### Scoring and analysis

We scored whether chimpanzees stabilized both trays (coded as 1) or only one tray (coded as 0). Following the preregistered analysis plan, we fitted a generalized linear mixed model (GLMM) with binomial error structure and logit link function [1] to investigate whether the chimpanzees were more likely to stabilize both trays in the test condition than in the control condition. We included as fixed effects condition, trial number (within condition, 1-16), and the order of conditions (test-first, control first). Additionally, we included subject ID as random intercept and condition as random slope within subject ID (the random slope of trial number was removed due to convergence issues following our preregistered contingency plans).

We used likelihood ratio tests (R function drop1 with argument 'test' set to "Chisq") to make inferences about fixed effects. The models were fitted in R (version 4.03 [2]) using the function glmer of the R package lme4 [3]. The 95% confidence intervals were calculated using the function bootMer of the R package lme4, using 1,000 parametric bootstraps and bootstrapping over the random effects.

We z-transformed the covariate trial number (to a mean of zero and a standard deviation of one). We determined variance inflation factors [4] for a standard linear model excluding the random effects using the R package car [5]. Collinearity was no issue (maximum variance inflation factor: 1.0). We also assessed model stability by removing each level of the random effect subject ID one at the time and refitting the model. This procedure revealed the model to be stable with respect to the fixed effects.

Additionally, we explored whether the chimpanzees were obtaining both food items more often than expected by chance in trials in which they only stabilized one platform. Therefore, we fitted, separately for the control and test conditions, intercept-only binomial GLMMs only including trials in which the chimpanzees stabilized only one tray. We used the Wald test of the intercept to determine whether the chimpanzees’ performance deviated significantly from the chance level of 0.5.

The data and R scripts associated with this manuscript are available on GitHub: https://github.com/cvoelter/dog\_unwilling\_unable

### Results

The chimpanzees were significantly more likely to stabilize both trays in the test condition than the control condition (χ² = 14.97, df = 1, *p* < 0.001; for the model coefficients, see Table S1). Trial number within condition or the order of conditions had no significant effect on the chimpanzees’ performance.

#### Table S1. Results of a binomial GLMM with both-trays-stabilized as response variable.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Estimate | SE | 95% CI | | χ² | df | p |
| (Intercept) | -0.54 | 0.35 | -1.31 | 0.15 |  |  |  |
| Condition1 | 1.36 | 0.28 | 0.84 | 2.1 | 14.97 | 1 | <0.001 |
| Trial number3 | -0.07 | 0.11 | -0.29 | 0.15 | 0.48 | 1 | 0.488 |
| Order of condition2 | 0.92 | 0.49 | -0.04 | 1.96 | 3.17 | 1 | 0.075 |

Notes: Reference categories: 1control; 2control-first. Covariate trial number was z-transformed to a mean of zero and a standard deviation of one; mean (sd) of the original variable was 38.5 (4.6).

