ManyDogs 1 Analyses

ManyDogs Project et al.

2023-05-17

Methods

Data analysis

We analyzed data from the project using R (Version 4.3.0; R Core Team 2023) and the R-packages *BayesFactor* (Version 0.9.12.4.4; Morey and Rouder 2022), *bayesplot* (Version 1.10.0; Gabry et al. 2019), *bayestestR* (Version 0.13.1; Makowski, Ben-Shachar, and Lüdecke 2019), *brms* (Version 2.19.0; Bürkner 2017, 2018, 2021), *broom.mixed* (Version 0.2.9.4; Bolker and Robinson 2022), *car* (Version 3.1.2; Fox and Weisberg 2019), *coda* (Version 0.19.4; Plummer et al. 2006), *data.table* (Version 1.14.8; Dowle and Srinivasan 2023), *flextable* (Version 0.9.1; Gohel and Skintzos 2023), *ggdist* (Version 3.3.0; Kay 2023a), *gghalves* (Version 0.1.4; Tiedemann 2022), *ggpubr* (Version 0.6.0; Kassambara 2023), *glue* (Version 1.6.2; Hester and Bryan 2022), *gtools* (Version 3.9.4; Bolker, Warnes, and Lumley 2022), *here* (Version 1.0.1; Müller 2020), *kableExtra* (Version 1.3.4; Zhu 2021), *knitr* (Version 1.42; Xie 2015), *lme4* (Version 1.1.33; Bates et al. 2015), *MCMCglmm* (Version 2.34; Hadfield 2010), *papaja* (Version 0.1.1; Aust and Barth 2022), *patchwork* (Version 1.1.2; Pedersen 2022), *performance* (Version 0.10.3; Lüdecke et al. 2021), *psych* (Version 2.3.3; William Revelle 2023), *tidybayes* (Version 3.0.4; Kay 2023b), 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).

Results

Pilot data

Demographics

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 (Mean = 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 choose the baited cup more in the Ostensive condition compared to the Non-ostensive condition ($X^2(1) = 5.11$, $p = 0.02$, $BF_{10} = 3.68$) (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

Demographics

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 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]).

Inter-Rater Reliability

The raters who recoded a subset of the trials had very high reliability with the original coding for choice (κ = 0.98, 95% CI [0.97, 0.98], N = 2486). Individual site reliability ranged from κ = 0.92-1.00.

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], $t(454) = 2.94$, $p = .003$, $BF_{10} = 3.7$) 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).

Condition Comparison The dogs did not choose the baited cup differently in the Ostensive and Non-ostensive conditions ($X^2(1) = 0.15$, $p = 0.70$) (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).

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	-0.06	0.31				
Condition	0.02	0.05	-0.07	0.12	0.15	1	0.70	0.04
Condition order first	-0.04	0.05	-0.14	0.06				

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

Table 1: Results of GLMM of the dogs' choice performance

effect	Estimate	SE	Lower CI	Upper CI	Chi-square	df	p	BF*
Trial number	-0.03	0.02	-0.07	0.02	1.35	1	0.25	0.03
Age	0.01	0.03	-0.05	0.06	0.13	1	0.72	0.02
Trainability score	-0.06	0.03	-0.12	0.01	2.62	1	0.11	0.09
Sex:desexed	-0.01	0.12	-0.24	0.22	0.01	1	0.94	0.09

