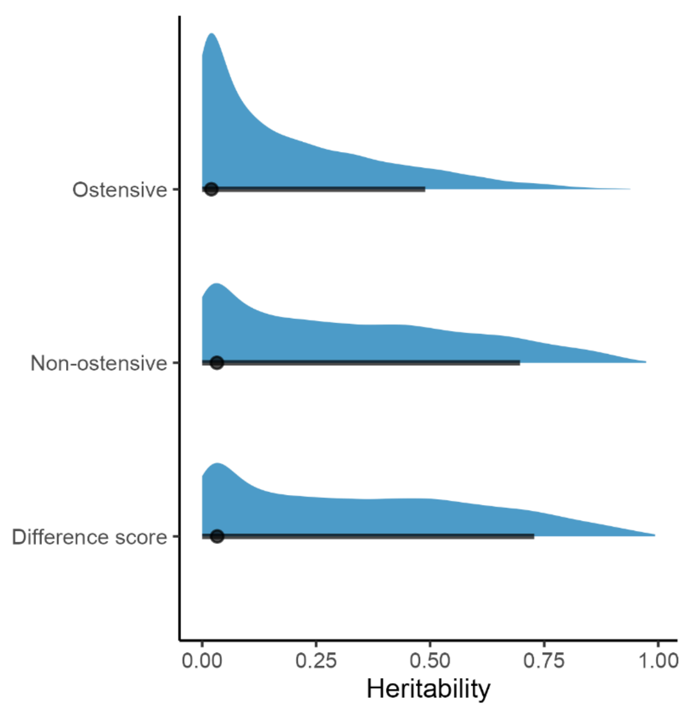
Plot of Dogs' Performance in Non-Ostensive and Ostensive Conditions for each Breed Group with N >= 8.



Note. Orange (Non-ostensive condition) and blue (Ostensive condition) bubbles represent the number of individuals at that performance level. Filled dots represent 95% within-subjects confidence intervals. The red dashed line shows the chance level of 0.5.

Figure 6

Posterior Distributions of Heritability Estimates for Models Including Dogs from Breeds with Three or More Individuals with Cognitive Data



Note. Points reflect the posterior mode and lines reflect the 90% highest continuous posterior interval for each model.

Within-Subject Reliability

To examine the extent to which individual performance was stable across trials and conditions, we performed a split-half reliability analysis. We first split the data into odd and even trials (irrespective of condition) and aggregated the odd and even trial performance (mean individual performance). However, we found no evidence for a correlation between their performance in odd and even trials (Pearson correlation: $r(452) = 0.07$, 95% CI [-0.03, 0.16], $p = .164$, $BF_{10} = 0.29$; Figure S8A). Additionally, we aggregated the Ostensive and Non-ostensive condition performance of each subject. While the correlation between the two conditions was small in magnitude, it was statistically significant, indicating a positive relationship between individuals' performance in the two conditions ($r(452) = 0.26$, 95% CI [0.18, 0.35], $p < .001$, $BF_{10} = 1.1\times10^6$; Figure S8B).

Response Strategies

Overall, it is not clear that subjects followed pointing cues often in this task. We were interested in exploring other strategies that the dogs could have employed. Two candidate strategies investigated in a previous pointing study are win-stay, lose-shift and win-shift, lose-stay (Byrne et al., 2020). That is, rather than following cues, the subjects could simply continue choosing the same cup or switch to the other cup depending on whether they received a reward on the previous trial. To test whether dogs were using these strategies, we calculated for each trial (except the first in a block) whether the dogs' performance followed a win-stay, lose-shift or a win-shift, lose-stay strategy based on their performance in the previous trial. We found that the win-stay, lose-shift strategy would have been negatively correlated with success (Pearson correlation: $r(452) = -0.67$, 95% CI [-0.71, -0.61], $p < .001$, $BF_{10} = 5.9\times10^{55}$; Figure S9A), and conversely a win-shift, lose-stay strategy would have been positively correlated with success ($r(452) = 0.67$, 95% CI [0.61, 0.71], $p < .001$, $BF_{10} = 5.9\times10^{55}$; Figure S9B). These correlations are likely caused by the pseudo-randomization of the baited side (the food was presented no more than two trials in a row on the same side). At a group level, the dogs did not engage in the win-stay, lose-shift (Mean = 0.50, 95% CI [0.49, 0.51], $t(453) = 0.06$, $p = .955$, $BF_{10} = 0.05$; Figure S9C) or the win-shift, lose-stay strategy (Mean = 0.50, 95% CI [0.49, 0.51], $t(453) = -0.06$, $p = .955$, $BF_{10} = 0.05$; Figure S9D) above chance levels (0.5).

An even simpler strategy would be to always choose the same side. Side biases are relatively common in animal choice experiments (Andrade et al., 2001; Miletto Petrazzini, Pecunioso, et al., 2020), including dog studies (Gácsi et al., 2009; Miletto Petrazzini, Mantese, et al., 2020). In our study, this would be a reasonable strategy because it would result in a reward on average every other trial. Overall, in 49.9% of the trials, the food was located on the right side, and dogs chose the right side in 51.1% of trials. Side biases were relatively common with 78.0% of dogs biased more than 10% away from the experienced chance levels (Figure S10). This bias varied substantially across sites (Figure S11).

No-Choice

For dogs included in this analysis, dogs did not choose a cup (no-choice) in 2.0±4.3% (mean ± SD) of the trials (per dog). This differed between conditions with more no-choices in the Non-ostensive (2.4%, 95% CI [1.8, 2.9]) condition compared to the Ostensive (1.4%, 95% CI [1.0, 1.8]) condition. It did not matter if dogs experienced the Non-ostensive condition first (1.8%, 95% CI [1.3, 2.3]) or the Ostensive condition first (2.0%, 95% CI [1.5, 2.4]). Condition order also did not influence whether dogs were included in the final analysis: 67.7% of dogs that received Non-ostensive first were included compared to 66.4% of dogs that received Ostensive first.

Discussion

In ManyDogs 1, we aimed to assess dogs' ability to follow human pointing, a well-known phenomenon in canine science. Specifically, we posed a question that is theoretically interesting to many researchers in the field: do dogs understand and act on human pointing as though it is a social communicative gesture? In this inaugural study, with a sample of 455 dogs assessed across 20 research groups and, using the same method as our pre-registered pilot study (N=61), we found that dogs do not robustly follow a human pointing gesture if it is performed in a momentary contralateral manner without gaze cues. Although dogs performed better than expected by chance in both the Ostensive and the Non-ostensive conditions (but not in the Odor Control condition), their success was overall very low. In both test conditions, the 95% confidence intervals for the proportion of correct responses were between 0.51 and 0.54. Moreover, dogs did not perform differently between the two conditions (Ostensive M = 0.53; Non-ostensive M = 0.52), and no other predictors (order of condition, trial number within condition, sex, neuter status, age, C-BARQ trainability score) significantly affected performance. This finding was consistent across sites with only one out of 20 sites showing a difference between conditions. Follow-up analyses also found no differences in performance across breed groups and no meaningful genetic influence on performance. Given the low performance levels, we investigated within-subject reliability and found no relationship between performance in even versus odd trials but a positive relationship in performance across the two conditions. Finally, we found no evidence of win-stay, lose-shift or win-shift, lose-stay strategies. Instead, it appears that many dogs simply chose one particular side for most trials, despite their willingness to approach both sides in warm-up phases. These side biases likely resulted from the dogs' minimal attention to the pointing cues. Below, we briefly overview our three hypotheses and further discuss how we interpret our results given both our pilot results and the previous literature.

Our main hypothesis set forth that the use of ostensive cues preceding a human-given pointing cue would facilitate significantly above chance rates of point-following in dogs, while points lacking ostension would not. We did not observe this pattern in our study; instead, response rates were similar and only slightly above chance in both conditions. The low performance observed in the ostensive condition is surprising given the higher performance observed in our pilot study (Ostensive M = 0.60, 95% CI [0.55, 0.65]) and previous work using a similar cue (e.g., Kaminski et al., 2012; Pongrácz et al., 2013). The low reliability in performance across trials further suggests that the point-following signal in our data was weak. Yet, we do not interpret our findings as suggesting that dogs are unable to follow pointing gestures. Though the dominant view in the literature has been that dogs accurately, spontaneously, and flexibly follow ostensive pointing gestures, our multi-lab data highlight limitations in their point-following behavior. Thus, we propose that not all pointing gestures are equally likely to elicit point-following.

In the present investigation, we used a single, specific pointing cue: contralateral, momentary (2 s), and without accompanying gaze cues to the target. This cue type was chosen through a consensus-building process among participating researchers; the resulting pointing style appears to have been particularly difficult for dogs, perhaps due to limited salience. There may also be other elements of the pointing style, such as speed, angle, or emphaticness that could be important and merit further investigation. We therefore suspect that the low point-following that we report is not necessarily generalizable to other pointing styles, while simultaneously, our finding raises questions about the reliability of point-following behavior that should be investigated further. This lack of point-following also limits the inferences we can draw from our primary comparison of ostensive and non-ostensive cues; the lack of difference between the conditions is not particularly meaningful in the absence of robust point-following.

There is some empirical evidence to support the claim that additional human-given signals may result in increased salience of pointing, and therefore higher rates of point-following in dogs. Kaminski et al. (2012) observed that dogs (N = 26) selected the correct cup at above-chance levels (ca. 72% correct choices on average) when gaze alternation between the dog and the momentary, cross-lateral cue was presented. Utilizing a similar method of pointing, Pettersson et al. (2011) found that pointing helpfully with a cooperative tone of voice led to dogs' above chance performance (N = 40, correct M: 59%), even in the first trial (15/20 dogs chose the baited and cued cup). By including additional social cues with a pointing gesture, the signal level of the compound cue may reach a response threshold above which it is salient enough for dogs to reliably follow.

However, three other studies have found convincing point-following in dogs without the use of additional human-given signals—specifically, gaze alternation. Lakatos et al. (2009) and Pongrácz et al.

(2013) both conducted studies in which a similar pointing gesture to the current study was used. In these studies, 15 and 46 dogs, respectively, were tested with a "long cross-pointing" gesture in a momentary manner, and without the experimenter following the hand gesture with her gaze. Interestingly, we consider the pointing gesture in these studies to have been particularly challenging in three respects: 1) the distance between the experimenter's tip of the finger and the cup was four times larger (120 cm) than in our study (30 cm), 2) the distance between the dog and the cup line was also greater (ca. 200 cm instead of 135 cm in our study), and 3) the signaling duration was only 1 s instead of 2 s. Nevertheless, the dogs in these two studies outperformed the current sample of dogs. The dogs chose the pointed cup significantly above the chance, with on average more than 70% (Lakatos et al., 2009) and 69% (Pongrácz et al., 2013). In the third study, the experimenter used a momentary "cross-forward pointing" in which she pointed with her contra-lateral white hand in the direction of the correct location, but her extended index finger did not protrude from her black silhouette, and still the dogs (N = 14) performed clearly above chance at 60% (Lakatos et al., 2007).

Dogs in the current study performed slightly above chance in the Ostensive condition, following pointing at much lower levels than in previous research. Contrary to our original prediction, there was no difference between pointing conditions, with dogs also performing marginally above chance in the Non-ostensive condition. The low point-following we observed does not align with previous investigations. In the study by Kaminski et al. (2012), domestic dogs treated the same human gesture differently if it was performed with a communicative intention or not. More recently, Byosiere et al. (2022) found that dogs of varying ages were more responsive to human ostensive signals than similar but non-ostensive ones. However, again, there are methodological differences between the studies that must be emphasized. In both studies, the experimenter did not only alternate gaze between the dog and the cup, but also established eye contact and called the dog's name at the onset of the trial. We, therefore, do not know if the gaze alternation between dog and cup is necessary to signal the dog to attend to the cup (referential intention) or if the initial ostensive cues (eye contact, verbal addressing) are sufficient to indicate that the communicative act is directed at them.

Our second hypothesis addressed whether dogs interpret pointing gestures as either an imperative or informative signal. As acknowledged in the introduction, our protocol and analytical plan were not designed to definitively discriminate between these two alternatives. However, we indicated that if dogs interpret pointing gestures as imperative signals, trainability could predict performance in both the ostensive and non-ostensive conditions. We found that trainability did not predict performance; given the low performance observed, it remains unclear whether dogs interpret pointing cues as imperative or informative signals. We could speculate that the above-chance responses in the Non-ostensive condition are suggestive of obligatory following, as put forward by Topál et al. (2014), or that dogs are relying on a previously learned gesture-reward association (e.g., Wynne et al., 2008). Our results do not lend compelling support to either of these explanations. If dogs felt obliged to respond or had learned that pointing is accompanied by treats, we would expect to see a stronger behavioral response.

The third, and final, hypothesis was included as a control to make sure that dogs were using the demonstrated visual cues when choosing between cups. We predicted that, as a group, dogs would not rely on odor cues to make a choice between cups and therefore would not select the baited cup in the odor control significantly above chance. This is what we observed. Across research sites dogs performed at chance in the odor control condition, confirming previous studies: dogs do not appear to spontaneously use olfactory cues to find hidden food in object-choice tasks, particularly when the task is presented in a visual modality (Szetei et al., 2003).

To our knowledge, this is the first dog pointing study to be provisionally accepted as a registered report in which the methods and analytic plan were formalized before data were collected. Registered reports typically yield positive results less frequently than other hypothesis-testing studies in the literature. For example, a study by Scheel et al. (2021) found that only 44% of registered reports resulted in positive findings, compared to 96% in standard reports. Similarly, high-powered replication studies in the social sciences have revealed the replicability to be limited by false positives and exaggerated effect sizes in the literature (Camerer et al., 2018). A plausible explanation for these findings is a publication bias in the field for positive outcomes. The findings of the present study seem to support these concerns and suggest that dogs' point-following is more limited than previously thought.

Limitations & Future Directions

We selected a multi-lab, single-protocol approach for ManyDogs' initial study. Similar to ManyBabies (e.g., The ManyBabies Consortium, 2020), the cost of behavioral data collection with dogs is high and this avenue of sharing investigation efforts also offered the opportunity to examine sources of variability between research sites. As we and other initiatives have discussed (Byers-Heinlein et al., 2020; Forscher et al., 2022; ManyDogs Project et al., 2022), this distributed, collaborative method to scientific investigation overcomes problems inherent to single-lab investigations but comes with its own challenges for conducting research activities.

