MD1 - CDL data analysis

February 21, 2021

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

To evaluate whether the dogs' performance deviated significantly from the hypothetical chance level of 0.5 in the ostensive, non-ostensive, and odor control condition we first aggregated the data across trials for each individual and condition. We then conducted 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].

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), the 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*). In order to check for collinearity issues, 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 the 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 smaller than .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], $t(60) = 4.41$, $p < .001$, $\text{BF}_{10} = 459.91$) but not in the non-ostensive condition ($M = 0.53$, 95% CI [0.49, 0.57], $t(60) = 1.47$, $p = .146$, $\text{BF}_{01} = 2.56$) or the odor control condition ($M = 0.46$, 95% CI [0.41, 0.51], $t(60) = -1.45$, $p = .151$, $\text{BF}_{01} = 2.63$) (Figure 1).

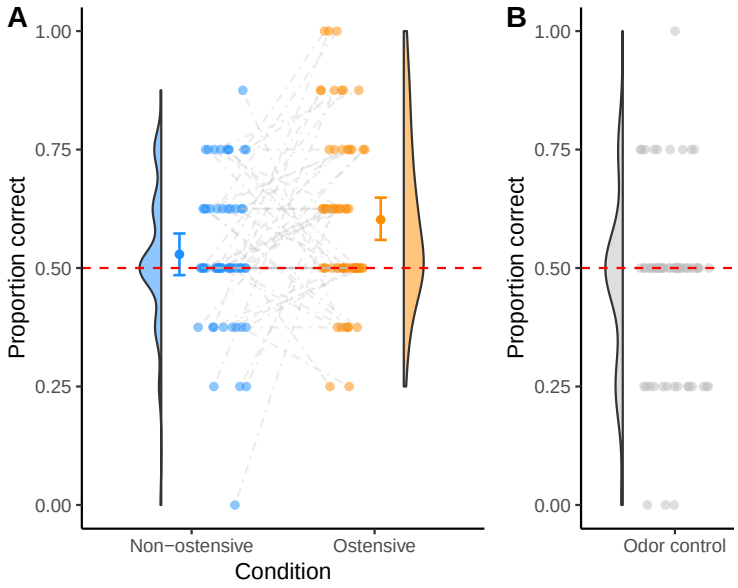

Figure 1: Violin and dot plot of dogs' performance (N=61) across the non-ostensive and ostensive test conditions (A) and the odor control condition (B) of Experiment 1. The red dashed lines show the hypothetical chance level of 0.5. The transparent dots represent the mean proportion correct for each individual. The grey lines connect dots representing the same individuals. The error bars are 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). The control predictors order of condition, trial number (within condition), sex, age, or the C-BARQ trainability score had no significant 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 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.

Study 2

Genetic analysis of among-breed heritability

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

Breed average genetic similarity will be represented by an identity-by-state (IBS) matrix calculated from publicly available genetic data collected using the Illumina CanineHD bead array [@Parker.et.al.2017]. The proportion of single-nucleotide polymorphisms (SNPs) identical by state between pairs of individual dogs will be calculated using PLINK [@Purcell.Chang.2018]. These values will then be averaged for every pair of breeds in order to generate a breed-average IBS matrix. This breed-average IBS matrix will be extrapolated to an individual-level IBS matrix for purposes of our analysis. For individuals of different breeds, the IBS value will be set to the average similarity between those breeds in the genetic dataset. For individuals of the same breed, the IBS value will be set to the average IBS value among members of that breed in the genetic dataset. This purpose of this approach is to simultaneously incorporate a measure of between- and within-breed genetic similarity, retaining the ability to model phenotypes at the individual, rather than breed-average level.

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