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

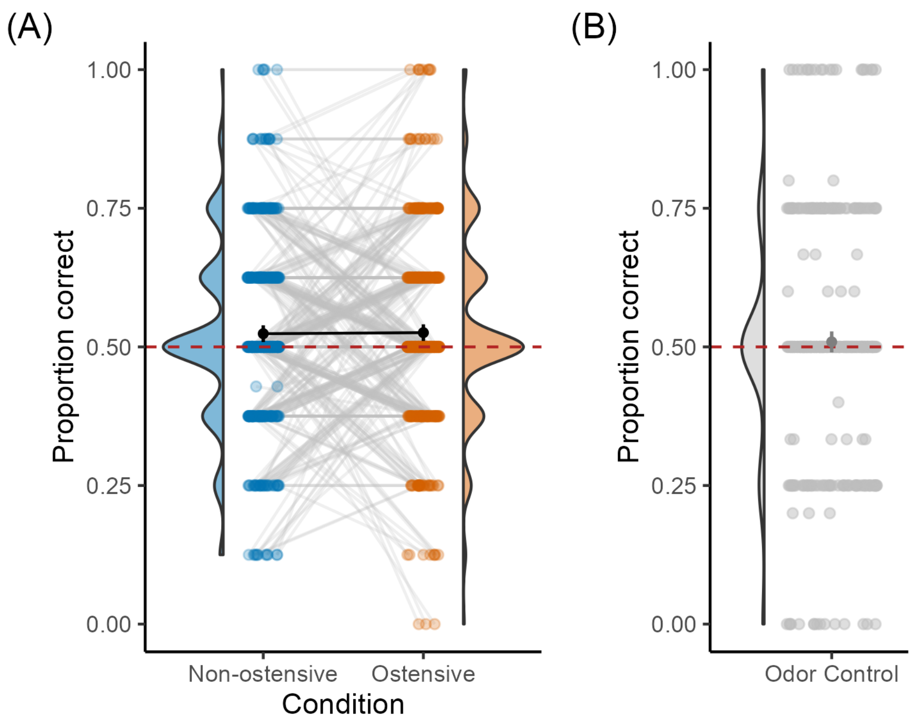

Figure 4: Violin and dot plot of dogs' performance (N = 455) across the (A) Non-ostensive and Ostensive conditions and the (B) Odor Control condition. The red dashed lines show the chance level of 0.5. Dots represent the mean proportion correct for each individual. The grey lines connect dots representing the same individuals. The error bars represent 95% within-subjects confidence intervals; the filled circles on top of the error bars show the means per condition.

Exploratory Analyses

Handler Bias One of our departures from pre-registration involved 8 of our 20 sites allowing at least some of the handlers/guardians to potentially view the cuing and baiting process of the trials. 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 or no potential for cuing. We then added this variable as a fixed effect to the GLMM investigating condition effects on responses. The dogs did not choose the baited cup differently across sites with or without the potential for cuing ($X^2(1) = 0.01$, $p = 0.92$).

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 with subject and site ID) and all possible random slope components. Condition had no effect on the dogs' choice performance ($X^2(1) = 0.50$, $p = 0.48$, $BF = 0.06$) (Figure 5). None of the control predictor variables (order of condition, trial number within condition, sex, neuter status, age, C-BARQ trainability score) had a 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.

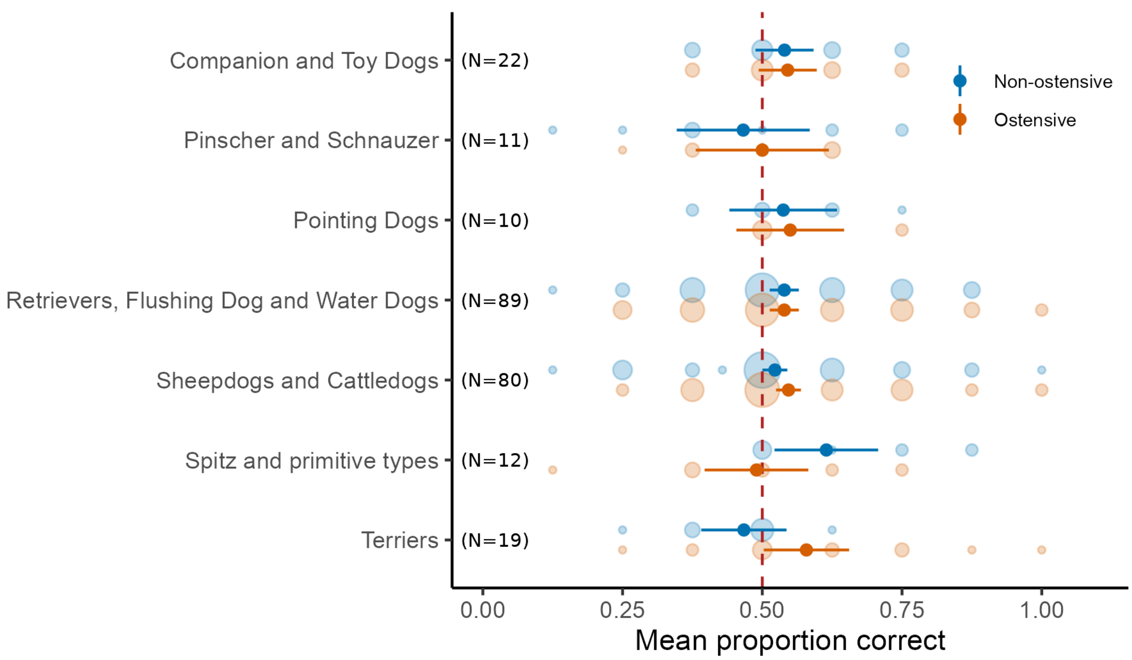

Figure 5: Plot of dogs' performance in Non-ostensive and Ostensive conditions for each breed group with N >= 8. Orange (Non-ostensive condition) and blue (Ostensive condition) bubbles represent the number of individuals at that performance level. Filled dots represent breed group means per condition, and error bars represent 95% within-subjects confidence intervals. The red dashed line shows the chance level of 0.5.

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 pure breed dogs with three more more individuals per breed (27 breeds; 0 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 the majority of cases, the posterior mode was near 0 (Non-ostensive: 0.03, 90% highest-density continuous interval [0, 0.72]; Ostensive: 0.02, 90% highest-density continuous interval [0, 0.51]), 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.