One of these challenges is inter-lab differences, both a strength and a limitation of this study design. On the one hand, the single-protocol, multi-lab approach allowed us to explore specific inter-lab variation in effects, an aim that is in line with the core motivations of large-scale replication projects (e.g., Klein et al., 2018). On the other hand, despite using a single protocol across different research environments, uncontrolled differences between sites are necessarily introduced. These might include factors inherent to the local site's participant recruitment procedure, testing environment, or personnel training on research protocols. Our steps to standardize the protocol included a written protocol and video demonstrations. We also reviewed and provided detailed feedback on two rounds of sample videos from each site to assist in protocol learning and implementation. While this was a helpful–and necessary–process by which research teams could receive feedback and move towards standardization, it was not sufficient oversight to prevent variations from being introduced to the protocol. For example, we discovered a major deviation from the pre-registered method in a majority of labs (i.e., handlers did not close their eyes during test trial baiting). Since a large number of sites reported this alteration in how they performed the protocol, it may reflect on the information organization of the project and suggests that instructions could have been clarified or simplified for better compliance. At the same time, it could be a broader-reaching indication of how difficult it is to standardize a protocol across different individuals, let alone across separate sites. In a standard, single-lab study, subtle yet important changes to the protocol execution may not be noticed or reported. If a single experimenter is performing the demonstration, there is no available comparison that would flag any gradual evolution in how a protocol is being performed from one subject to another. In the process of scaling up and comparing notes across sites, we were given a unique opportunity to shed light on how behavioral experiments are executed in canine science (potentially across other areas of behavioral science as well), which makes us aware of additional sources of uncontrolled variability that are present in behavioral studies. Future work at the individual lab level as well as other big team science studies should consider how to minimize this tendency and take steps to do additional experimental assessments aimed at maintaining the protocol's integrity.

Though this study represents an important investigation and replication of point-following behavior in dogs, we are not able to extrapolate strong conclusions about this behavior across other contexts, given that we were constrained by the type of pointing gesture that we examined. We used one pair of contrasting pointing cues across sites which allowed us to assess replicability of point-following in different cultural contexts, but this necessarily limited the breadth of our exploration into how dogs perceive and respond to human pointing. Pointing can take many forms (e.g., Soproni et al., 2002; Lakatos et al., 2009), and be presented in conjunction with a variety of additional cues, including (but not limited to) vocalizations, facial expressions, and gazing. We observed statistically significant but only slightly-above-chance responses across all research sites. This is not to say that the findings of this study do not provide important insights about dog-human interaction. Quite the opposite, our results suggest that there are limits to dogs' point-following and that simply pointing at an object does not necessarily elicit an accurate behavioral response. Future studies of point-following in dogs should build on these results and examine more closely the role that accessory cues have in motivating a following response, in particular gazing at or alternating gaze between the subject and target, and build on our understanding of the conditions in which dogs respond reliably to pointing.

Another challenge that we faced during this study was the COVID-19 pandemic. Lockdowns interrupted our piloting in 2020 and continued to negatively impact data collection and slow progress throughout. The combination of restrictions on in-person research activities and the uncertainty of not knowing when the next wave of the virus could cause another lockdown made it extremely challenging for sites to plan data collection and resulted in long delays and disruptions. Following health guidelines, some experimenters wore face masks, which may have had an impact on dogs' behavior in a social interaction task, potentially disrupting attention to the experimenter's face and therefore facial processing. More generally, pandemic-related social isolation has affected both dogs and owners and recent findings indicate there has been a widespread increase in problem behaviors, such as bite incidence and severity (Habarth-Morales et al., 2022; Pitak-Arnnop et al., 2022) and worsening separation-related problems or anxiety (Ribeiro et al., 2023; Sherwell et al., 2023). The extent of these effects across populations are difficult to estimate, but this early work already indicates far-reaching impacts on dog-human social interaction.

Anecdotally, many of the research groups in our study reported that more dogs than in their past (pre-pandemic) studies were anxious upon arriving at the research site and during the early stages of the experiment, in some cases being unwilling to eat food in the experiment room. Unless dogs acclimated well, these dogs were often sent home without completing the study and were excluded from the reported data analyses. Indeed, out of 704 dogs that showed up across research sites to participate in the study, 235 dogs did not complete the protocol for various reasons (e.g., anxiety, lack of motivation). This is a rather remarkable dropout rate of over 33% of the total sample that attempted the task. Compared to studies that use a similar cue (0% excluded in Lakatos et al., 2009; 3% excluded in Kaminski et al., 2011; 18% and 3% in Experiments 2 & 3 excluded in Pongrácz et al., 2013) and even in non-pointing canine science studies, this rate is unusually high (one exception reported a dropout rate of 56.4%; Stevens et al., 2022). It is important to note that this rate is a proportion of the total tested (i.e., a dog is included in the sample after having attempted at least one warm-up trial), which is not how other studies typically report this information. We believe that using this approach provides greater transparency about the costs of data collection in canine science, while underscoring the practicality of big team science for gathering large samples through distributing the data collection load across a network. The high rate of noncompletion we observed took a toll on our final sample breed composition as well. This was a limiting factor on our ability to do a broader, robust breed comparison and likely contributed to the wide distribution of heritability estimates. Nevertheless, the inclusion of a large number of mixed breed dogs and relative consistency across the breed groups sampled suggests that our results were not merely due to sampling breeds that may be less receptive to human signals (Wobber et al., 2009; Gnanadesikan et al., 2020; Junttila et al., 2022).

The data we have collected over the course of ManyDogs 1 is rich in additional information, both about research subjects and the sites at which they participated in the study. Beyond the planned analyses of our registered report, we were able to include key additional (exploratory) analyses to account for our final sample and its limitations, as well as to explore the unexpected findings of our study in more detail. Throughout the study we have collected large amounts of additional demographic information (including C-BARQ scores) about individual dogs, and this along with research site related data are open for use in secondary analyses. We are making this data available with the publication with the hope that future investigations of dog-human social interaction will be able to incorporate it and examine additional hypotheses about socio-cultural factors that play a role in dogs' social behavior in a research study context.

At the larger consortium level, like other ManyX initiatives, we faced challenges of accessibility and functionality when integrating and scaling up software platforms (e.g., GitHub, Qualtrics, and Google Drive) to manage information sharing and data collection across many research groups (e.g., see Byers-Heinlein et al., 2020). The absence of dedicated administrative support and personnel was a main factor in deciding to automate sections of this process, a necessary tradeoff that imposed a higher learning curve on researchers who were not accustomed to integrating these tools into their typical workflows. Future ManyDogs studies can improve on this limitation by streamlining information resources, obtaining funding to support administrative oversight, and implementing a comprehensive "collaborator training" into the onboarding process that would support skill acquisition for individual collaborators joining the study. Looking at accessibility in a different domain, along with related themes such as inclusion and diversity, the first ManyDogs study is not representative of the global population of domestic dogs, nor the researchers that study them. We were successful in uniting teams across nine countries, which is a strong beginning for the consortium. However, there are many other research groups in other parts of the world, and a priority for future projects is to grow the collaborative network to connect with research teams outside of the Americas and Europe. This would facilitate the study of groups of dogs with different roles in society besides companionship (e.g., working dogs or free-ranging dogs).

Conclusion

As an international consortium, the ManyDogs project represents a group of researchers interested in conducting big team canine science with the aims to 1) enhance replicability, 2) provide a platform for testing questions that require large and/or diverse samples, 3) quantify differences across labs, and 4) foster collaboration. We realized these aims in our first empirical study, ManyDogs 1, uniting a geographically distant network of canine science researchers in a common project, then collectively investigating a question of theoretical importance in our field. Twenty research groups across nine countries came together to implement a single-protocol study, each working with different populations of dogs to evaluate dogs' ability to follow human-given ostensive and non-ostensive pointing. The empirical results of this study do not provide evidence that ostensive cues (subject-focused gaze and dog-directed speech) positively impact rates of point-following. Rather, we saw no difference between Ostensive and Non-ostensive conditions and overall point-following was quite low. Although statistically significant, dogs' point-following was not substantially different from what would be expected by chance. These findings suggest that there are limits to point-following behavior in dogs. The outcomes of ManyDogs 1 demonstrate that big team science is of critical importance within the field of canine science and strongly supported by researchers globally. Through this consortium, we believe we can continue working to address questions within the field of canine science that have previously been unanswerable or inconclusive, without the mobilization of a large number of researchers. There is a broad-scale and collaborative interest in expanding and maintaining an interdisciplinary network in which diverse individuals studying various canine models can contribute to the growth of scientific knowledge within canine science. Thus, we believe that the future of the ManyDogs project is bright.

Acknowledgements

Many individuals contributed to this project that are not part of the author list. First, we are very grateful to the guardians who volunteered to participate in this study with their dogs. Next, we thank the additional researchers who provided logistic support with site enrollment, subject recruitment, protocol piloting, and completed the reliability coding at research sites (ordered alphabetically by site): Abigail Flyer, Chandler From, and Tyler Rodriquez (ACCC); Lyndsey Elrod (ACPS); Kerry Hunt and Daisy Jolly (AHWRC); Rachel Ceccanecchio and Noah Liedtke-Silva (BCCC); Lauren Bonner, Anwyn Gatesy-Davis, Lindsay Gould, and Samantha Kritzer (BDL); Dr. Mariana Bentosela (CBRG); Carmelo Mattia La Ferrara, Agnese Marrazzo, and Andrea Reitano (CC); Miranda Dresen, Bailey Wilson, and London Wolff (CCHIL); Theadora Block, Brenda Kennedy, and Emlyn Vezzosi (CCI); Yanyu Lei and Em Sundby (CCL); Gorana Birovljević and Tea Štrbac (CCSL); Emily Halfyard and Jenna K. Perry (CRU); Samantha Kefer and Emily Sandberg (DCCC); David Locklin (LDB); Hannah Burrows and the staff of Grassroots K9 (SCL); Angie Lee and Gillian Schmitt (TDC).

Author Contributions: Several of the authors have maintained a high level of dedication and effort over the years of this project. J. Espinosa served as the project lead, J.R. Stevens supported leadership and developed data infrastructure, and S.-E. Byosiere supported leadership and collaborator training. Other researchers that were instrumental in the success of the project are G.E. Gnanadesikan and M.H. Pelgrim

for overseeing protocol development and training, D. Buchsbaum for establishing the video database, M. Bogese for organizing collaborator registration, and E. Hare for assisting in data management.

CRediT Contributions

Conceptualization: J. Espinosa, E.E. Bray, D. Buchsbaum, S.-E. Byosiere, M. Byrne, G.E. Gnanadesikan, D.J. Horschler, A.M. Johnston, E.L. MacLean, M.H. Pelgrim, and Z.A. Silver. **Data curation**: J.Espinosa, J.R. Stevens, E. Hare, and C.J. Völter. **Formal analysis**: J.R. Stevens, E. Hare, E.L. MacLean, and C.J. Völter. **Investigation**: D. Alberghina., H.E.E. Alway, J.D. Barela, E.E. Bray, S.-E. Byosiere, C.M. Cavalli, L.M. Chaudoir, C. Collins-Pisano, H.J. DeBoer, L.E.L.C. Douglas, S. Dror, M.V. Dzik, B. Ferguson, L. Fisher, H.C. Fitzpatrick, M.S. Freeman, S.N. Frinton, M.K. Glover, J.E.P. Goacher, M. Golańska, M. Hickey, H.-L. Jim, D.M. Kelly, V.A. Kuhlmeier, L. Lassiter, L. Lazarowski, J. Leighton-Birch, K. Maliszewska, V. Marra, L.I. Montgomery, M.S. Murray, E.K. Nelson, L. Ostojić, S.G. Palermo, A.E. Parks Russell, M.H. Pelgrim, S.D. Pellowe, A. Reinholz, L.A. Rial, E.M. Richards, M.A. Ross, L.G. Rothkoff, H.Salomons, J.K. Sanger, A.R. Schirle, S.J. Shearer, J.M. Silverman, A. Sommese, T. Srdoc, H. St. John-Mosse, K. Vékony, Y.A. Worth, L.M.I. Zipperling, B. Żołędziewska, and S.G. Zylberfuden. **Methodology**: J. Espinosa, E.E. Bray, D. Buchsbaum, M. Byrne, G.E. Gnanadesikan., L. Huber, M.H. Pelgrim, and C.J. Völter. **Project administration**: J. Espinosa, J.R. Stevens, M. Bogese, D. Buchsbaum, S.-E. Byosiere, G.E. Gnanadesikan, E. Hare, and M.H. Pelgrim. Resources: L. Santos. **Software**: J.R. Stevens and E. Hare. **Supervision**: D. Alberghina, J.D. Barela, E.E. Bray, D. Buchsbaum, S.-E. Byosiere, C.M. Cavalli, M.S. Freeman, M.K. Glover, G.E. Gnanadesikan, B. Hare, D.J. Horschler, L. Huber, A.M. Johnston, J. Kaminski, D.M. Kelly, V.A. Kuhlmeier, L. Lazarowski, L. Ostojić, A. Reinholz, A. Sommese, C.J. Völter, and C.J. Walsh. Validation: J.R. Stevens, E. Hare, and A.C. Vega. **Visualization**: J.R. Stevens, E.L. MacLean, and C.J. Völter. **Writing - original draft**: J. Espinosa, J.R. Stevens, S.-E. Byosiere, M.S. Freeman, G.E. Gnanadesikan, L. Huber, A.M. Johnston, E.L. MacLean, M.H. Pelgrim, Z.A. Silver, and C.J. Völter. **Writing - review & editing**: J. Espinosa, J.R. Stevens, D. Alberghina, E.E. Bray, D. Buchsbaum, S.-E. Byosiere, M.S. Freeman, G.E. Gnanadesikan, C.-N.A. Guran, D.J. Horschler, and L. Huber. **Registered report preparation**: J. Espinosa, J.R. Stevens, E.E. Bray, D. Buchsbaum, S.-E. Byosiere, M.S. Freeman, G.E. Gnanadesikan, C.-N.A. Guran, D.J. Horschler, L. Huber, A.M. Johnston, E.L. MacLean, M.H. Pelgrim, Z.A. Silver, and C.J. Völter. **Final manuscript preparation**: J. Espinosa, J.R. Stevens, S.-E. Byosiere, G.E. Gnanadesikan, L. Huber, A.M. Johnston, E.L. MacLean, and C.J. Völter.

