Preliminary data

2021-05-04

Analysis

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 in an associated online repository at https://github.com/ManyDogsProject/manydogs1_vienna.git.

Reliability analysis

A second rater naive to the hypotheses and theoretical background of the study scored the data of 18 randomly selected dogs (ca. 30% of the preregistered 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).

Hypothesis tests

To evaluate whether dogs' performance 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. We calculated Bayes factors using the `ttestBF()` function from the *BayesFactor* package [@Morey.2018] in using R [@R-base].

To compare the performance between the test conditions (ostensive, non-ostensive) we fitted a Generalised Linear Mixed Model (GLMM) with binomial error distribution and logit link function using the function 'glmer' of the lme4 package [@Bates.2015]. We included the predictor variables condition, order of condition (ostensive first, non-ostensive first), trial number within condition, sex, age (in years), dogs' trainability score based on the C-BARQ questionnaire [@Hsu.2003]. Additionally, we included the random intercept of subject ID and the random slopes of condition and trial number within subject ID.

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). Confidence intervals for the predictors were derived based on 1000 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.2019). Collinearity was no issue (maximum VIF: 1.02). 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 (using a function kindly provided by Roger Mundry). This procedure revealed the model to be stable with respect to the fixed effects.

As inference criterion, we used p-values below .05 (two-tailed) for the one-sample t-tests. For the GLMM, we used likelihood ratio tests (R function `drop1()` with argument 'test' set to "Chisq", @Barr.2013) with p-values below 0.05 as criterion to make inferences about fixed effects.

Bayes factors for the models were calculated from Bayesian models using the `brm()` function from the *brms* package [@R-brms]. We used default, non-informative priors, 12,000 iterations per chain (of which 2,000 were warm-up iterations), and 4 chains for the Bayesian models. We then used the `bayes_factor()` function to compare models, with 10 repetitions using bridge sampling [@R-bridge_sampler]. For the GLMM, Bayes factors represent the evidence for the full model over the full model without the fixed effect under investigation.

Results

One-sample t-test to compare against chance level

The dogs (N = 61) performed significantly better than expected by chance in the ostensive condition ($M = 0.60$, 95% CI $[0.55, 0.65]$, $t(60) = 4.41$, $p < .001$, $BF_{10} = 459.91$) but not in the non-ostensive condition ($M = 0.53$, 95% CI $[0.49, 0.57]$, $t(60) = 1.47$, $p = .146$, $BF_{10} = 0.39$) or the odor control condition ($M = 0.46$, 95% CI $[0.41, 0.51]$, $t(60) = -1.45$, $p = .151$, $BF_{10} = 0.38$) (Figure 1).
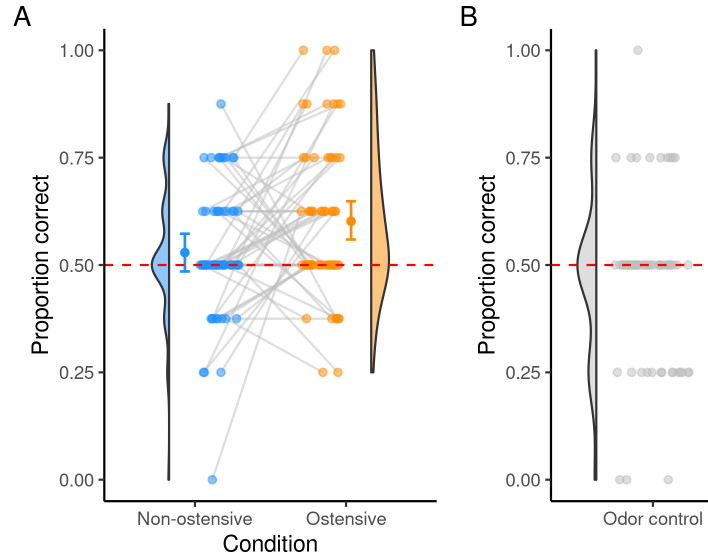

Figure 1: Violin and dot plot of dogs' performance (N=61) across the non-ostensive and ostensive test conditions (A) and the odor control condition (B) of Experiment 1. The red dashed lines show the chance level of 0.5. Dots represent the mean proportion correct for each individual. The grey lines connect dots representing the same individuals. The error bars represent 95% confidence intervals based on a parametric bootstrap; the filled circles on top of the error bars show the fitted model.

Generalised linear mixed model

The dogs were significantly more likely to choose the baited cup in the ostensive condition compared to the non-ostensive condition ($\chi^2(1) = 5.11$, p = .024) (Figure 1A). 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

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

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