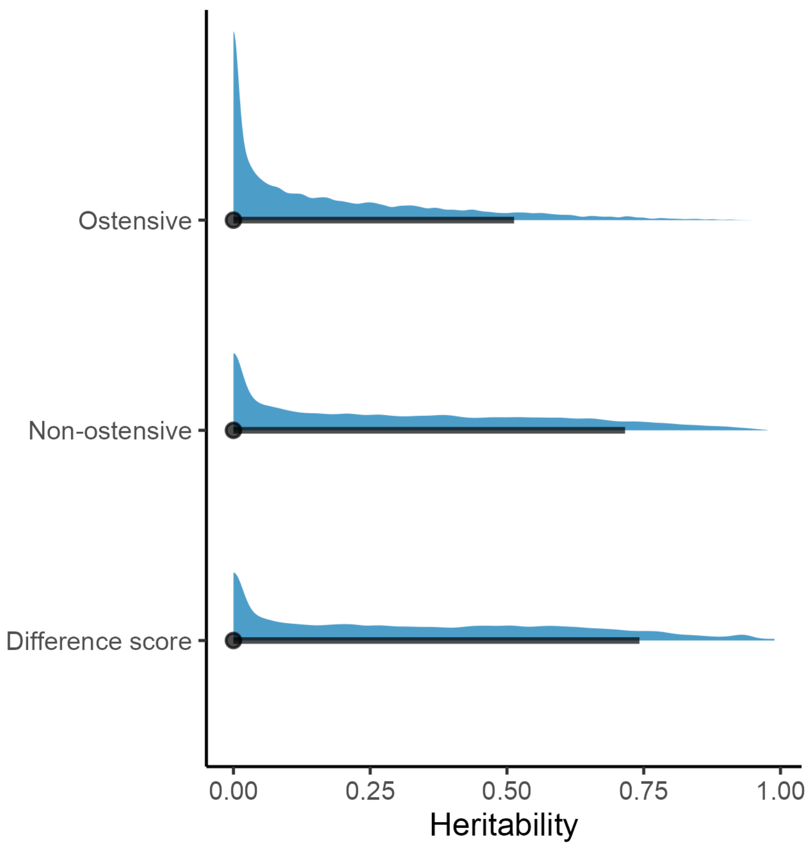

Figure 6: Posterior distributions of heritability estimates for models including dogs from breeds with three or more individuals with cognitive data. Points reflect the posterior mode and lines reflect the 90% highest continuous posterior interval for each model.

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 [-0.02, 0.16], $p = .148$, $BF_{10} = 0.31$; 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], $p < .001$, $BF_{10} = 9.9{\times}10^5$; 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], $p < .001$, $BF_{10} = 5.4\times10^{55}$; Figure S9A) and conversely a win-shift, lose-stay strategy was positively correlated with success ($r(452) = 0.67$, 95% CI [0.61, 0.71], $p < .001$, $BF_{10} = 5.4\times10^{55}$; 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], $t(453) = 0.06$, $p = .950$, $BF_{10} = 0.05$; Figure S9C) or the win-shift, lose-stay strategy (Mean = 0.50, 95% CI [0.49, 0.51], $t(453)$ = -0.06, $p = .950$, $BF_{10} = 0.05$; Figure S9D) above chance levels (0.5).

An even simpler strategy would be to simply 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 choose 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: 65.7% of dogs that received Non-ostensive first were included compared to 64.2% of dogs that received Ostensive first.

Supplementary materials

Table S1: Site information and demographics

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)	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)	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)	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)	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)	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)	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)	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)	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)	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)	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)	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)	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)
Comparative Cognitive Science Lab (CCSL)	Ostojić	Rijeka, Croatia	2	20	10	5.1±2.8 (1-10.1)	5:15	16	P = 20 G = 0	Lab	Exempt
Consultorio Comportamentale (CC)	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)	Reinholz	Warsaw, Poland	4	32	26	5.2±2.9 (0.7-12)	12:20	18	P = 32 G = 0	Lab	Exempt
Dog Cognition Centre (DCC)	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)	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)	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)	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)	Sommese	Budapest, Hungary	3	18	13	3.8±3.1 (0.7-12.6)	7:11	12	P = 18 G = 0	Lab	Exempt
Thinking Dog Center (TDC)	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.