Funding: Some of the collaborators were supported in their research activities by individually acquired awards. G.E. Gnanadesikan was supported by the National Science Foundation Graduate Research Fellowship Program (DGE-1746060). B. Hare and H. Salomons were supported by the Office of Naval Research (N00014-16-12682) and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NIH-1R01HD097732). D.J. Horschler was supported by the National Science Foundation SBE Postdoctoral Research Fellowship under Grant No. 2104589.; J.R. Stevens was supported by the National Science Foundation (NSF-1658837); V.A. Kuhlmeier was supported by the Natural Sciences and Engineering Council of Canada (NSERC) operating grant to V. Kuhlmeier; D.M. Kelly was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) - RGPIN-2017-04944; M.K. Glover was supported by Ontario Graduate Scholarship; C.M. Cavalli and L.A. Rial were supported by Préstamo BID, PICT 2018 N° 1581, Agencia Nacional de Promoción Científica y Tecnológica; M.S. Murray was supported by PSC-CUNY Award TRADB-53-86; S.-E. Byosiere was supported by a PSC-CUNY Award (Research Award -TRADB-53-86), jointly funded by The Professional Staff Congress and The City University of New York.; A. Reinholz was supported by the Faculty of Psychology, University of Warsaw, from the funds awarded by the Ministry of Science and Higher Education in the form of a subsidy for the maintenance and development of research potential in 2023 (501-D125-01-1250000 zlec. 5011000615).

Conflict of Interest: The authors declare no conflict of interest in designing or carrying out this study.

Data Availability: Our pre-registration of the design and analysis plan, data, analysis scripts, supplementary materials, and reproducible research materials are available at the Open Science Framework (https://osf.io/9r5xf/).

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. https://doi.org/10.1177/25152459211039218

Andrade, C., Alwarshetty, M., Sudha, S., & Suresh Chandra, J. (2001). Effect of innate direction bias on T-maze learning in rats: Implications for research. *Journal of Neuroscience Methods*, *110*(1), 31–35. https://doi.org/10.1016/S0165-0270(01)00415-0

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. https://doi.org/10.1007/s10071-020-01448-2

Aust, F., & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with R Markdown*. https://github.com/crsh/papaja

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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1111/j.1467-7687.2005.00440.x

Ben-Aderet, T., Gallego-Abenza, M., Reby, D., & Mathevon, N. (2017). Dog-directed speech: Why do we use it and do dogs pay attention to it? *Proceedings of the Royal Society B: Biological Sciences*, *284*(1846), 20162429. https://doi.org/10.1098/rspb.2016.2429

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

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*(14), 3132-3136.e5. https://doi.org/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. (2020). Cognitive characteristics of 8- to 10-week-old assistance dog puppies. *Animal Behaviour*, *166*, 193–206. https://doi.org/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. (2021). Dog cognitive development: A longitudinal study across the first 2 years of life. *Animal Cognition*, *24*(2), 311–328. https://doi.org/10.1007/s10071-020-01443-7

Bürkner, P.-C. (2017a). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2017b). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/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. https://doi.org/10.1037/cap0000216

Byosiere, S.-E., Mundry, R., Range, F., & Virányi, Z. (2022). Selective responding to human ostensive communication is an early developing capacity of domestic dogs. *Developmental Science*, *2022*, e13361. https://doi.org/10.1111/desc.13361

Byrne, M., Bray, E. E., MacLean, E. L., & Johnston, A. M. (2020). Evidence for Win-Stay-Lose-Shift in Puppies and Adult Dogs. *Proceedings of the 42nd Annual Conference of the Cognitive Science Society,* (pp. 2884-2889). https://www.cognitivesciencesociety.org/cogsci20/papers/0717/0717.pdf

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

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), s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8

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

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

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. https://doi.org/10.1007/s10071-017-1098-2

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. https://doi.org/10.1098/rsos.170349

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

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. https://doi.org/10.1371/journal.pone.0256607

Forscher, P. S., Wagenmakers, E.-J., Coles, N. A., Silan, M. A., Dutra, N., Basnight-Brown, D., & IJzerman, H. (2022). The Benefits, Barriers, and Risks of Big-Team Science. *Perspectives on Psychological Science*, 17456916221082970. https://doi.org/10.1177/17456916221082970

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

Gácsi, M., Kara, E., Belényi, B., Topál, J., & Miklósi, Á. (2009). The effect of development and individual differences in pointing comprehension of dogs. *Animal Cognition*, *12*(3), 471–479. https://doi.org/10.1007/s10071-008-0208-6

Gnanadesikan, G. E., Hare, B., Snyder-Mackler, N., & MacLean, E. L. (2020). Estimating the heritability of cognitive traits across dog breeds reveals highly heritable inhibitory control and communication factors. *Animal Cognition*, *23*(5), 953–964. https://doi.org/10.1007/s10071-020-01400-4

Gohel, D., & Skintzos, P. (2023). *flextable: Functions for tabular reporting*.

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. https://doi.org/10.1111/cdev.13026

Habarth-Morales, T. E., Rios-Diaz, A. J., & Caterson, E. J. (2022). Pandemic Puppies: Man's Best Friend or Public Health Problem? A Multidatabase Study. *Journal of Surgical Research*, *276*, 203–207. https://doi.org/10.1016/j.jss.2022.02.041

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

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. https://doi.org/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. https://doi.org/10.2460/javma.2003.223.1293

Jeannin, S., Gilbert, C., Amy, M., & Leboucher, G. (2017). Pet-directed speech draws adult dogs' attention more efficiently than Adult-directed speech. *Scientific Reports*, *7*(1), 4980. https://doi.org/10.1038/s41598-017-04671-z

Junttila, S., Valros, A., Mäki, K., Väätäjä, H., Reunanen, E., & Tiira, K. (2022). Breed differences in social cognition, inhibitory control, and spatial problem-solving ability in the domestic dog (*Canis familiaris*). *Scientific Reports*, *12*(1), 22529. https://doi.org/10.1038/s41598-022-26991-5

Kaminski, J., & Nitzschner, M. (2013). Do dogs get the point? A review of dog-human communication ability. *Learning and Motivation*, *44*(4), 294–302. https://doi.org/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. https://doi.org/10.1111/j.1467-7687.2011.01120.x

Kassambara, A. (2023). *ggpubr: "ggplot2" based publication ready plots*. https://rpkgs.datanovia.com/ggpubr/

Kay, M. (2023). *ggdist: Visualizations of distributions and uncertainty*. https://doi.org/10.5281/zenodo.3879620

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. https://doi.org/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. https://doi.org/10.1027/1864-9335/a000178

Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., … Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, *1*(4), 443–490. https://doi.org/10.1177/2515245918810225

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. https://doi.org/10.1007/s10071-010-0359-0

Lakatos, G., Dóka, A., & Miklósi, Á. (2007). The Role of Visual Cues in the Comprehension of the Human Pointing Signals in Dogs. *International Journal of Comparative Psychology*, *20*(4), 341-350. https://doi.org/10.46867/ijcp.2007.20.04.03

Lakatos, G., Soproni, K., Dóka, A., & Miklósi, Á. (2009). A comparative approach to dogs' (*Canis familiaris*) and human infants' comprehension of various forms of pointing gestures. *Animal Cognition*, *12*(4), 621–631. https://doi.org/10.1007/s10071-009-0221-4

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). Assessment, testing and comparison of statistical models using R. *Journal of Open Source Software*, *6*(59), 3112. https://doi.org/10.31234/osf.io/vtq8f

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. https://doi.org/10.1098/rspb.2019.0716

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). BayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

ManyDogs Project, Alberghina, D., Bray, E., Buchsbaum, D., Byosiere, S.-E., Espinosa, J., Gnanadesikan, G., Guran, C.-N. A., Hare, E., Horschler, D., Huber, L., Kuhlmeier, V. A., MacLean, E., Pelgrim, M. H., Perez, B., Ravid-Schurr, D., Rothkoff, L., Sexton, C., Silver, Z., & Stevens, J. R. (2022). *ManyDogs Project: A Big Team Science Approach to Investigating Canine Behavior and Cognition*. PsyArXiv. https://doi.org/10.31234/osf.io/j82uc

ManyPrimates, Aguenounon, G., Allritz, M., Altschul, D., Ballesta, S., Beaud, A., Bohn, M., Bornbusch, S., Brandão, A., Brooks, J., Bugnyar, T., Burkart, J., Bustamante, L., Call, J., Canteloup, C., Cao, C., Caspar, K., da Silva, D., de Sousa, A., … Zablocki-Thomas, P. (2022). The Evolution of Primate Short-Term Memory. *Animal Behavior and Cognition*, *9*(4), 428–516. https://doi.org/10.26451/abc.09.04.06.2022

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. https://doi.org/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*(sup1), 99–105. https://doi.org/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. https://doi.org/10.1007/s100710050016

Miletto Petrazzini, M. E., Mantese, F., & Prato-Previde, E. (2020). Food quantity discrimination in puppies (*Canis lupus familiaris*). *Animal Cognition*, *23*(4), 703–710. https://doi.org/10.1007/s10071-020-01378-z

Miletto Petrazzini, M. E., Pecunioso, A., Dadda, M., & Agrillo, C. (2020). Does Brain Lateralization Affect the Performance in Binary Choice Tasks? A Study in the Animal Model Danio rerio. *Symmetry*, *12*(8), 1294. https://doi.org/10.3390/sym12081294

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. https://doi.org/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

Müller, K. (2020). *here: A simpler way to find your files*. https://CRAN.R-project.org/package=here

Neuliep, J. W. (1990). Editorial Bias Against Replication Research. *Journal of Social Behavior and Personality*, *5*(4), 85–90.

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. https://doi.org/10.1016/j.celrep.2017.03.079

Pedersen, T. L. (2020). *patchwork: The composer of plots*. https://CRAN.R-project.org/package=patchwork

Pettersson, H., Kaminski, J., Herrmann, E., & Tomasello, M. (2011). Understanding of human communicative motives in domestic dogs. *Applied Animal Behaviour Science*, *133*(3), 235–245. https://doi.org/10.1016/j.applanim.2011.05.008

Pitak-Arnnop, P., Auychai, P., Subbalekha, K., Sirintawat, N., Tangmanee, C., Meningaud, J.-P., Messer-Peti, R., & Neff, A. (2022). COVID-19 pandemic and lockdown increase the frequency and severity of periorbital dog bite injuries in children. *Journal of Oral Biology and Craniofacial Research*, *12*(5), 666–672. https://doi.org/10.1016/j.jobcr.2022.08.005

Pongrácz, P., Gácsi, M., Hegedüs, D., Péter, A., & Miklósi, Á. (2013). Test sensitivity is important for detecting variability in pointing comprehension in canines. *Animal Cognition*, *16*(5), 721–735. https://doi.org/10.1007/s10071-013-0607-1

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. https://doi.org/10.1037/0735-7036.118.4.375

Quinn, G. P., & Keough, M. J. (2002). *Experimental Design and Data Analysis for Biologists*. Cambridge University Press. https://doi.org/10.1017/CBO9780511806384

R Core Team. (2021). *R: A language and environment 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. https://doi.org/10.1016/j.applanim.2009.05.012

Revelle, W. (2021). *psych: Procedures for psychological, psychometric, and personality research*. https://CRAN.R-project.org/package=psych

Ribeiro, L. de S., Soares, G. M., Arnold, E., & Nobre e Castro, M. C. (2023). Effects of the COVID-19 pandemic on the behavior and physical health of dogs in Rio de Janeiro state: Reflections on the quality of life of dogs and their owners. *Journal of Veterinary Behavior*, *60*, 37–43. https://doi.org/10.1016/j.jveb.2022.12.001

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

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An Excess of Positive Results: Comparing the Standard Psychology Literature With Registered Reports. *Advances in Methods and Practices in Psychological Science*, *4*(2), 25152459211007468. https://doi.org/10.1177/25152459211007467

Scheider, L., Kaminski, J., Call, J., & Tomasello, M. (2013). Do domestic dogs interpret pointing as a command? *Animal Cognition*, *16*(3), 361–372. https://doi.org/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. https://doi.org/10.3389/fpsyg.2012.00558

Sherwell, E.-G., Panteli, E., Krulik, T., Dilley, A., Root-Gutteridge, H., & Mills, D. S. (2023). Changes in Dog Behaviour Associated with the COVID-19 Lockdown, Pre-Existing Separation-Related Problems and Alterations in Owner Behaviour. *Veterinary Sciences*, *10*(3), 195. https://doi.org/10.3390/vetsci10030195

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), 122–126. https://doi.org/10.1037/0735-7036.115.2.122

Soproni, K., Miklósi, Á., Topál, J., & Csányi, V. (2002). Dogs' (*Canis familiaris*) responsiveness to human pointing gestures. *Journal of Comparative Psychology*, *116*(1), 27–34. https://doi.org/10.1037/0735-7036.116.1.27

Stan Development Team. (2020). *RStan: The R interface to Stan*. http://mc-stan.org/
Stevens, J. R. (2017). Replicability and reproducibility in comparative psychology. *Frontiers in Psychology*, *8*, 862. https://doi.org/10.3389/fpsyg.2017.00862
Stevens, J. R., Mathias, M., Herridge, M., Hughes-Duvall, K., Wolff, L. M., & Yohe, M. (2022). Do owners know how impulsive their dogs are? *Animal Behavior and Cognition*, *9*(3), 261–286. https://doi.org/10.26451/abc.09.03.02.2022
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. https://doi.org/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*, Article 1. https://doi.org/10.1038/srep05304
Tauzin, T., Csík, A., Kis, A., Kovács, K., & Topál, J. (2015). The order of ostensive and referential signals affects dogs' responsiveness when interacting with a human. *Animal Cognition*, *18*(4), 975–979. https://doi.org/10.1007/s10071-015-0857-1
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. https://doi.org/10.1016/j.cub.2011.12.018
The ManyBabies Consortium. (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. https://doi.org/10.1177/2515245919900809
Tiedemann, F. (2020). *gghalves: Compose half-half plots using your favourite geoms*. https://CRAN.R-project.org/package=gghalves
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. https://doi.org/10.1017/S0140525X05000129
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. https://doi.org/10.1016/B978-0-12-407818-5.00011-5
Topál, J., Miklósi, Á., Gácsi, M., Dóka, A., Pongrácz, P., Kubinyi, E., Virányi, Z., & Csányi, V. (2009). The dog as a model for understanding human social behavior. In *Advances in the Study of Behavior* (Vol. 39, pp. 71–116). Academic Press. https://doi.org/10.1016/S0065-3454(09)39003-8
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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1371/journal.pbio.2003693
Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686
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. https://doi.org/10.1111/j.1365-2656.2009.01639.x
Wobber, V., Hare, B., Koler-Matznick, J., Wrangham, R., & Tomasello, M. (2009). Breed differences in domestic dogs' (*Canis familiaris*) comprehension of human communicative signals. *Interaction Studies*, *10*(2), 206–224. https://doi.org/10.1075/is.10.2.06wob
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. https://doi.org/10.1016/j.anbehav.2008.03.010
Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman and Hall/CRC. https://yihui.org/knitr/
Xie, Y., Allaire, J. J., & Grolemund, G. (2018). *R markdown: The definitive guide*. Chapman and Hall/CRC. https://bookdown.org/yihui/rmarkdown

Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook*. Chapman and Hall/CRC. https://bookdown.org/yihui/rmarkdown-cookbook

Zhu, H. (2021). *kableExtra: Construct complex table with "kable" and pipe syntax*. https://CRAN.R-project.org/package=kableExtra

Supplemental Material

Methods

Videos of the experimental procedure and questionnaires are available on the Open Science Framework (https://osf.io/9r5xf/).

