Felix's experiment

December 6, 2023

1 Methods

To estimate the effect of condition (solo, human-computer dyad, human-human dyad), random dot pattern coherence, stimulus duration, stimulus number, block number, and day number on the performance of participants we fitted a series if six models, three of them with data from the solo condition and the human-computer dyad condition, and three of them with data from the human-human dyad condition. The reason for splitting the data was that we needed to control for the identity of two humans in the human-human dyad condition but only for the identity of one human in the other conditions. In the following we first describe the analysis of the data from the solo and the human-computer dyad condition, and then the analysis of the data from the human-human dyad condition.

1.1 Solo and human-computer dyad condition; model 1

We fitted three Generalized Linear Mixed Models (GLMM; Baayen, 2008) which differed in their response variable and the size of the data set analyzed but had identical fixed effects structures and largely identical random effects structures. We fitted one model for each of the probability of a target hit (model 1a), joystick eccentricity (model 1b), and joystick accuracy (model 1c) as the response. All three aimed at estimating the extent to which the respective response variable was affected by the fixed effects of condition (solo or human-computer dyad), random dot pattern coherence, stimulus duration, stimulus number, block number, and day number. As we hypothesized that the effect of coherence could depend on the condition we also included the interaction between these two predictors into the fixed affects part of the model. To avoid pseudo-replication and account for the possibility that the response was influenced by several layers of non-independence we included three random intercepts effects, namely those of the ID of the participant, the ID of test day (nested in participant; thereafter 'day ID'), and the ID of the block (nested in participant and day; thereafter 'block ID'). The reason for including the latter two was that it could be reasonably assumed that the performance of participants varied between test days and also between blocks tested on the same day. To avoid an 'overconfident model' and keep type I error rate at the nominal level of 0.05 we included all theoretically random slopes (Schielzeth and Forstmeier, 2009; Barr et al., 2013). These were those of condition, coherence, their interaction, stimulus duration, stimulus number, block number, and day number within participant, coherence, stimulus duration, stimulus number, and block number within day ID, and finally coherence, stimulus duration, and stimulus number within block ID. Originally we also included estimates of the correlations among random intercepts and slopes into each model, but do to convergence and identifiability problems (recognizable by absolute correlation parameters being close to 1; Matuschek et al., 2017) we had to exclude all or several of these estimates from the full models (see Table SI 1 for detailed information).

For each model we conducted a full-null model comparison which aims at avoiding 'cryptic multiple testing' and keep type one error rate at the nominal level of 0.05 (Forstmeier and Schielzeth, 2011). As we had a genuine interest in all predictors present in the fixed effects part of each model the null models comprised only the intercept in the fixed effects part but were otherwise identical to the respective full model. This full-null model comparison utilized a likelihood ratio test (Dobson, 2002). Tests of individual effects were also based on likelihood ratio test, comparing a full model with a each in a set of reduced models which lacked fixed effects one at a time.

1.2 Implementation

We fitted all models in R (version 4.3.2; R Core Team, 2023). In model 1a and 2a we included the response as a two columns matrix with the number of targets hit and not hit in the first and second column respectively (Baayen, 2008). The models were then fitted with a binomial error structure and logit link function (McCullagh and Nelder, 1989). In essence, such models model the proportion of targets hit. We are aware that in principle one would need an 'observation level random effect' which would link the number of targets hit and not hit in a given stage. However, in a large proportion of stages (24.4%) there was only a single target that appeared, making it unlikely that a respective random effect can be fitted successfully. Model 2 a, b and 3 a, b we fitted with a beta error distribution and logit link function (Bolker, 2008). Models fitted with a beta error distribution cannot cope with values in the response being exactly 0 or 1. Hence, when such values were present in a given response variable we transformed then as suggested by Smithson and Verkuilen (2006).

Models 1a and 2a we fitted using the function glmer of the package lme4 (version 1.1-34; Bates et al., 2015), and models 2 and 3 we fitted using the function glmmTMB of the equally named package (version 1.1.8; Brooks et al., 2017). We determined model stability by dropping levels of the random effects factors, one at a time, fitting the full model to each of the subsets, and finally comparing the range of fixed effects estimates obtained from the subsets with those obtained from the model fitted on the respective full data set. This revealed all models to be of good stability. We estimated 95% confidence limits of model estimates and fitted values by means of parametric bootstraps (N=1000 bootstraps; function bootMer of the package lme4 for model 1 and function simulate of package glmmTMB for models 2 and 3). In none of the models the response was overdispersed (maximum dispersion parameter: 1.0).