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	-0.12	0.37				
Condition	0.29	0.13	0.04	0.56	5.11	1	0.02	3.68
Condition order first	0.06	0.13	-0.21	0.30				
Trial number	-0.10	0.07	-0.23	0.03	2.32	1	0.13	0.48
Sex	-0.08	0.14	-0.36	0.19	0.37	1	0.54	0.40
Age	-0.01	0.07	-0.15	0.12	0.03	1	0.85	0.16
Training score	0.09	0.07	-0.04	0.23	1.96	1	0.16	0.42

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

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 dogs' choice performance (main experiment: with breed group as random effect)

effect	Estimate	SE	Lower CI	Upper CI	Chi-square	df	p	BF*
(Intercept)	0.19	0.12	-0.06	0.42				
Condition	0.05	0.07	-0.07	0.18	0.50	1	0.48	0.06
Condition order first	-0.13	0.07	-0.28	0.00				
Trial number	-0.02	0.03	-0.09	0.04	0.53	1	0.47	0.03
Age	0.08	0.05	-0.03	0.19	2.17	1	0.14	0.07
Trainability score	0.00	0.05	-0.10	0.09	0.00	1	0.96	0.04
Sex:desexed	-0.04	0.19	-0.43	0.37	0.04	1	0.84	0.15

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

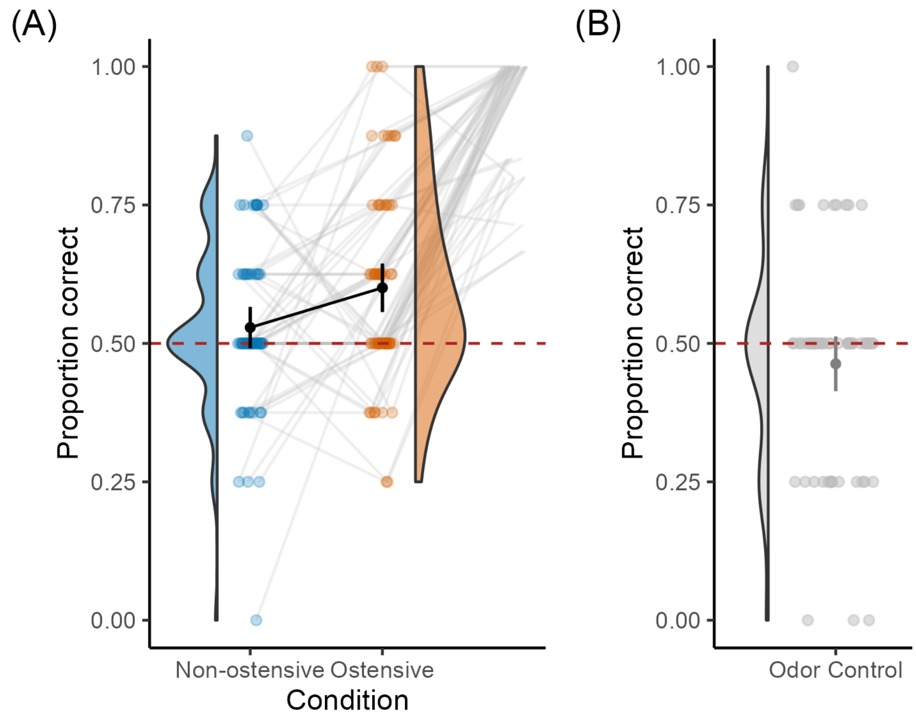

Figure S6: Violin and dot plot of dogs' performance (N = 61) across the (A) Non-ostensive and Ostensive conditions and the (B) Odor Control condition for the preliminary data. The red dashed lines show the chance level of 0.5. Dots represent the mean proportion correct for each individual. The grey lines connect dots representing the same individuals. The error bars represent 95% within-subjects confidence intervals; the filled circles on top of the error bars show the means per condition.