ManyDogs 1 Detailed Experimental Protocol

Order of Experiment

1. Warm-ups
2. Condition 1 (8 trials)
3. 1-minute play break
4. Refamiliarization (2 trials of 2-cup warm-ups)
5. Condition 2 (8 trials)
6. 1-minute play break
7. Odor control (4 trials)

(Note: purpose of odor control is not to determine whether a given dog can use odor but whether a given population of dogs or a certain lab's dogs can do so.)

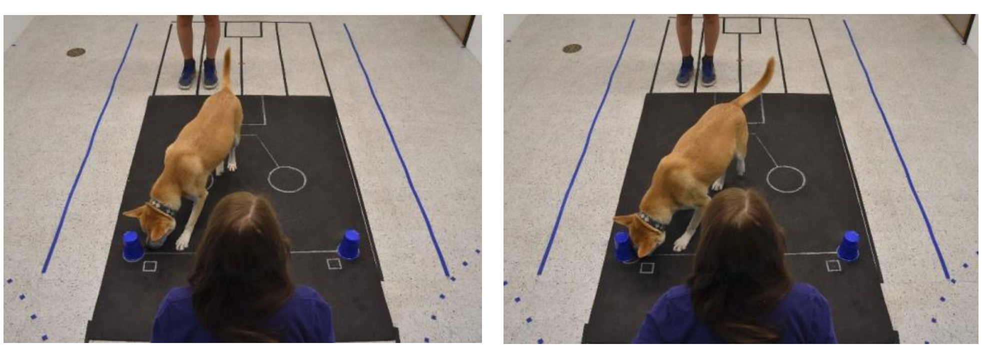
Overall Guidelines for Object Choice Tasks with Cups

"Choice"

A choice is defined as the dog physically touching the cup with their snout or a front paw (NOT an ear, back leg, or tail) (Figure S1). If the dog hits the cup with a back leg or their tail and moves the cup more than a few inches or exposes the treat, redo the trial (experimental error). Each trial allows the dog 25 seconds (s) from the release to make a choice. When a choice is made, Handler (H) or Experimenter (E) says "choice" in a soft, neutral tone (primarily H's job, E can call if H doesn't--usually because the dog is occluding the choice from the handler's angle); if the dog chose correctly, E may help them access the treat if necessary. E then blocks the dog from making a second choice and H retrieves/recalls the dog. If the guardian is handling, E may need to help return the dog to the guardian. If the dog chooses the wrong cup, E does not make a point of showing the unchosen cup to the dog or hiding the removal of the treat from the dog's view. E should, as efficiently as possible, remove the cups after a choice has been made and retrieve any uneaten treats while H is recalling and resetting the dog at start.

Figure S1

Photographic Examples of (a) Not Quite a Choice and (b) a Choice



Presenting the Reward: "Name/Puppy, look!"

E presents the reward by holding the treat up so the dog can see it, and saying "[Name], look!" or "Puppy, look!" During warm-ups, if the dog is not attending, E can try to get their attention by: repeating "name/puppy look", whistling, or leaning closer to the dog and waving her hand around in front of the dog (including such that they can smell the treat & lick E's fingers).

Handler Guidelines

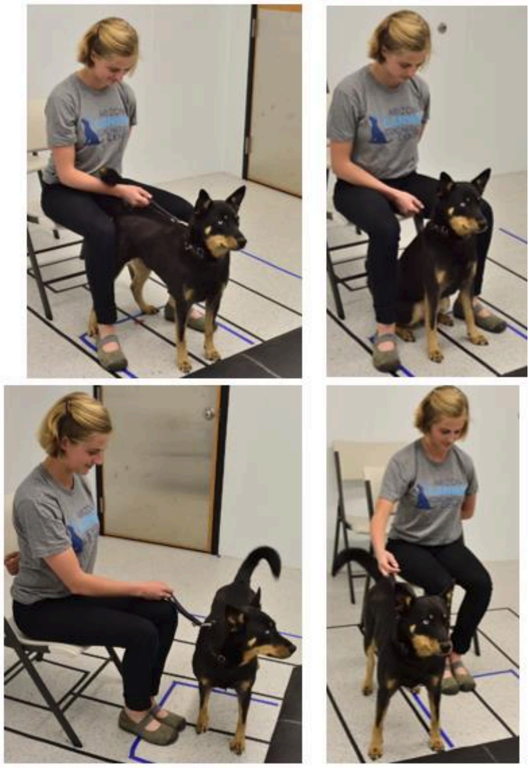
The handler (H) should preferably be a trained lab member who can help with various aspects of the procedure. If this is not possible for whatever reason, the guardian may handle.

H sits in a chair behind the dog, centered behind the starting line (see Figure S2). At the beginning of each trial H centers the dog with the dog's front paws inside a 50 x 50 cm target box, and preferably facing toward E, with their head over the start line on the mat; the dog will likely be between H's knees or ankles, but this is not required. While it is preferred that the dog is standing in a straight line between E and H, if the dog is resistant to being moved, it is fine if their back legs are to the side of the box, so long as they are attending to E. H holds the tab/leash straight back and looks down with their eyes closed. Once E has given the "now" signal, H says "okay!" (or other release command) and drops the dog's tab/leash/collar. For each trial, when E says "now", H releases the leash and says "okay" or other verbal release specified by the guardian. If the dog does not move from the start line within three seconds, E has the option to say "nudge", at which point H nudges the dog (gentle, centered tap on the butt or shoulders). Once the dog leaves the start line, H should not touch the dog until the end of the trial. H should praise the dog on each trial (including test trials) for successful recalls (i.e., after the dog turns away in response to H's recall command), but like E, should not praise the dog for making a choice during test trials.

- Trained handlers: may open their eyes after E says "now" and the dog has been released. Trained handlers are expected to time the trial length (25 s), call choices, and retrieve/recall the dog after each trial.
- Guardian handlers: After E says "choice" or "time", the guardian can open their eyes and recall/retrieve the dog to return to them for the next trial. E may need to help return the dog to the start line.

Figure S2

The Handler's Positioning



Note. Top two photos are preferred. Bottom left is acceptable. Bottom right is not allowed; the dog should be repositioned so that both front paws are in the blue (50 x 50 cm) box.

Experimenter Guidelines

E's resting position = kneeling, looking down at their knees/the floor immediately in front of their knees, and with hands on the side of their thighs and elbows tucked into rib cage (see main Figure 2).
The experimenter (E) begins each phase by kneeling, centered behind the experimenter's line (~30 cm behind the cup line). The cup line and locations should be **clear of all stimuli at the start of each trial**. This allows the dog to observe cups being placed at the start of trials across all phases, demonstrating that there are no hidden treats under the cups, and they only need to track the one that E shows during baiting.

During warm-ups and ostensive trials, E presents the reward from the center position, says "[Name], look!", then baits the appropriate cup and repeats the "[Name], look!", and either performs the appropriate social cue or kneels in resting position.

At the end of the baiting and cueing procedure, from resting position E says "now" in a soft neutral tone to signal H to release the dog's leash and say "okay" (or another release command that a given dog is reported to have).

If the dog does not move from the start line within three seconds, E can say "nudge", at which point H nudges the dog. (Nudges should be recorded at the end of each trial, or from video post session). If the dog does not move from the start line after another 3 seconds, E may repeat "now", and the handler "OK" once more. The order of "nudge" and the second verbal release can be reversed if it seems appropriate with a given dog (i.e., if they are not comfortable with a stranger touching their hindquarters and are likely to startle or based on what worked in past trials with that dog).

If the dog chooses correctly, they may move the cup such that they access the treat themselves, but if they are unable to get the treat, E tips/lifts the cup to let them access the treat. E may need to prevent the dog from making a second choice (either block the dog's path or remove the cups).

Note, on every trial, **E will give two attention directing verbal/vocal signals** prior to signally H to release the dog. For warm ups, the first verbal cue is just prior to showing the treat, and the second just prior to baiting. On test trials, the first cue occurs just prior to showing the treat, and the second just prior to executing the point.

Warm-Ups

All warm-ups should take place in 15 minutes or less. After 15 minutes, the warm-ups should be halted and the dog moved on to the test trials. If warm-ups end because the time limit has been reached, this should be indicated on the coding sheet and noted during data entry.

The dog is required to pass a warm-up criterion prior to completing any other tasks with cups. The false-baited cups should be used throughout all warm-ups. Warm-up trials ensure that the dogs are motivated to search for the treat and to prevent side biases. These trials consist of three phases: (1) no-cup visible placement and free-form cup familiarization game, (2) one-cup alternating visible placement, and (3) two-cup visible placement.

Positioning and Setup

H holds the dog at the starting line. E kneels in the center, approximately 30 cm behind the cup line or 1.6 m from the dog (main Figures 1-2). When baiting, E should lift the cup approximately 10 cm off the ground (visible but not too high).

E Guidelines for Warm-Ups

E should praise the dog for retrieving treats throughout warm-ups. E should look down in resting position after baiting while dog makes their choice.

Phase 1 – No-Cup Visible Placement and Free-Form Cup Game

Visible Treat Placement

First, there are two repetitions of treat-only ("no-cup") trials to make sure the dog likes the treat and is willing to eat it off the testing area floor.

1. 1st visible placement. E says "[Name], look!", then presents the treat by holding it out between thumb and forefinger in front of their body, and tries to make eye contact with the dog.
2. E repeats "[name], look!" as they visibly place the treat on the floor halfway between the dog and experimenter.
3. E returns to resting position with their hands on the sides of their thighs, looks down at the treat, and says "now" in a neutral tone, signaling H to say "okay" and release the dog's leash. The dog is then allowed to approach the treat and retrieve it.
4. 2nd visible placement. Repeat steps 1-3, but place the treat directly in front of E on the cup line to encourage the dog to approach closer to E.

NOTE: If the dog does not approach the treat within 25 s, the trial is repeated. If the dog approaches/climbs on E or sits and waits for E, the trial may be repeated until the dog retrieves the treat immediately on its own.

Free-Form Cup Games

After the dog has retrieved the treat successfully from each visible placement, E plays a free-form cup game with the dog to familiarize them with finding treats under cups (Figure S3).

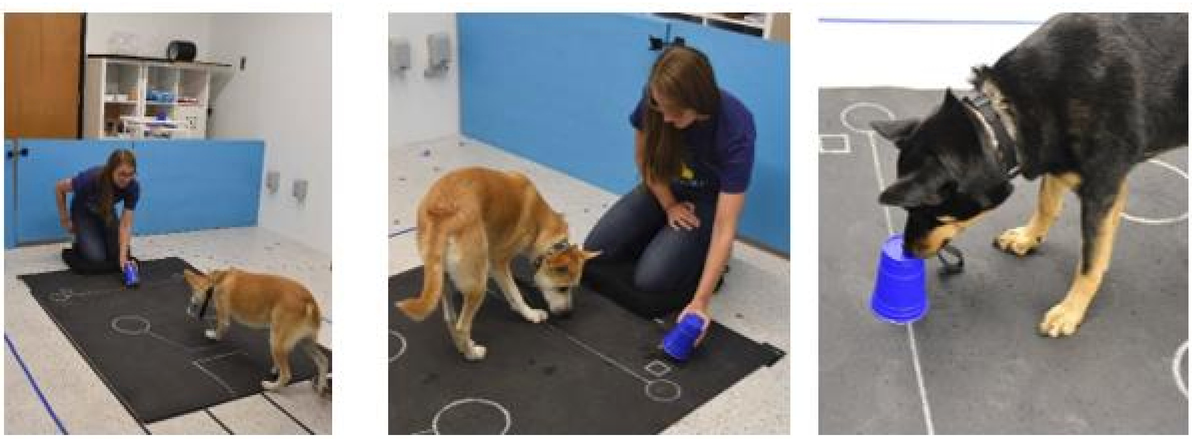
1. The dog is held at the start line.
2. Is at resting position and places a single cup in front of them on the cup line.
3. E says "[name], look!" and presents a treat, before hiding it under the cup.
4. E says "now" in a neutral tone which cues H to say "okay" and releases the dog to search and interact with the cup.
5. When the dog touches the cup, E can lift the cup away and the dog is allowed to eat the treat underneath.
6. E should set up a minimum of **3 cup touches/trials**, but can use more if necessary for a given dog that hesitates to interact with the cups. This game may take 1-2 minutes. The goals are to build an association between the cups and rewards and to shape the behavior of touching the cups.

