Supplementary Information for

Elephants have a nose for quantity

Authors: Joshua M. Plotnik[a,b,c,d,1], Daniel L. Brubaker[c], Rachel Dale[c], Lydia N. Tiller[c], Hannah S. Mumby[e,f] and Nicola S. Clayton[b]

Affiliations:

a) Dept. of Psychology, Hunter College, City University of New York, New York, NY, USA
b) Dept. of Psychology, University of Cambridge, Cambridge, U.K.
c) Think Elephants International, New York, NY, USA
d) Golden Triangle Asian Elephant Foundation, Chiang Saen, Thailand
e) Dept. of Zoology, University of Cambridge, Cambridge, U.K.
f) College for Life Sciences, Wissenschaftskolleg zu Berlin, Berlin, Germany

[1]Corresponding Author: Joshua.Plotnik@gmail.com, phone: 1-212-396-6442

This PDF file includes:

Supplementary Materials and Methods
Statistical Analyses

Tables S1 to S4
Captions for Movies S1 to S2

Other supplementary materials for this manuscript include the following:

Movies S1 to S2

www.pnas.org/cgi/doi/10.1073/pnas.1818284116

Supplementary Materials and Methods

Statistical Analyses. To analyze the elephants' success in choosing the higher quantity by both ratio and disparity, we constructed two logistic regression mixed models. The outcome of each trial (success or failure in selecting the bucket with more food) was a binary variable and therefore analyses were based on logistic regressions, which model likelihood of success using the function *glmer* in the *lme4* package in R (Ref. S1). First, we tested for success by the magnitude of the difference in food quantities. We constructed a full model of success by quantity of food (transformed to a linear term by dividing the lower by the higher value, i.e. 1:2=0.5) as a continuous linear term and a continuous quadratic term. We tested the significance of covariates by removing them from the full model and testing for significance using a likelihood ratio test of the full model against the model not including the term. We tested for the following effects in the full model: sex (binary factor), age (continuous, mean centered), trial number (continuous variable from 1 to 14 to test for learning effect), and whether the trial was a control (factor with two levels: metal bucket trial and double-blind trial). We also included a random term allowing for variation in intercept by ID to account for individual variation. For any individual factor term that was significant as a covariate, we then tested for an interaction between the term and the linear quantity variable. A subset of the experiments was replicated using single and double or triple versions of the same ratios (n=252). Within this subset, we again implemented a full model using the same covariates as above and tested for significance.

We then tested for whether the absolute difference between the two food quantities was associated with success, not taking into account the quantities of food themselves; for example, both 1:2 and 5:6 have a disparity of 1 unit. The full model was of success by disparity (linear term; 1-5). We tested the following terms by removing them from the full model and testing for

significance using a likelihood ratio test of the full model against the model not including the term: sex (binary factor), age (continuous, mean centered), trial number (continuous variable from 1 to 14 to test for learning effect), and whether the trial was a control (factor with two levels: metal bucket trial and double-blind trial). We also included a random term allowing for variation in intercept by ID to account for individual variation. For any individual factor term that was significant as a covariate, we then tested for an interaction between the term and the linear disparity variable.

Non-significant terms in the logistic regression mixed models: Overall success by ratio. None of the covariate terms significantly improved the model fit based on likelihood ratio tests with and without each term included. The term for quadratic food quantity (LRT χ^2=0.14, df=1, p=0.70) was non-significant, indicating a linear association between the overall ratio decimal value and success. Trial number was not significant (LRT χ^2=1.52, df=1, p=0.22), indicating there was no learning effect across trials. Age of elephant (LRT χ^2=0.42, df=1, p=0.52) and whether the trial was a control (LRT χ^2=0.29, df=2, p=0.87) were also non-significant, showing they did not affect the association. The term for subject sex did not significantly improve the fit of the full model (LRT χ^2=3.12, df=1, p=0.08).

Non-significant terms in the logistic regression mixed models: Association between success and disparity. As with the ratio model, none of the covariate terms significantly improved the model fit based on likelihood ratio tests with and without each term included. Trial number was not significant (LRT χ^2=1.65, df=1, p=0.20), indicating there was no learning effect across trials. Age of elephant (LRT χ^2=0.43, df=1, p=0.51) and whether the trial was a control (LRT χ^2=0.33, df=2, p=0.85) were also non-significant, showing they did not affect the

association. The term for subject sex did not significantly improve the fit of the full model (LRT χ^2=3.01, df=1, p=0.08).