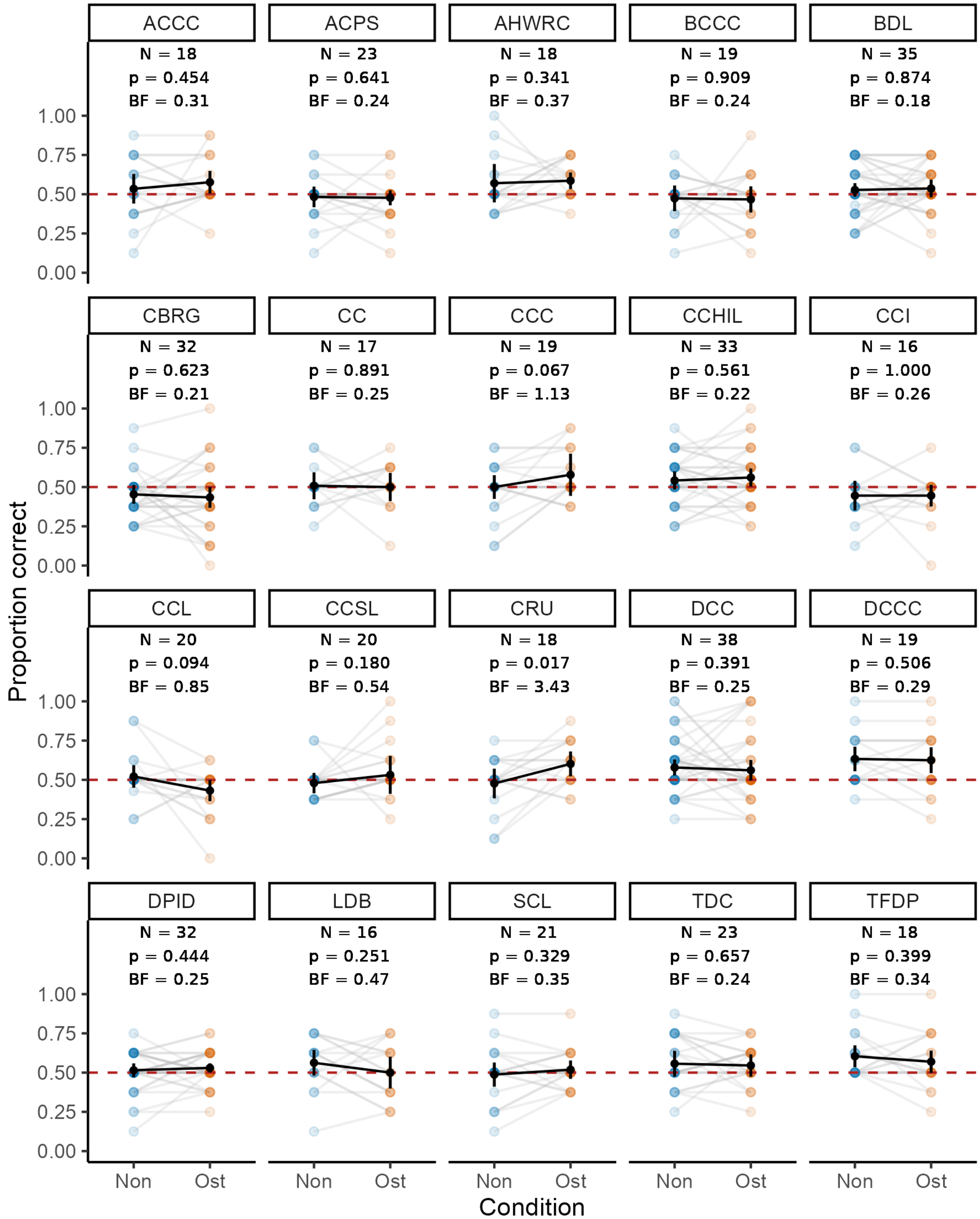

Figure S7: Condition effects on performance per site. Dot plot of dogs' performance across the Non-ostensive and Ostensive conditions across sites. Dots represent the mean proportion correct for each individual. The grey lines connect dots representing the same individuals. The error bars represent 95% within-subjects confidence intervals; the filled circles on top of the error bars show the means per condition.

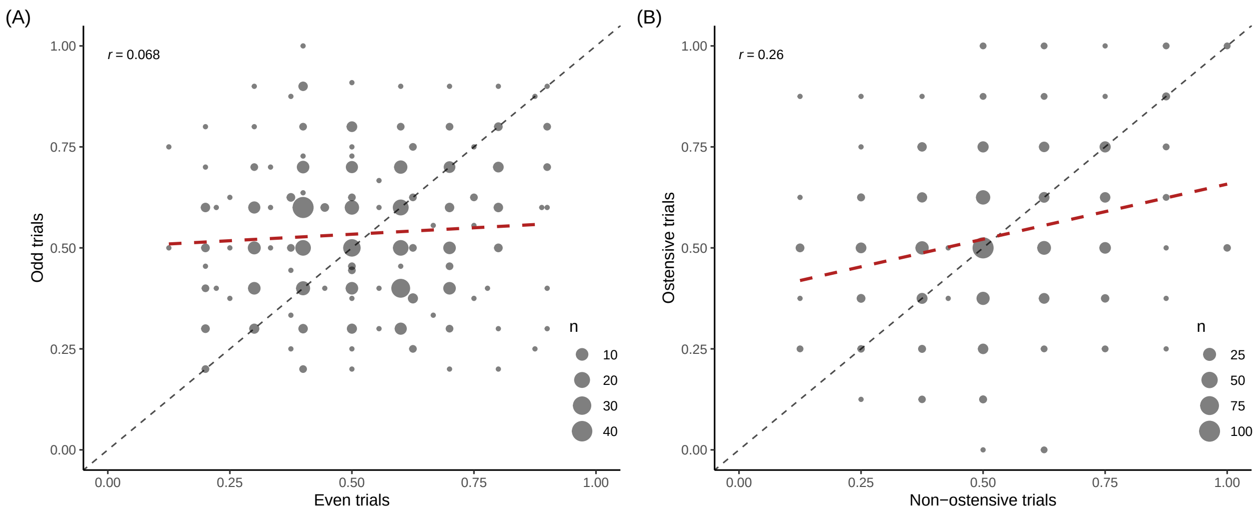

Figure S8: Split-half reliability. (A) The dogs' mean performance in odd and even test trials and (B) in the Non-ostensive and Ostensive conditions. The bubbles represent the number of individuals at that performance levels; the red dashed line shows the linear regression line. The black dashed line shows the identity line.