NOTES:

- As needed, E can tilt, lift, or tap the cup to draw the dog's attention to the treat and encourage them to touch the cup (but do not point).
- These hiding instances happen in quick succession without recalls to keep the dog fully engaged. H does not recall the dog until E indicates that cup games are complete.
- The cup can be placed in various locations on or near the cup line.
- Avoid baiting one of the search locations more than the other during this phase as that could potentially create bias in the dog.
- Although we have set a minimum passing criterion for cup games, more trials can be done if the experimenter feels they would be beneficial. Please note that this is the only stage of the process with this flexibility; additional trials are not allowed in subsequent phases of warm-ups, so now is the time to make sure the cup-reward association is strong.

Figure S3

Free-Form Cup Games



Note. Building the association between touching the cup and getting treat, no matter where the cup is.

Phase 2 – One-Cup Alternating Visible Displacement

Moving directly from free-form cup games once the dog confidently approaches and touches the cup, E begins one cup alternating warm-ups. This phase of warm-ups familiarizes the dog with the setup and assures that the dog is motivated to find the treat. The dog is required to successfully retrieve the treats on 4 trials, twice on each side, to move on (within a maximum of seven trials). These trials do not need to be consecutive. Refer to the sequences pre-set in the coding sheet for treat location order.

The dog is held at the start line by H, E is at resting position.

1. E places one cup next to the location indicated by the pre-set order.
2. E says "[Name], look!", then presents the treat by holding it out between thumb and forefinger in front of their body.
3. E repeats "[name], look!", trying to make eye contact with the dog while placing the treat at either the R or L position, lifting the cup to cover the treat (see Figure S3).
   3.1. Recommendation: Bait the cup with the opposite hand to avoid moving the cup in front of the treat and blocking the dog's sight line. I.e., if baiting the R location, hold the treat in the L hand and use the R hand to lift the cup and cover the treat deposited by the L hand.
4. After baiting, E kneels at the E location in resting position (looking at knees or at floor right in front of knees) and after a brief pause of ~2 seconds says "now" in a neutral tone to cue H to give the verbal release "okay" and allow the dog to search.
   4.1. If the dog does not approach the cup in 25 s, E may call the dog over, show the treat if necessary, encourage them to touch the cup on their own and reward them for touching it by lifting the cup away, and allow them to eat the treat that was under the cup.
5. When the dog makes contact with the cup, E/H verbally marks the touch in a neutral tone.
   5.1. Reminder: the dog should be praised by E/H on warm-ups.
6. H recalls the dog to reset for the next trial.
7. E removes the cup until the dog is facing them from the start line and can observe the cup being placed for the next trial.
8. Steps 1-7 are repeated for 4 correct trials, or until the dog has had 7 total trials.
   8.1. Repeating these trials serves as a correction procedure for spontaneous side biases and ensures that dogs gain experience finding the treats in both locations.
   8.2. Once the dog has correctly touched the cup on four trials, move on to phase 3 (please do not do any extra trials at this stage, move on directly once they reach four correct touches).

Phase 3 – Two-Cup Visible Displacement

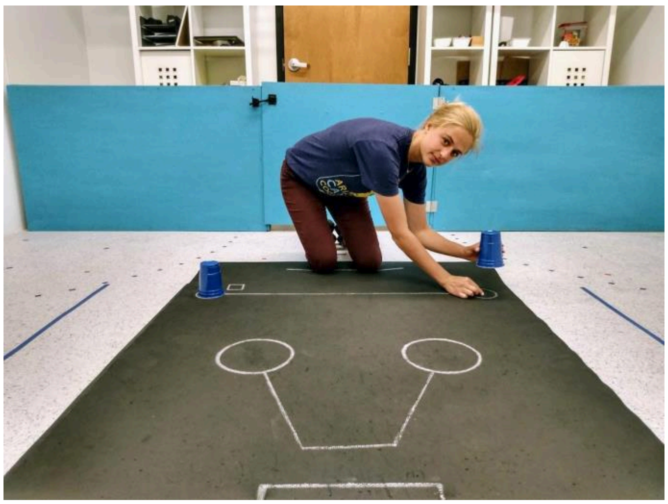
A predetermined sequence of baiting the R and L locations is set out in the coding sheet for E to follow. This phase of warm-ups assures that the dog is not choosing cups randomly and is fixing attention on the experimenter's actions (Figure S4). Dogs are required to choose correctly on the first presentation (i.e. column A of the datasheet) for four out of six consecutive trials to advance to test trials within a maximum of 20 trials (including repeated trials, but not refamiliarizations).

1. With the dog watching from the start line, E places both cups at the R and L locations on the floor.
2. E says "[Name], look!", then presents the treat by holding it out between the thumb and forefinger in front of their body.
3. E repeats "[name], look!", trying to make eye contact with the dog while placing the treat at either the R or L position, lifting the cup already in place to cover the treat (Figure S4).
4. After baiting, E kneels at the E location in resting position and after a brief pause of ~2 seconds says "now" in a neutral tone to cue H to give the verbal release "okay" and allow the dog to search.
5. The dog is then allowed 25 s to make a choice. If the dog touches the baited cup before the empty cup, the dog is allowed to have the treat.
6. Upon the dog making a correct choice, E praises the dog and H recalls and resets for the next trial.
   6.1. If the dog chooses the incorrect cup, E can either say "choice/touch" or "miss" in a neutral tone, (if the verbal marker being used depends on the dog's choice, say "miss", if using the same verbal marker regardless of choice, say "choice/touch" "hit/miss"). The dog is not allowed to access the other cup with the treat, and that specific trial is repeated *until the dog chooses correctly*. If the dog does not choose any cup within 25 s, the trial is repeated. See the Refamiliarization procedure below.

7. E removes the cups while the dog is being reset, waiting to place them at the R and L locations for the next trial until the dog is at the start line and facing forward.
8. Steps 1-7 are repeated until the dog:
   8.1. meets the criteria of 4 correct trials within the sliding window of 6 trials (at which point, please proceed to test trials; do not perform additional trials at this stage).
   8.2. has completed a maximum of 20 trials *including* repeated trials.
   8.3. has *not* met criteria but a total of 15 minutes has elapsed since the beginning of the warm-ups.

Figure S4

Two-Cup Visible Displacement from the Dog's Point of View



Note. Baiting and eye-contact as in both one-cup and two-cup visible displacement.

Refamiliarization Procedure

If the dog does not choose a side for two trials (i.e., No-Choice, NC) in a row, go back to the previous step of warm-ups. Specifically, if the dog no-choices twice in a row during phase 2 (one cup), resume cup games until the dog seems reengaged; then resume trials. If the dog does not choose on two trials in a row during phase 3 (two cups), conduct 2 trials of phase 2 (one-cup alternating).

Abort Criteria/Procedure

- If the dog stops interacting with cup games for about 1 min, try increasing treat value. If the dog still does not engage, abort the session.
- If the dog no-choices twice in a row, refamiliarize with the procedure as explained above.
- If there are 4 NCs total within either phase 2 or phase 3 of warm-ups (including refamiliarizations), abort the session. (In other words, the count resets between phase 2 and phase 3.)
- If 7 trials of phase 2 or 20 trials of phase 3 have been conducted (including repeated trials, but not refamiliarizations), and the pass criterion has not been met, move on to test trials, noting that warm-up criteria were not met.
- If warm-up tasks take up the allotted 15 minutes, move on to test trials noting that warm-up pass criteria were not met.

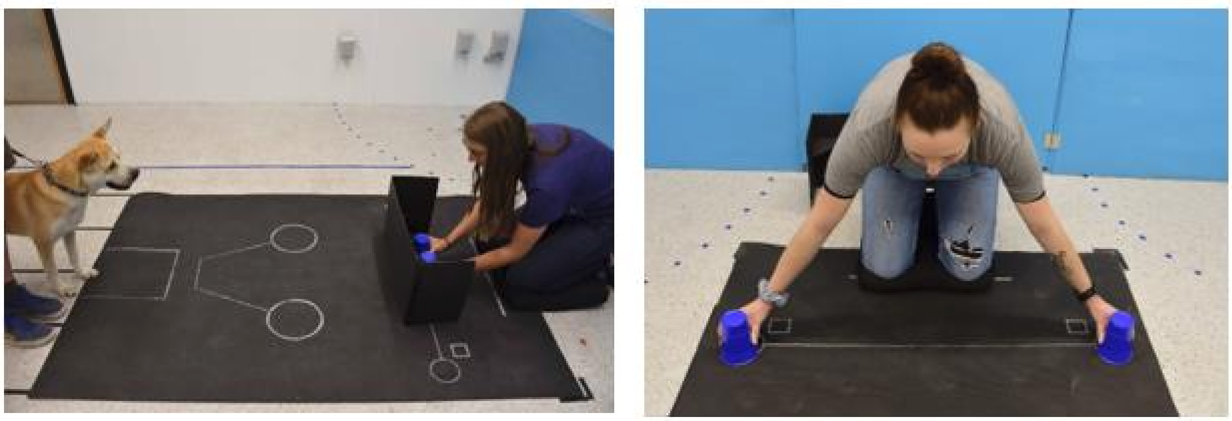
NOTE: Dogs may continue to test trials if they have been consistently making choices, but not if they have not passed due to no-choice responses. Dogs who abort warm-ups may not ultimately be included in the final main analysis but will used for exploratory analyses. If there is limited time (or if in a situation where many other dogs are available in a fixed time), a dog who does not pass warm-ups may abort at this stage (but the collected data must still be reported).

Test Trials with Social Cues

In the test trials, baiting will be occluded. E places cups on the line connecting the final cup locations with the occluder ~10 cm behind, blocking the dog's view of the cups. E presents the subject with the reward (either ostensively or non-ostensively), places the treat in front of one of the cups (but near the center), lifts both cups at the same time, and places them over the treat and empty space (Figure S5A-B). E places the occluder behind them and slides the cups into the positions marked on the mat, while looking straight down (Figure S5B). (Note: it is better for the cups to not end up precisely in the circles than to look to one side or the other.)

Figure S5

The Protocol for Test Trials with Social Cues



A   B

Note. (A) Occluded Baiting. The dog cannot see the cups or the treat while baiting, but the occluder is small enough that the dog can see the experimenter. (B) Moving the cups from their occluded position out to their final positions (in the cup circles) while looking down. Note that the cups are held close to the base for maximum stability and control - they are less likely to tip over with this hold.

There are two styles of social cues being presented in counterbalanced block order between participants; "ostensive" social cueing, and "non-ostensive" social cueing. Each style is presented as 1 block of 8 trials, and the order of the blocks is determined according to the 4 coding sheets that have been prepared for this study. Dogs should complete both blocks (total of 16 test trials) and sites should use equal numbers of the 4 coding sheets so that equal numbers of dogs will be assigned to each order (ostensive first vs. non-ostensive first, to control for learning or other order effects). In the minimum sample, this will result in 4 dogs completing both blocks of test trials on each of the 4 coding sheets, for a total of 16 dogs (8 non-ostensive first and 8 ostensive first).

NOTE: test trials are *not* repeated if the dog chooses the un-baited cup (i.e., does not follow the point). They are only allowed to make 8 cup choices in total for each test block. The only time test trials are repeated if the dog does a "no-choice."

Pointing

In both conditions, the point should consist of the index finger of the contralateral hand extended approximately 20 cm from the cup. The point should be a direct movement of the hand from the resting position on the side of the thigh, across the body, to the direction of the point (without lifting the hand above the sternum or activating the elbow more than absolutely necessary). E's torso should remain facing squarely forward (not twisted towards one side, by keeping a soft bend in the elbow throughout the point it is easier to keep shoulders from rotating). Note: the physicality of the point may vary slightly from experimenter to experimenter, but each site's experimenters should endeavor to use the exact same style of physical point and torso position for each condition (with only the ostensive vocal and gaze components differing) so as to be consistent within their sample of dogs.

We have provided video examples of how the point should be performed, and as part of the set-up process we will be reviewing experimenter practice videos from each site to provide feedback and assistance with mastering the gesture and general aspects of the protocol as necessary.

The timing of the pointing gesture should be as follows:

From resting position, E:

- Creates contralateral point gesture, taking 1 second to move hand from side thigh to completed point.
- Holds point gesture stationary for 2 seconds.
- Reverses arm motion to return to resting position, taking 1 second to retract point gesture.
- Waits in resting position (looking down at knees/floor in front of knees, arms tucked in at sides) for 1 second after retracting point before signaling H.

In resting position E says "now" in a low monotone voice to signal H to say "okay" (or other release command) and release the leash. Until this point, H should be holding the leash/collar/dog but looking down with their eyes closed so that H is unaware which side is being indicated. Once H releases the dog, the dog has 25s to make a choice.

On test trials E does not praise the dog. If the dog requires refamiliarizations, E may praise on refamiliarizations. H should always praise the dog when they are successfully recalled to the starting box, not when they make a choice during test trials.

At the beginning of test trials, E starts at the E location with false baited cups and occluder within easy reach but **not on or hiding the cup line**. This is to ensure that the dog can see when the stimuli are placed and know that no other treats are being hidden other than the single treat presented on each trial.

H holds the dog at the start box.

1. At the start of each trial, E places both cups to either side of the midpoint of the cup line connecting the R and L locations, while the dog watches from the start box.
2. E places the occluder ~10cm in front of the cups where they are set on the cup line, hiding the cups from the dog's view.
3. E presents the treat over the top of the occluder with their dominant hand in one of the following ways, determined by the condition:
   3.1. **Ostensive condition:** E says "[name], look!", and presents the treat making eye contact,
   3.2. **Non-ostensive condition:** E clears their throat, then presents the treat without making eye contact (looking down at the floor)
4. E places the treat on the floor between the cups
5. Working behind the occluder, E simultaneously picks up both cups at the same time and places one over the treat, following the pre-set sequence of R and L on the coding sheet, and places the other (empty) cup next to it.
   5.1. Both cups should be moved in perfect synchrony to avoid any unintentional cuing to the dog.
6. E removes the occluder from in front of the cups and places it behind them.
7. E simultaneously slides the cups apart to the R and L positions marked on the mat while looking at the center of the cup line.