Supplementary Tables S1 – S4 (continued on next page)

Condition		**Experimental + metal bucket + double-blind**											**res. odor**	**solid lid**
Ratio	Sex	1:6	1:5	1:4	1:3	2:5	1:2	3:5	2:3	3:4	4:5	5:6	3.5:3.5	0:6
Bleum	F	11/14	10/14	8/14	6/14	12/14	5/14	11/14	9/14	10/14	7/14	5/14	15/24	13/24
Lanna	F	12/14	12/14	8/14	11/14	11/14	8/14	8/14	5/14	9/14	8/14	9/14	12/24	12/24
Ploy	F	13/14	13/14	12/14	9/14	13/14	9/14	9/14	11/14	10/14	9/14	9/14	12/24	10/24
Poonlarb	F	9/14	11/14	6/14	8/14	12/14	11/14	8/14	7/14	4/14	8/14	7/14	10/24	13/24
Pepsi	M	13/14	9/14	12/14	13/14	9/14	10/14	8/14	11/14	11/14	7/14	8/14	14/24	10/24
Phuki	M	14/14	11/14	11/14	12/14	12/14	11/14	12/14	10/14	12/14	11/14	10/14	12/24	15/24
Mean		12	11	9.5	9.83	11.5	9.00	9.33	8.83	9.33	8.33	8	12.5	12.2
Wilcoxon		**21,6***	**21,6***	17,6	**18,6***	**21,6***	16,6	**21,6***	12,5	15,6	##	9,5	**19,6****	**21,6****
P-value		**0.016**	**0.016**	0.063	**0.047**	**0.016**	0.063	**0.016**	0.094	0.078	##	0.156	**0.031**	**0.016**

Table S1. Success at each ratio for each elephant across conditions. Ratios are presented in order of increasing value, from left to right. First, raw data are given for each ratio by elephant out of 14 total trials across the experimental (10 sets, one trial of each ratio per set), 'metal bucket' (2 sets, one trial of each ratio per set) and 'double-blind' conditions (2 sets, one trial of each ratio per set). For the 'residual odor' condition, the elephants participated in 2 sets of 12 trials each of 14 g vs. 14 g (i.e., the ratio 3.5:3.5, which represents the halfway point between ratio numbers 1 and 6). For the 'solid lid' control condition, they participated in 2 sets of 12 trials each of the ratio 0:6. Bold numbers in the experimental + metal bucket + double-blind ('e+m+db') conditions indicate that elephants performed significantly better than chance at that given ratio at $P<0.05$ (*Wilcoxon one-sample test, one-tailed; ##the N excluding ties was too small for an analysis). The elephants performed significantly better in the 'e+m+db' conditions at 1:6 than they did in the solid lid control, and significantly better as a group when comparing their mean performance across all ratios in the 'e+m+db' conditions to their performance in the 'residual odor' condition (**Wilcoxon paired-samples test, one-tailed). The two numbers (x,y) for the Wilcoxon test indicate the W value and the N excluding ties, respectively (Ref. S2). Exact P-values are provided for each test, and one-tailed tests were used due to the 'a priori' predictions that elephants would use smell alone to select larger quantities, and that their performance would significantly exceed chance and/or performance on control trials in which olfactory information was regulated or occluded.

	Bleum	Lanna	Pepsi	Phuki	Ploy	Poonlarb
Age	13	24	12	45	13	27
Sex	F	F	M	M	F	F
Raw Count	94/154	101/154	111/154	126/154	116/154	91/154
P-Value	0.008	0.0001	<0.0001	<0.0001	<0.0001	0.029

Table S2. Demographics and summary of results for each elephant. Name, age, sex and total number of trials in which each of the six elephants selected the bucket with more food over 154 total trials (experimental + metal bucket + double-blind conditions). P-values are for binomial tests performed on the raw count.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.314	0.467	2.815	0.005
Ratio value	-2.166	1.902	-1.139	0.255
Ratio value^2	0.700	1.845	0.379	0.705
Sex (male compared to female)	0.534	0.264	2.020	0.043
Age (mean centered)	0.083	0.127	0.652	0.514
Trial number	0.037	0.030	1.233	0.218
Double-blind condition	-0.160	0.323	-0.495	0.621
Metal bucket control	-0.125	0.280	-0.445	0.656
Random effect		Variance	Std. Dev.	
Elephant ID	(Intercept)	0.047	0.216	

Table S3. Effect sizes and significance of terms in the full logistic regression mixed model for association between success and ratio in food quantities. Estimates are expressed in logits.

Fixed effects	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.271	0.258	-1.049	0.294
Disparity	0.296	0.058	5.106	<0.001
Sex (male compared to female)	0.536	0.267	2.006	0.045
Age (mean centered)	0.085	0.128	0.659	0.510
Trial number	0.038	0.030	1.282	0.200
Double-blind condition	-0.172	0.324	-0.531	0.596
Metal bucket control	-0.134	0.281	-0.476	0.634
Random effect		Variance	Std. Dev.	
Elephant ID	(Intercept)	0.048	0.220	

Table S4. Effect sizes and significance of terms in the full logistic regression mixed model for association between success and disparity in food quantities. Estimates are expressed in logits.

Movie S1. A bull elephant, Phuki, participates in a magnitude discrimination experimental condition trial with two experimenters (co-authors R.D. and L.N.T.). After the elephant has an opportunity to investigate both buckets, the experimenters withdraw the table, unlock the buckets, and re-present the table so that Phuki can make a single choice.

Movie S2. Similar to Movie S1, a female elephant, Lanna, participates in a magnitude discrimination experimental condition trial with two experimenters (co-authors R.D. and L.N.T.). In this clip, Lanna, unlike Phuki, only investigates each bucket once before making a choice.

SI References

S1. Bates D, Maechler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67:1–48.

S2. Mosteller F, Rourke REK (1973) *Sturdy Statistics: Nonparametrics and Order Statistics* (Addison-Wesley, Reading, MA).