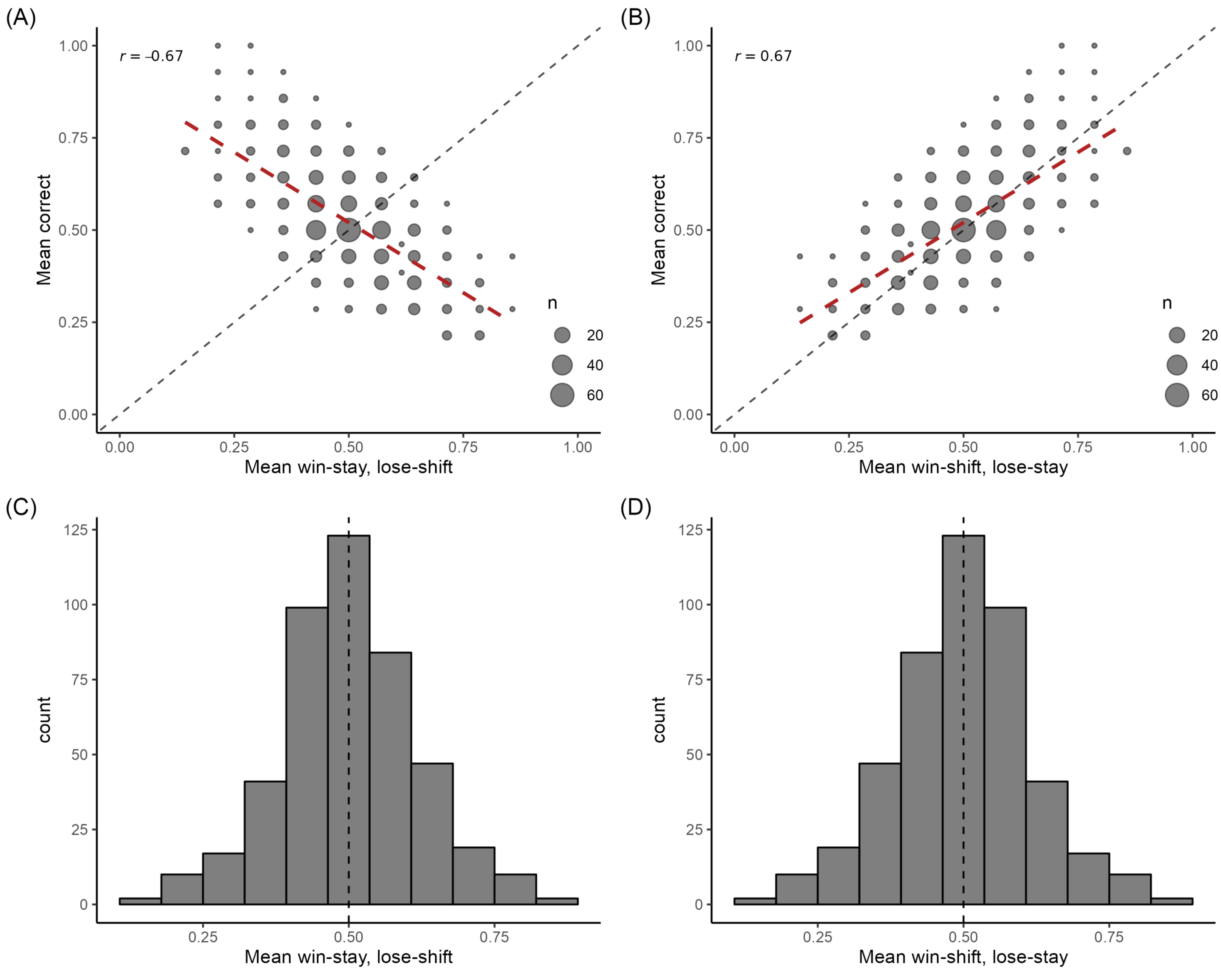

Figure S9: Strategy responses. The dogs' choice strategies in the Non-ostensive and Ostensive conditions as a function of (A) a win-stay, lose-shift strategy or (B) a win-shift, lose-stay strategy. The bubbles represent the number of individuals at that mean performance level. The red dashed line shows 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.