7.1. Avoid sliding the cups too quickly and hold near the cup opening to avoid pushing a cup over while it's in motion.

8. Then E places both hands on the sides of their thighs in resting position, with elbows tucked in close to their ribcage (main Figure 2).
9. E gives the social cue:
   9.1. **Ostensive** condition: E makes eye contact with the dog, E says in a high-pitched voice "[Name], look!", before pointing toward the cup with the reward. E maintains eye contact with the dog with torso squared forward throughout verbal and point cues following the timing detailed above (main Figure 3). E maintains a calm, neutral facial expression.
   9.2. **Non-ostensive** condition: E keeps gaze averted from the dog by looking down, E clears throat and points toward the cup with the reward, looking down at their elbow and keeping torso squared forward, following the timing detailed above (main Figure 3). E maintains a calm, neutral facial expression.
10. After completing the point gesture, E assumes resting position (looking down at knees/floor in front of knees, hands on the side of thighs with elbows tucked in) and holds for 1 s, then from resting position E says "now" in a soft neutral tone which cues H to give the verbal release "okay" and let the dog go to search.
11. When the dog touches one of the cups, E/H will give the verbal marker of "choice", "touch" or "miss" (but no praise) in a neutral tone and, if necessary, assist the dog to search by tilting the chosen cup back or lifting slightly to allow them to search underneath.
    11.1. If the dog chooses the baited cup they receive the treat hidden there. While showing the dog what is under the chosen cup, E simultaneously prevents the dog from making more than one choice, removing the unchosen cup from access, if necessary. (With tab-leash handling and a quick handler, this may not be necessary.)
    11.2. If the dog chooses the un-baited cup, they do not receive anything on that trial. E and H should ensure that they do not try to choose the other cup (removing from reach, if necessary).
        11.2.1. Remember: Incorrect choice trials are not repeated.
        11.2.2. NOTE: it is important to avoid moving from resting position to assist/remove the unchosen cup until the dog has made a choice, this could result in unintentional cuing by moving early.
        11.2.3. NOTE: the verbal marker used should be consistent within the research group, but it is up to the group to determine if they want to use the same verbal marker independent of dog's choice (i.e., "Choice), or if they want a different verbal marker for correct and incorrect choices (i.e., "choice" when correct, "miss" when incorrect).
12. After making a choice, the dog is recalled by H, praised for recall and reset at the start line for the next trial.
13. Repeat steps 1-13 for a total of 8 trials of one type of social cue
14. Following the 8th trial, take a 1-minute play break.
15. After the play break, resume testing with a short refamiliarization of 2-3 2-cup warm-up trials to re-engage the dog.
16. Immediately after the 2-cup refamiliarization, begin block 2 of test trials (the condition order in each block of test trials is counterbalanced on the coding sheets, do not deviate from this pre-set order).
17. Follow steps 1-14 to complete block 2 of test trials.

Abort Criteria

If the dog does not make a choice within 25 s, the trial is repeated. This applies to all trial types within the protocol. If the dog no-choices (NC) twice in a row, refamiliarize with two trials of two-cup alternating warm-ups. Also abort a given condition after 4 total NCs, including those on refamiliarization trials. If a dog aborts one condition, continue on to the play break, refamiliarization, and the other condition.

Odor Control

Use the occluder and clean, un-baited cups for this task without a treat taped into the cup. Only **one** verbal cue is given when presenting the treat (similar to ostensive trials, using eye-contact and "[dog name], look"), and no pointing cue is administered on the four odor control trials. Baiting procedures and choice criteria are the same as in test trials. The dependent measure for this task is the percent of trials that a dog chooses the baited cup.

NOTE: the point of odor control here is not to determine whether a given dog can use odor but whether a given population of dogs or a certain lab's dogs can do so.

1. E places the cups on the center of the cup line, just to either side of the midpoint while the dog observes from the start box.
2. E places the occluder ~10 cm in front of the cups hiding them from the dog's view.
3. E presents the treat similar to the ostensive test condition above, i.e., says "[Name] look!" and makes eye contact with the dog when presenting the treat.
4. E places the treat under one of the cups behind the occluder (simultaneously moving the cups to avoid accidental cuing) according to the sequence on the coding sheet.
5. E removes the occluder from in front of the cups and places it behind them.
6. E simultaneously moves the cups from the center to their lateral positions while looking down at the midpoint of the cup line
7. After moving the cups to the R and L locations, E pauses in resting position for ~2s then says a neutral "now", signaling H to say "okay" and release the leash for the dog to make a choice, maintaining their downward gaze for the trial duration.
8. The dog is allowed to make one choice and receives the hidden treat if they touch the baited cup before the empty cup.
   8.1. E should remove the unchosen cup from access as in test trials.
   8.2. E should praise the dog for a correct choice.
   8.3. **If the dog no-choices, move on to the next trial.**
9. H recalls the dog to the start box
   9.1. While the dog is being reset, E clears the cups and any treats off the floor.
10. Steps 1-9 are repeated for four trials according to the sequence of R and L on the coding sheet.

ABC 2023, 10(3): 232-286

Animal Behavior and Cognition

DOI: https://doi.org/10.26451/abc.10.03.03.2023

Table S1

Site	Lead	Location	Data collection team	Included dogs	Purebred	M_{age}±SD (range), yr	Sex (M:F)	Desexed	Housing*	Testing	Protocol
Animal Health and Welfare Research Centre (AHWRC)	Marianne Freeman	Winchester, United Kingdom	2	18	9	3.3 ± 2.6 (0.7-9)	8:10	8	P = 18 G = 0	Lab	USCEC 6321 (ERSG)
Arizona Canine Cognition Center (ACCC)	Evan MacLean	Tuscon, AZ, USA	3	18	10	6.4 ± 3.4 (1.3-12.7)	9:9	16	P = 18 G = 0	Lab	16-175 (IACUC)
Auburn Canine Performance Sciences (ACPS)	Lucia Lazarowski	Auburn, AL, USA	3	23	17	4 ± 2.1 (0.9-8.9)	10:13	19	P = 9 G = 14	Lab, Facility	2020-3725 (IACUC)
Boston Canine Cognition Center (BCCC)	Angie Johnston	Boston, MA, USA	3	19	9	4.4 ± 3.1 (0.9-12.1)	8:11	18	P = 19 G = 0	Lab	2020-001-01 (IACUC)
Brown Dog Lab (BDL)	Daphna Buchsbaum	Providence, RI, USA	4	35	8	4.4 ± 2.7 (0.9-11)	19:16	33	P = 35 G = 0	Lab	20-05-0002 (IACUC)
Canid Behavior Research Group (CBRG)	Camila Cavalli	Buenos Aires, Argentina	3	32	15	5.3 ± 2.7 (1-9.4)	12:20	25	P = 32 G = 0	Home	124-22 (CICUAL)
Canine Cognition Center at Yale (CCC)	Laurie Santos	New Haven, CT, USA	3	19	12	4.3 ± 2.9 (0.7-11.4)	8:11	17	P = 19 G = 0	Lab	2020-11448 (IACUC)
Canine Cognition and Human Interaction Lab (CCHIL)	Jeffrey Stevens	Lincoln, NE, USA	5	33	15	4.5 ± 2.8 (0.5-11.8)	19:14	32	P = 33 G = 0	Lab	2132 (IACUC); 20491 (IRB)
Canine Companions (CCI)	Emily Bray	Santa Rosa, CA, USA	3	16	5	1.6 ± 0.1 (1.3-1.8)	4:12	6	P = 1 G = 15	Facility	16-175 (IACUC)
Canine Research Unit (CRU)	Carolyn Walsh	St. John's, NL, Canada	3	18	12	4.6 ± 2.9 (1.6-12.9)	10:8	18	P = 18 G = 0	Lab	22-01-CW (ACC)
Clever Dog Lab† (CDL)	Ludwig Huber	Vienna, Austria	3	61	60	5.13 ± 3.31 (1-12)	26:35	25	P = 61 G = 0	Lab	081/05/2020 (ETK)
Comparative Cognition Lab (CCL)	Debbie Kelly	Winnipeg, MB, Canada	3	20	16	4.8 ± 3.2 (0.7-12)	11:9	20	P = 20 G = 0	Lab	F21-019 (AC11704) (ACC)

Site	Lead	Location	Data collection team	Included dogs	Purebred	M_{age}±SD (range), yr	Sex (M:F)	Desexed	Housing*	Testing	Protocol
Comparative Cognitive Science Lab (CCSL)	Ljerka Ostojić	Rijeka, Croatia	2	20	10	5.1 ± 2.8 (1-10.1)	5:15	16	P = 20 G = 0	Lab	Exempt (Narodne novine nr. 102/17 and 32.19)
Consultorio Comportamentale (CC)	Daniela Alberghina	Messina, Italy	2	17	12	5.4 ± 3.4 (1.1-11.6)	9:8	4	P = 17 G = 0	Lab	065_2021 (CE)
Department of Psychology and Individual Differences (DPID)	Anna Reinholz	Warsaw, Poland	4	32	26	5.2 ± 2.9 (0.7-12)	12:20	18	P = 32 G = 0	Lab	Exempt (Dz. U. 2021 poz. 1331)
Dog Cognition Centre (DCC)	Juliane Kaminski	Portsmouth, United Kingdom	5	38	21	5.7 ± 3.6 (1.4-14.1)	23:15	33	P = 38 G = 0	Lab	522D (AWERB)
Duke Canine Cognition Center (DCCC)	Brian Hare	Durham, NC, USA	3	19	7	5.1 ± 3.5 (0.3-12.5)	8:11	15	P = 19 G = 0	Lab	A150-20-07 (IACUC)
Leader Dogs for the Blind[‡] (LDB)	Sarah-Elizabeth Byosiere	Rochester, MI, USA	2	16	10	1.3 ± 0 (1.2-1.3)	7:9	12	P = 1 G = 13 O = 2	Facility	DR-Dog Percept 11/21 (IACUC)
Social Cognition Lab (SCL)	Valerie Kuhlmeier	Dundalk, ON, Canada	2	21	17	3.5 ± 4.5 (0.3-20.8)	14:7	10	P = 16 G = 5	Lab, Facility	2022-2264 (UACC)
The Family Dog Project (TFDP)	Andrea Sommese	Budapest, Hungary	3	18	13	3.8 ± 3.1 (0.7-12.6)	7:11	12	P = 18 G = 0	Lab	Exempt (1998. évi XXVIII. Törvény, 3. 1/9)
Thinking Dog Center (TDC)	Sarah-Elizabeth Byosiere	New York City, NY, USA	3	23	14	4.1 ± 2.8 (0.6-9.1)	8:15	22	P = 23 G = 0	Lab	DR-Dog Percept 11/21 (IACUC)

*For housing types, P = Private, G = Group, and O = Other; [†]Clever Dog Lab participated only in the pilot data collection; [‡]Leader Dogs for the Blind data collection carried out by TDC with Leader Dog personnel assistance.

Animal Behavior and Cognition
ABC 2023, 10(3): 232-286
DOI: https://doi.org/10.26451/abc.10.03.03.2023

Table S2

Results of GLMM of the Dogs' Choice Performance (Pilot Experiment)

effect	Estimate	SE	Lower CI	Upper CI	Chi-square	df	p	BF
(Intercept)	0.12	0.13	-0.13	0.35				
Condition	0.29	0.13	0.03	0.55	5.11	1	0.02	3.92
Condition order	0.06	0.13	-0.20	0.32	0.19	1	0.67	0.38
Trial number	-0.10	0.07	-0.23	0.03	2.32	1	0.13	0.50
Sex	-0.08	0.14	-0.34	0.19	0.37	1	0.54	0.42
Age	-0.01	0.07	-0.13	0.12	0.03	1	0.85	0.17
Training score	0.09	0.07	-0.05	0.23	1.96	1	0.16	0.43