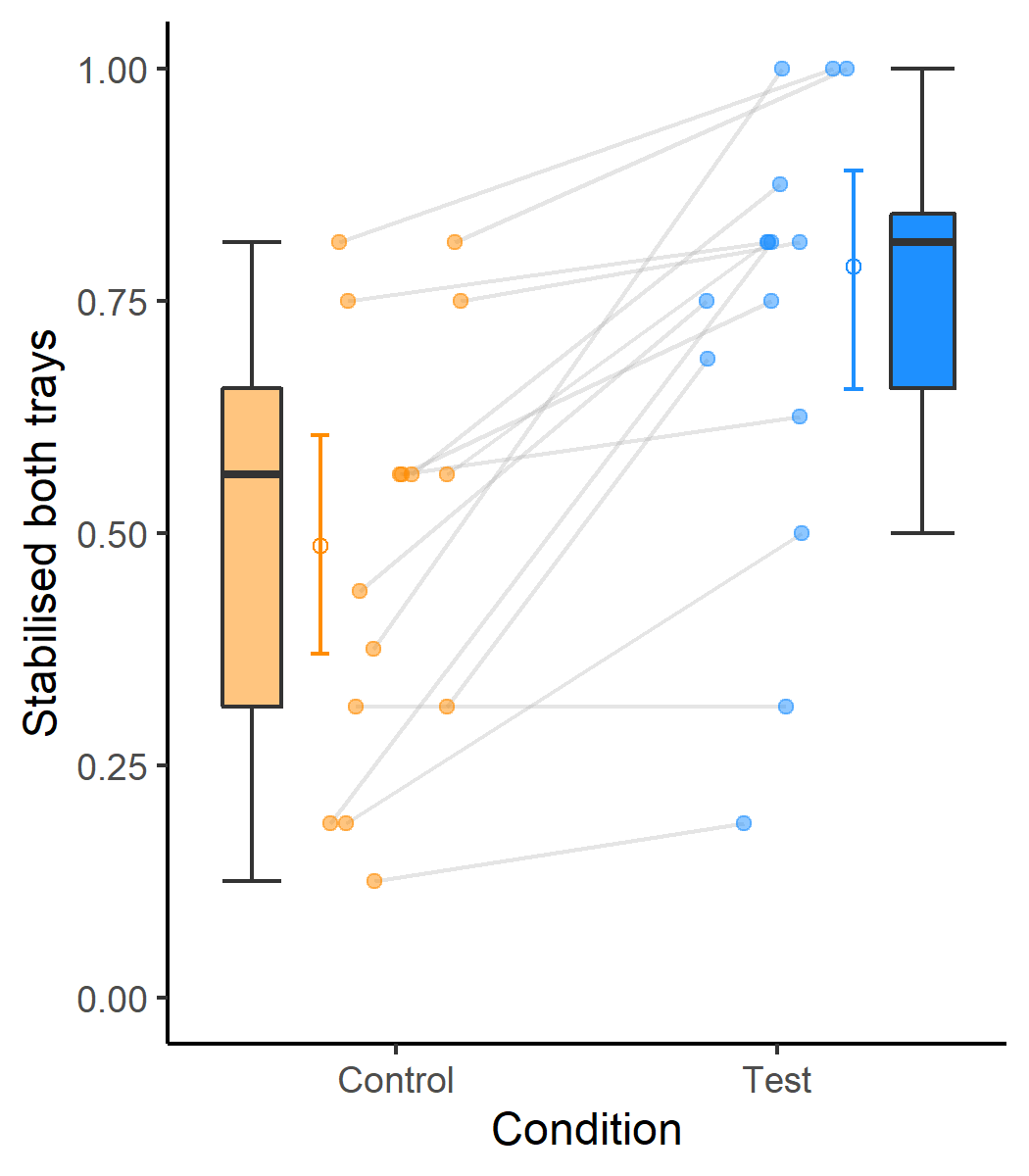


Figure 1 Box and dot plot showing the proportion of trials in which the chimpanzees stabilized both platforms across the two conditions. The dots represent individual mean values; the lines connect values of the same individuals. The error bars represent bootstrapped 95% confidence intervals; the open circles show the fitted values.

When the chimpanzees stabilized only one platform, they were significantly more likely than expected by chance to obtain both food items in the control condition (Mean ± SE: 0.86 ± 0.04; z = 4.51, p < 0.001) but not in the test condition (Mean ± SE: 0.47 ± 0.09; z = -0.85, p = 0.395).

References

1. McCullagh P, Nelder JA. 1989 *Generalized Linear Models, Second Edition*.

2. R Core Team. 2020 *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. See https://www.R-project.org/.

3. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, Dai B, Eigen C. 2016 Package ‘lme4’. *Comprehensive R Archive Network (CRAN)*.

4. Field A. 2005 Discovering Statistics Using SPSS, (Second Edition). *Journal of Advanced Nursing*. (doi:10.1111/j.1365-2648.2007.04270\_1.x)

5. Fox J, Weisberg S. 2011 An R Companion to Applied Regression, second edition. *Sage, Thousand Oaks CA*.