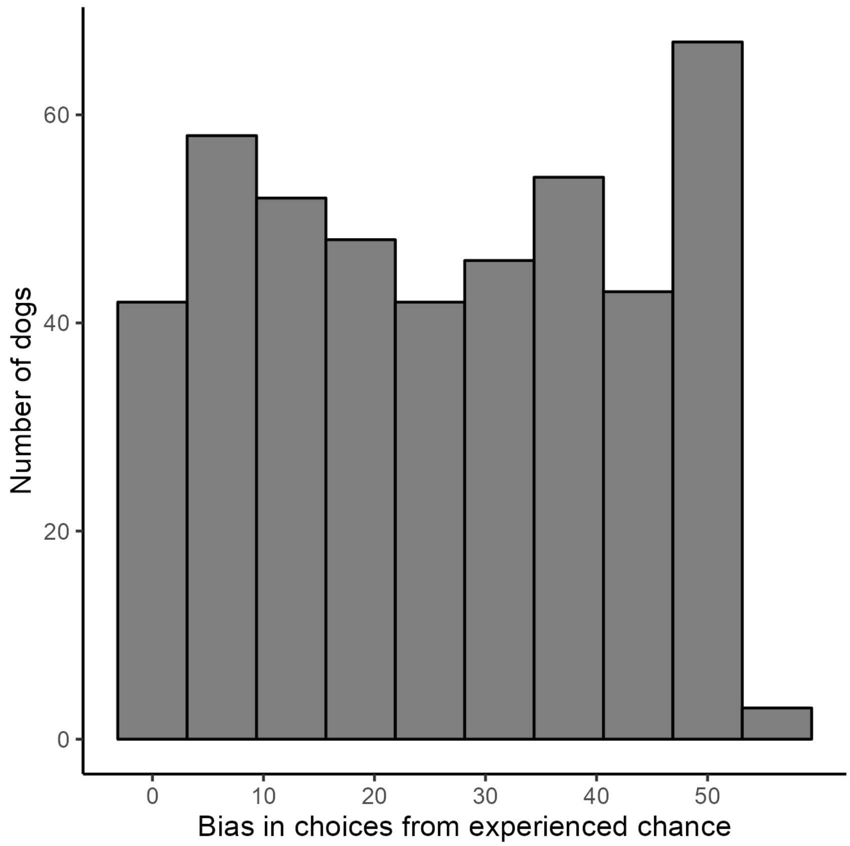

Figure S10: Overall side bias. We calculated bias by taking the absolute value of the difference between each dog's percent located right and percent choice right. No side bias would be 0 and total side bias would be 50 or more.