Table S3

One-Sample T-Tests for Each Condition and Site

Site	Ostensive	Non-ostensive	Odor Control
ACCC	$M = 0.58$, 95% CI $[0.50, 0.66]$, $t(17) = 2.01$, $p = .061$, $\mathrm{BF}_{10} = 1.24$	$M = 0.53$, 95% CI $[0.44, 0.63]$, $t(17) = 0.74$, $p = .472$, $\mathrm{BF}_{10} = 0.31$	$M = 0.48$, 95% CI $[0.37, 0.59]$, $t(17) = -0.36$, $p = .725$, $\mathrm{BF}_{10} = 0.26$
ACPS	$M = 0.47$, 95% CI $[0.41, 0.53]$, $t(22) = -1.14$, $p = .266$, $\mathrm{BF}_{10} = 0.39$	$M = 0.48$, 95% CI $[0.42, 0.54]$, $t(22) = -0.57$, $p = .575$, $\mathrm{BF}_{10} = 0.25$	$M = 0.53$, 95% CI $[0.47, 0.60]$, $t(21) = 1.14$, $p = .266$, $\mathrm{BF}_{10} = 0.40$
AHWRC	$M = 0.60$, 95% CI $[0.54, 0.66]$, $t(17) = 3.50$, $p = .003$, $\mathrm{BF}_{10} = 15.70$	$M = 0.55$, 95% CI $[0.46, 0.63]$, $t(17) = 1.20$, $p = .248$, $\mathrm{BF}_{10} = 0.45$	$M = 0.52$, 95% CI $[0.38, 0.65]$, $t(15) = 0.25$, $p = .806$, $\mathrm{BF}_{10} = 0.26$
BCCC	$M = 0.47$, 95% CI $[0.38, 0.55]$, $t(18) = -0.81$, $p = .426$, $\mathrm{BF}_{10} = 0.32$	$M = 0.47$, 95% CI $[0.40, 0.54]$, $t(18) = -0.78$, $p = .448$, $\mathrm{BF}_{10} = 0.31$	$M = 0.52$, 95% CI $[0.42, 0.61]$, $t(18) = 0.34$, $p = .735$, $\mathrm{BF}_{10} = 0.25$
BDL	$M = 0.54$, 95% CI $[0.49, 0.59]$, $t(34) = 1.47$, $p = .152$, $\mathrm{BF}_{10} = 0.48$	$M = 0.53$, 95% CI $[0.47, 0.59]$, $t(34) = 1.10$, $p = .280$, $\mathrm{BF}_{10} = 0.32$	$M = 0.53$, 95% CI $[0.47, 0.59]$, $t(33) = 1.07$, $p = .292$, $\mathrm{BF}_{10} = 0.31$
CBRG	$M = 0.43$, 95% CI $[0.36, 0.51]$, $t(31) = -1.73$, $p = .094$, $\mathrm{BF}_{10} = 0.71$	$M = 0.45$, 95% CI $[0.40, 0.50]$, $t(31) = -1.88$, $p = .070$, $\mathrm{BF}_{10} = 0.90$	$M = 0.45$, 95% CI $[0.38, 0.53]$, $t(31) = -1.27$, $p = .215$, $\mathrm{BF}_{10} = 0.39$
CC	$M = 0.51$, 95% CI $[0.43, 0.58]$, $t(16) = 0.21$, $p = .835$, $\mathrm{BF}_{10} = 0.25$	$M = 0.50$, 95% CI $[0.44, 0.56]$, $t(16) = 0.00$, $p > .999$, $\mathrm{BF}_{10} = 0.25$	$M = 0.45$, 95% CI $[0.33, 0.58]$, $t(15) = -0.80$, $p = .439$, $\mathrm{BF}_{10} = 0.34$
CCC	$M = 0.57$, 95% CI $[0.50, 0.65]$, $t(18) = 2.00$, $p = .061$, $\mathrm{BF}_{10} = 1.22$	$M = 0.49$, 95% CI $[0.40, 0.57]$, $t(18) = -0.33$, $p = .749$, $\mathrm{BF}_{10} = 0.25$	$M = 0.62$, 95% CI $[0.47, 0.78]$, $t(7) = 1.87$, $p = .104$, $\mathrm{BF}_{10} = 1.12$
CCHIL	$M = 0.56$, 95% CI $[0.50, 0.62]$, $t(32) = 2.00$, $p = .054$, $\mathrm{BF}_{10} = 1.08$	$M = 0.54$, 95% CI $[0.49, 0.60]$, $t(32) = 1.51$, $p = .140$, $\mathrm{BF}_{10} = 0.52$	$M = 0.52$, 95% CI $[0.42, 0.61]$, $t(32) = 0.33$, $p = .744$, $\mathrm{BF}_{10} = 0.20$
CCI	$M = 0.45$, 95% CI $[0.36, 0.53]$, $t(15) = -1.39$, $p = .186$, $\mathrm{BF}_{10} = 0.57$	$M = 0.45$, 95% CI $[0.36, 0.53]$, $t(15) = -1.39$, $p = .186$, $\mathrm{BF}_{10} = 0.57$	$M = 0.50$, 95% CI $[0.42, 0.58]$, $t(15) = 0.00$, $p > .999$, $\mathrm{BF}_{10} = 0.26$
CCL	$M = 0.43$, 95% CI $[0.36, 0.50]$, $t(19) = -2.07$, $p = .053$, $\mathrm{BF}_{10} = 1.33$	$M = 0.52$, 95% CI $[0.45, 0.59]$, $t(19) = 0.63$, $p = .533$, $\mathrm{BF}_{10} = 0.28$	$M = 0.50$, 95% CI $[0.46, 0.54]$, $t(19) = 0.00$, $p > .999$, $\mathrm{BF}_{10} = 0.23$
CCSL	$M = 0.55$, 95% CI $[0.47, 0.63]$, $t(19) = 1.32$, $p = .202$, $\mathrm{BF}_{10} = 0.50$	$M = 0.49$, 95% CI $[0.44, 0.54]$, $t(19) = -0.52$, $p = .606$, $\mathrm{BF}_{10} = 0.26$	$M = 0.52$, 95% CI $[0.47, 0.57]$, $t(11) = 1.00$, $p = .339$, $\mathrm{BF}_{10} = 0.44$

CRU	$M = 0.60$, 95% CI $[0.53, 0.67]$, $t(17) = 2.96$, $p = .009$, $\mathrm{BF}_{10} = 5.91$	$M = 0.48$, 95% CI $[0.40, 0.56]$, $t(17) = -0.53$, $p = .604$, $\mathrm{BF}_{10} = 0.28$	$M = 0.47$, 95% CI $[0.35, 0.59]$, $t(15) = -0.56$, $p = .580$, $\mathrm{BF}_{10} = 0.29$
DCC	$M = 0.56$, 95% CI $[0.49, 0.63]$, $t(37) = 1.82$, $p = .077$, $\mathrm{BF}_{10} = 0.77$	$M = 0.59$, 95% CI $[0.53, 0.65]$, $t(37) = 3.23$, $p = .003$, $\mathrm{BF}_{10} = 13.20$	$M = 0.51$, 95% CI $[0.45, 0.56]$, $t(34) = 0.18$, $p = .856$, $\mathrm{BF}_{10} = 0.18$
DCCC	$M = 0.59$, 95% CI $[0.50, 0.68]$, $t(18) = 2.11$, $p = .049$, $\mathrm{BF}_{10} = 1.44$	$M = 0.61$, 95% CI $[0.53, 0.69]$, $t(18) = 3.03$, $p = .007$, $\mathrm{BF}_{10} = 6.86$	$M = 0.53$, 95% CI $[0.42, 0.64]$, $t(15) = 0.62$, $p = .544$, $\mathrm{BF}_{10} = 0.30$
DPID	$M = 0.52$, 95% CI $[0.48, 0.56]$, $t(31) = 1.18$, $p = .245$, $\mathrm{BF}_{10} = 0.36$	$M = 0.50$, 95% CI $[0.46, 0.55]$, $t(31) = 0.17$, $p = .869$, $\mathrm{BF}_{10} = 0.19$	$M = 0.49$, 95% CI $[0.43, 0.55]$, $t(24) = -0.29$, $p = .776$, $\mathrm{BF}_{10} = 0.22$
LDB	$M = 0.50$, 95% CI $[0.40, 0.60]$, $t(15) = 0.00$, $p > .999$, $\mathrm{BF}_{10} = 0.26$	$M = 0.56$, 95% CI $[0.47, 0.65]$, $t(15) = 1.46$, $p = .164$, $\mathrm{BF}_{10} = 0.62$	$M = 0.50$, 95% CI $[0.35, 0.65]$, $t(15) = 0.00$, $p > .999$, $\mathrm{BF}_{10} = 0.26$
SCL	$M = 0.52$, 95% CI $[0.46, 0.57]$, $t(20) = 0.68$, $p = .505$, $\mathrm{BF}_{10} = 0.28$	$M = 0.49$, 95% CI $[0.41, 0.57]$, $t(20) = -0.32$, $p = .754$, $\mathrm{BF}_{10} = 0.24$	$M = 0.55$, 95% CI $[0.45, 0.66]$, $t(19) = 1.04$, $p = .312$, $\mathrm{BF}_{10} = 0.37$
TDC	$M = 0.55$, 95% CI $[0.49, 0.60]$, $t(22) = 1.82$, $p = .083$, $\mathrm{BF}_{10} = 0.89$	$M = 0.57$, 95% CI $[0.49, 0.64]$, $t(22) = 1.91$, $p = .069$, $\mathrm{BF}_{10} = 1.03$	$M = 0.51$, 95% CI $[0.39, 0.63]$, $t(21) = 0.16$, $p = .878$, $\mathrm{BF}_{10} = 0.23$
TFDP	$M = 0.57$, 95% CI $[0.48, 0.66]$, $t(17) = 1.66$, $p = .116$, $\mathrm{BF}_{10} = 0.76$	$M = 0.60$, 95% CI $[0.53, 0.68]$, $t(17) = 2.95$, $p = .009$, $\mathrm{BF}_{10} = 5.74$	$M = 0.56$, 95% CI $[0.46, 0.65]$, $t(17) = 1.29$, $p = .215$, $\mathrm{BF}_{10} = 0.50$

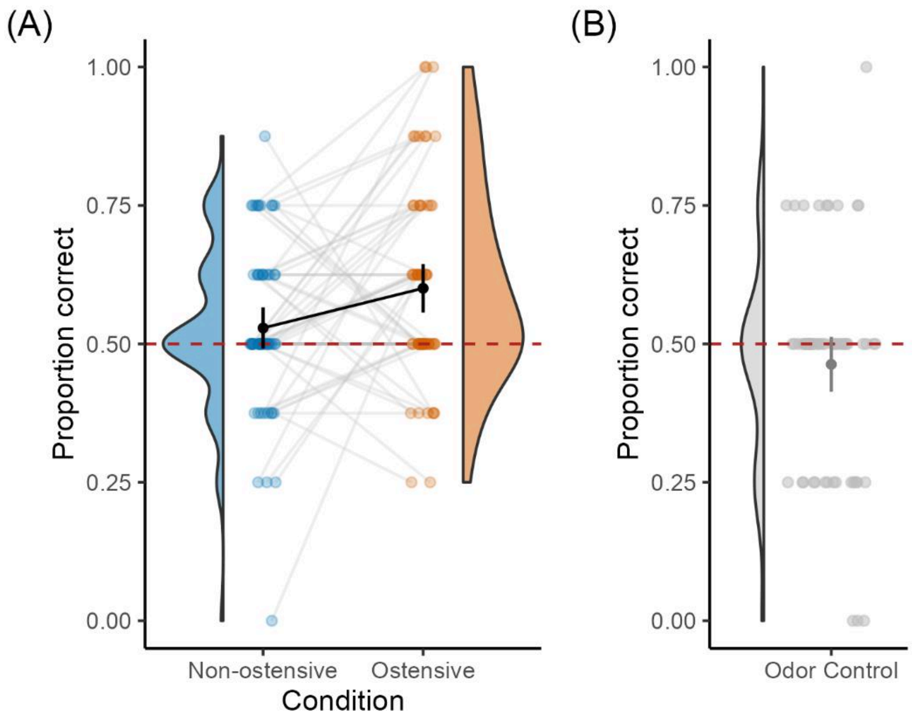
Table S4

Results of GLMM of the Dog's Choice Performance (Main Experiment: With Breed Group as a Random Effect)

effect	Estimate	SE	Lower CI	Upper CI	Chi-square	df	p	BF
(Intercept)	0.18	0.12	-0.08	0.42				
Condition	0.05	0.07	-0.09	0.18	0.52	1	0.47	0.07
Condition order	-0.13	0.07	-0.26	0.01	3.51	1	0.06	0.21
Trial number	-0.02	0.03	-0.09	0.04	0.56	1	0.45	0.03
Age	0.08	0.05	-0.01	0.19	2.61	1	0.11	0.08
Trainability score	0.00	0.05	-0.09	0.09	0.00	1	0.96	0.04
Sex:desexed	-0.04	0.19	-0.43	0.32	0.03	1	0.86	0.16

Figure S6

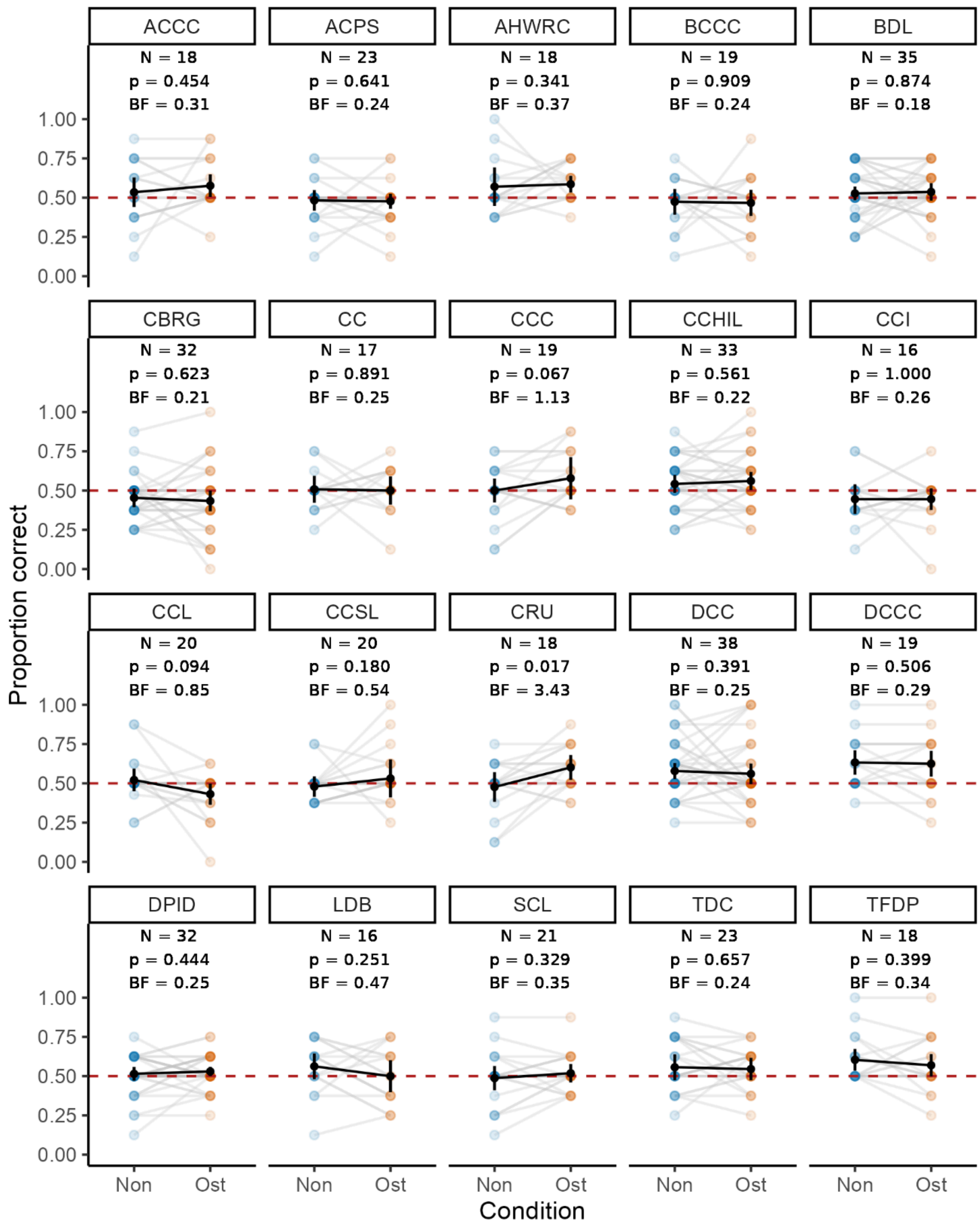
Violin and Dot Plot of Dogs' Performance (N = 61) Across the (A) Non-ostensive and Ostensive Conditions and the (B) Odor Control Condition for the Preliminary Data



Note. The red dashed lines show the chance level of 0.5. Dots represent the mean proportion correct for each individual. The grey lines connect dogs representing the same individuals. The error bars represent 95% within-subjects confidence intervals; the filled circles on the top of the error bars show the means per condition.