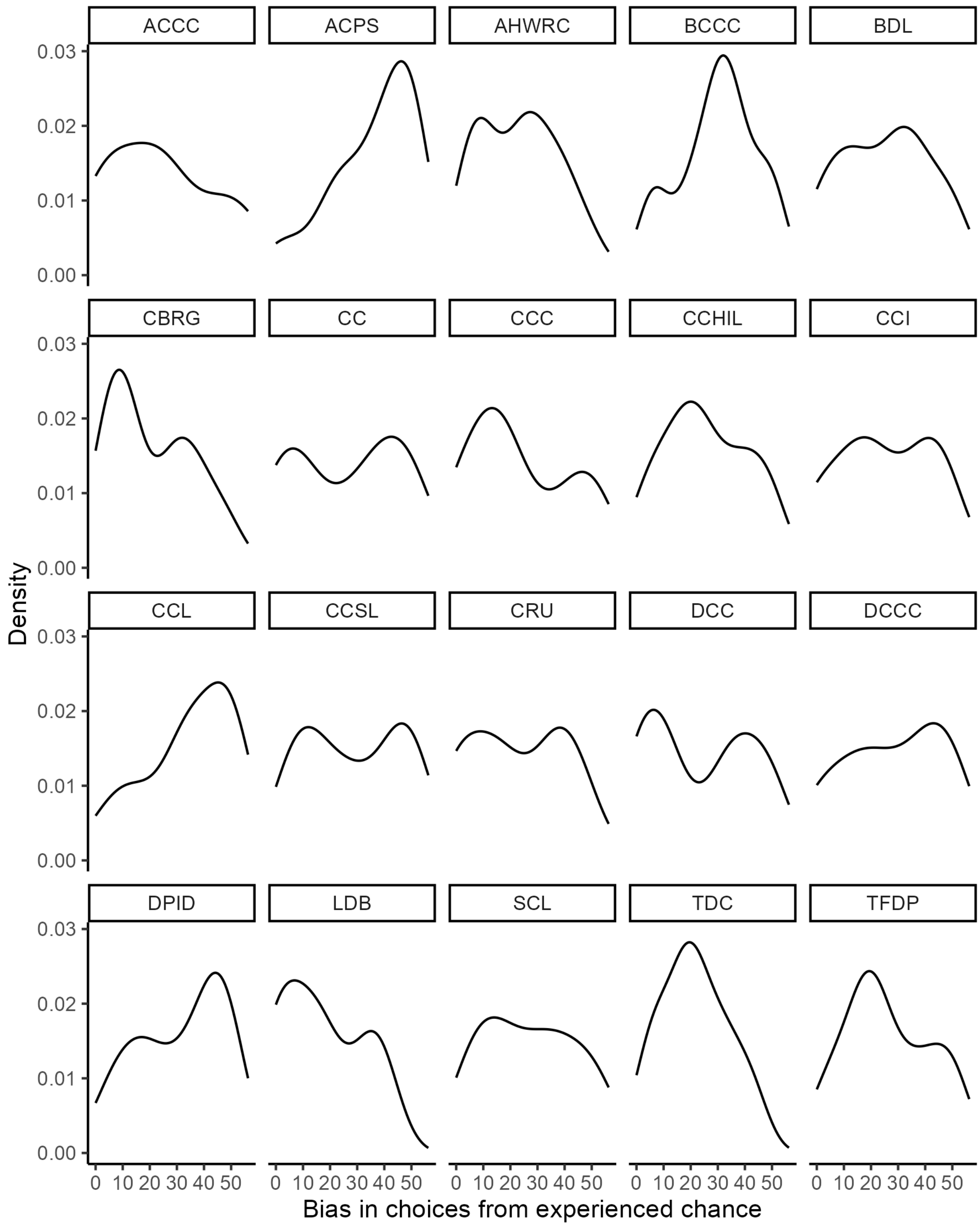

Figure S11: Side bias per site. We calculated bias by taking the absolute value of the difference between each dog's percent located right and percent choice right. No side bias would be 0 and total side bias would be 50. The density function is a smoothed version of a histogram that ensures the total area under the curve is 1, which helps equate for labs with different sample sizes.

References

Aust, Frederik, and Marius Barth. 2022. *papaja: Prepare Reproducible APA Journal Articles with R Markdown*. https://github.com/crsh/papaja.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Bolker, Ben, and David Robinson. 2022. *Broom.mixed: Tidying Methods for Mixed Models*. https://github.com/bbolker/broom.mixed.

Bolker, Ben, Gregory R. Warnes, and Thomas Lumley. 2022. *Gtools: Various r Programming Tools*. https://github.com/r-gregmisc/gtools.

Bürkner, Paul-Christian. 2017. "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.

———. 2018. "Advanced Bayesian Multilevel Modeling with the R Package brms." *The R Journal* 10 (1): 395–411. https://doi.org/10.32614/RJ-2018-017.

———. 2021. "Bayesian Item Response Modeling in R with brms and Stan." *Journal of Statistical Software* 100 (5): 1–54. https://doi.org/10.18637/jss.v100.i05.

Dowle, Matt, and Arun Srinivasan. 2023. *Data.table: Extension of 'Data.frame'*.

Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*. Third. Thousand Oaks CA: Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Gabry, Jonah, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman. 2019. "Visualization in Bayesian Workflow." *J. R. Stat. Soc. A* 182: 389–402. https://doi.org/10.1111/rssa.12378.

Gohel, David, and Panagiotis Skintzos. 2023. *Flextable: Functions for Tabular Reporting*.

Hadfield, Jarrod D. 2010. "MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package." *Journal of Statistical Software* 33 (2): 1–22. https://www.jstatsoft.org/v33/i02/.

Hester, Jim, and Jennifer Bryan. 2022. *Glue: Interpreted String Literals*.

Kassambara, Alboukadel. 2023. *Ggpubr: 'Ggplot2' Based Publication Ready Plots*. https://rpkgs.datanovia.com/ggpubr/.

Kay, Matthew. 2023a. *ggdist: Visualizations of Distributions and Uncertainty*. https://doi.org/10.5281/zenodo.3879620.

———. 2023b. *tidybayes: Tidy Data and Geoms for Bayesian Models*. https://doi.org/10.5281/zenodo.1308151.

Lüdecke, Daniel, Mattan S. Ben-Shachar, Indrajeet Patil, Philip Waggoner, and Dominique Makowski. 2021. "performance: An R Package for Assessment, Comparison and Testing of Statistical Models." *Journal of Open Source Software* 6 (60): 3139. https://doi.org/10.21105/joss.03139.

Makowski, Dominique, Mattan S. Ben-Shachar, and Daniel Lüdecke. 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.

Morey, Richard D., and Jeffrey N. Rouder. 2022. *BayesFactor: Computation of Bayes Factors for Common Designs*. https://richarddmorey.github.io/BayesFactor/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*.

Pedersen, Thomas Lin. 2022. *Patchwork: The Composer of Plots*.

Plummer, Martyn, Nicky Best, Kate Cowles, and Karen Vines. 2006. "CODA: Convergence Diagnosis and Output Analysis for MCMC." *R News* 6 (1): 7–11. https://journal.r-project.org/archive/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Tiedemann, Frederik. 2022. *Gghalves: Compose Half-Half Plots Using Your Favourite Geoms*. https://github.com/erocoar/gghalves.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

William Revelle. 2023. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University. https://CRAN.R-project.org/package=psych.

Xie, Yihui. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman;

Hall/CRC. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.*