Figure S7

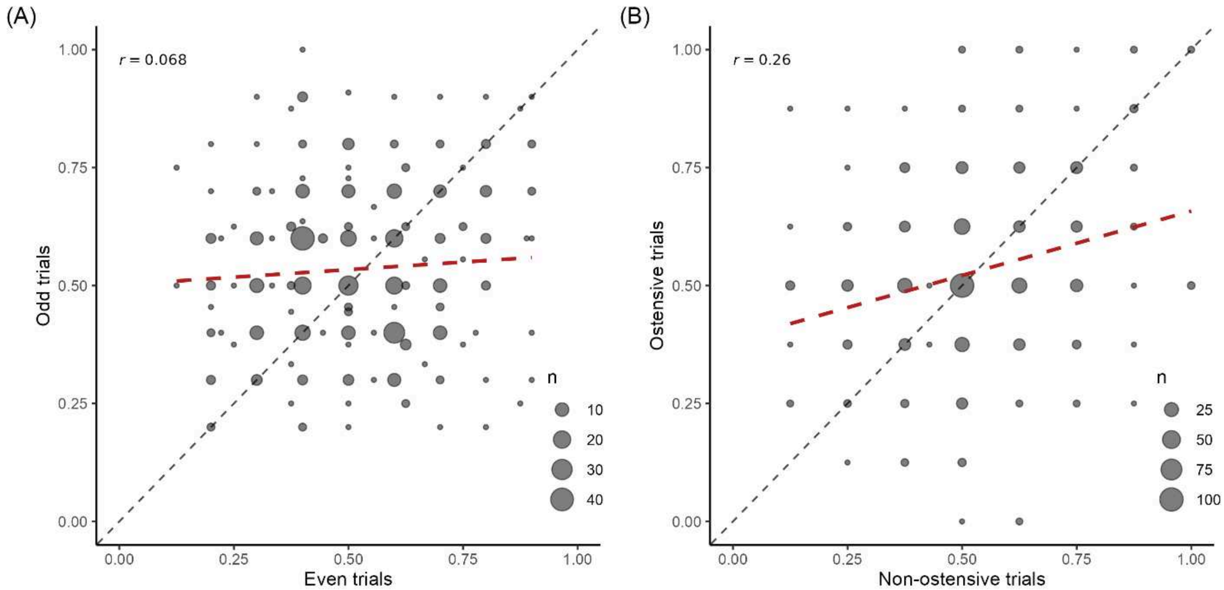
Condition Effects on Performance Per Site



Note. Dot plots of dogs' performance across the Non-ostensive and Ostensive conditions across sites. 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 means per condition.

Figure S8

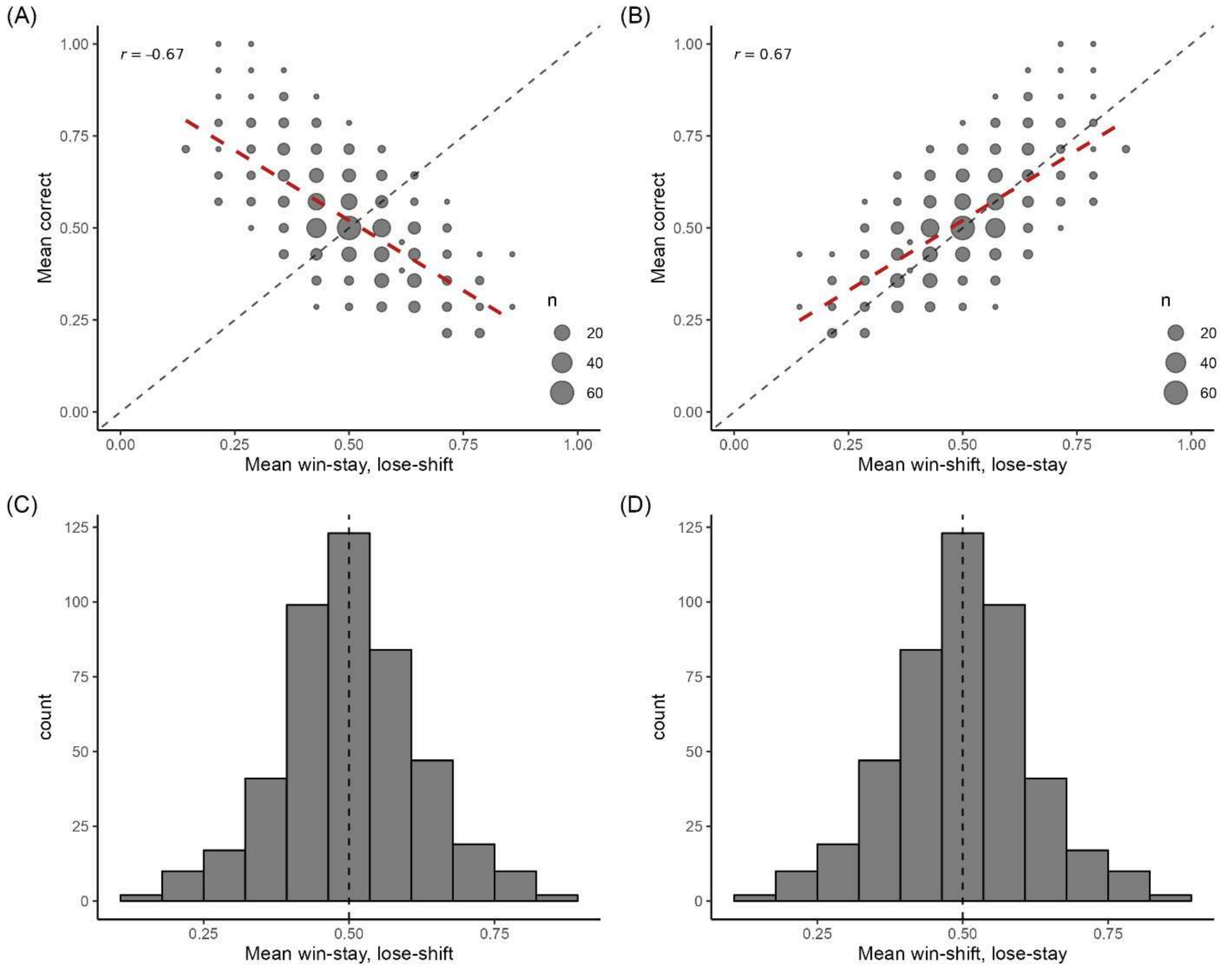
Split-Half Reliability. (A) The Dogs' Mean Performance in Odd and Even Test Trials and (B) in the Non-ostensive and Ostensive Conditions



Note. The bubbles represent the number of individuals at that performance levels; the red dashed line shows the linear regression line. The black dashed line shows the identity line.

Figure S9

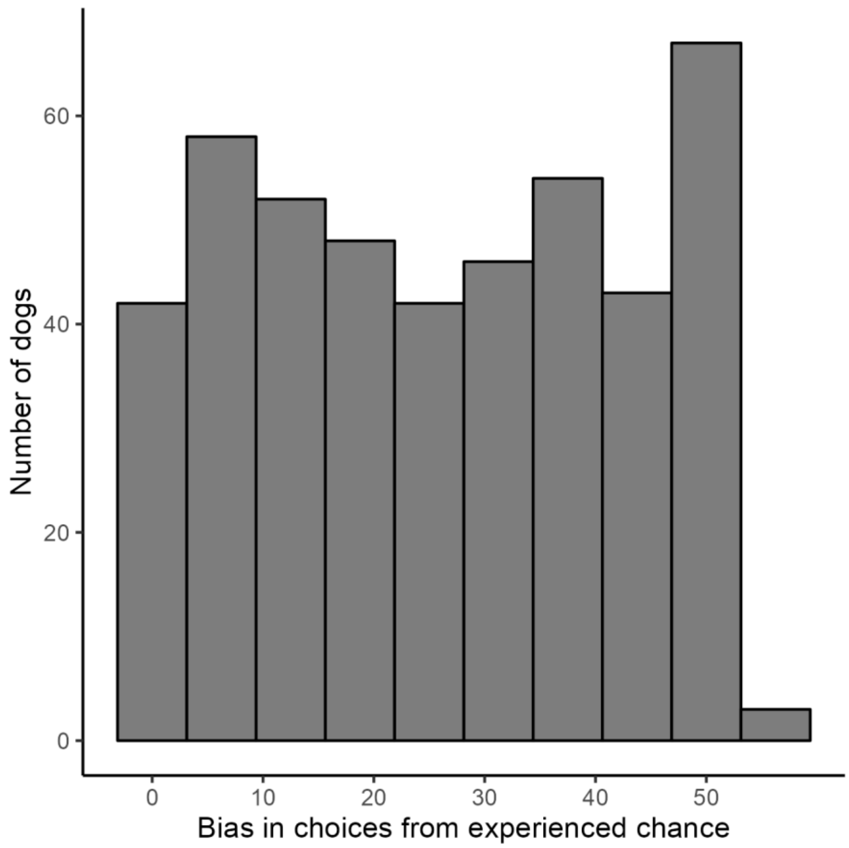
Strategy Responses. The Dogs' Choice Strategies in the Non-ostensive and Ostensive Conditions as a Function of (A) a Win-Stay, Lose-Shift Strategy or (B) a Win-Shift, Lose-Stay Strategy



Note. The bubbles represent the number of individuals at that mean performance level. The red dashed line shows the linear regression line; the black dashed line shows the identity line. (C) and (D): Histogram of dogs' distribution according to the extent to which they performed in line with the (C) win-stay, lose-shift or (D) win-shift, lose-stay strategies. The vertical line shows the chance level of 0.5.

Figure S10

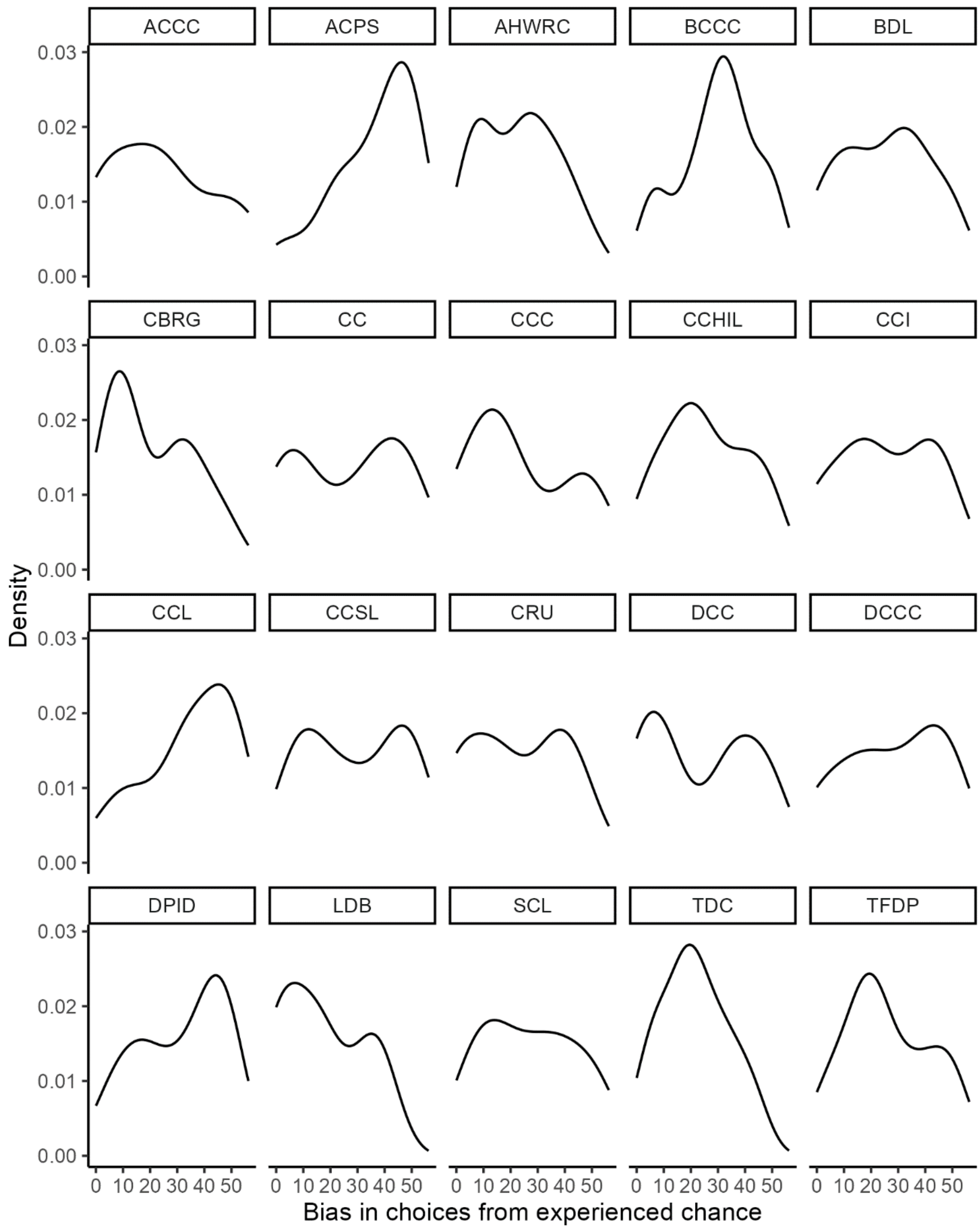
Overall Side Bias



Note. We calculated bias by taking the absolute value of the difference between each dog's percent located right and percent choice right. No side bias would be 0 and total side bias would be 50 or more.

Figure S11

Side Bias Per Site



Note. We calculated bias by taking the absolute value of the difference between each dog's percent located right and percent choice right. No side bias would be 0 and total side bias would be 50. The density function is a smoothed version of a histogram that ensures the total area under the curve is 1, which helps equate for labs with different sample sizes.