Table SI 1. Random effects structure and sample size of the models.

model	random effects structure	sample size
1a (prob. of hit)	(1 + cond*coh + dur + stim + block + day|ID) + (1 + coh + dur + stim + block||dayID) + (1 + coh + dur + stim||blockID)	118218 stimulus states 39 participants 140 dayIDs 295 block IDs
1b (eccentricity)	(1 + cond*coh + dur + stim + block + day|ID) + (1 + coh + dur + stim + block||dayID) + (1 + coh + dur + stim|blockID)	143503 stimulus states 39 participants 140 dayIDs 295 block IDs
1c (accuracy)	(1 + cond*coh + dur + stim + block + day||ID) + (1 + coh + dur + stim + block||dayID) + (1 + coh + dur + stim||blockID)	33581 stimulus states 39 participants 140 dayIDs 295 block IDs

| indicates that estimates for the correlations among random intercepts were included, || that they were not; abbreviations: cond: condition; coh: coherence; dur: stimulus duration; stim: stimulus number; block: block number; day: day number; cond*coh means condition, coherence, and their interaction

2 Results

2.1 Solo and human-computer dyad condition; model 1

In case of target hit probability model (model 1a) we found a clearly significant full-null model comparison (likelihood ratio test: $\chi^2 = 131.291, df = 7, P<0.001$), and also the interaction between condition and coherence was clearly significant (Table SI 2). Plotting the model and the data, revealed the target hit probability increased steeper in the solo condition as compared to the computer condition (Fig. SI 1). Furthermore, target hit probability increased with stimulus duration (Fig. 2), stimulus number Fig. SI 3, block number (Fig. SI 4), and over the days the experiment commenced (Fig. SI 5).

Also for the model with eccentricity being the response (model 1b) we found a clearly significant full-null model comparison ($\chi^2 = 88.59, df = 7, P<0.001$). Again also the interaction between coherence and condition was clearly significant (Table SI 3), and, with the exception of block number, also all other terms in this model revealed significance. Plotting the data and the fitted model revealed that eccentricity was about the same in the in the solo and the computer condition when the coherence of the random dot pattern was low but increased a little steeper in the solo condition (Fig. SI 6). Furthermore, eccentricity slightly increased with increasing stimulus duration (Fig. SI 7), slightly decreased with stimulus number (Fig. SI 8), slightly but not significantly increased with block number (Fig. SI 9), and increased clearly over the course of the experiment (Fig. SI 10).

Finally, also in case of the accuracy model (model 1c), the full-null model comparison was clearly significant ($\chi^2 = 286.962, df = 7, P<0.001$). Wit the exception of stimulus number, all effects in this model revealed significance (Table SI 4). Plotting the data and the fitted model revealed that in the solo condition accuracy was lower than in the computer condition when the coherence of the random dot pattern was low but that accuracy was about the same in both conditions when coherence was large (Fig SI 11). Furthermore, accuracy increased slightly with stimulus duration (Fig. SI. 12), did hardly change with increasing stimulus number (Fig. SI 13), increased slightly with block number (Fig. SI 14), and clearly over the course of the experiment (Fig. SI 15).

Table SI 2. Results of the full model with hit probability being the response

term	estimate	SE	CL_{lower}	CL_{upper}	χ^2	df	P	min	max
intercept	0.562	0.027	0.524	0.598				0.553	0.577
condition	0.243	0.016	0.216	0.274				0.235	0.268
coherence	0.137	0.010	0.119	0.154				0.128	0.142
stim. duration	0.078	0.005	0.069	0.087	93.943	1	<0.001	0.076	0.079
stim. nr.	0.007	0.004	-0.001	0.016	38.527	1	<0.001	0.006	0.008
block nr.	0.016	0.005	0.006	0.026	8.255	1	0.004	0.013	0.018
day nr.	0.111	0.042	0.068	0.147	44.490	1	<0.001	0.090	0.168
condition:coherence	-0.070	0.012	-0.094	-0.048	26.386	1	<0.001	-0.076	-0.058

indicated are estimates, together with their standard errors, confidence limits, significance tests, and the range of estimates when excluding individuals one at a time; all covariates were z-transformed, mean and standard of the original variables were 0.338 and 0.320 (coherence), 245.840 and 142.113 (stim. nr.), 1907.223 and 359.506 (stim. dur.), 1.640 and 0.720 (block nr.), and 6.832 and 8.573 (day nr.) respectively; condition was dummy coded with CPRsolo being the reference level

Table SI 3. Results of the full model with eccentricity being the response

term	estimate	SE	χ^2	df	P
intercept	0.994	0.175			
condition	-0.150	0.069			
coherence	0.447	0.030			
stim. nr.	-0.024	0.009	5.921	1	0.015
stim. duration	0.027	0.004	27.135	1	<0.001
block nr.	0.034	0.025	1.763	1	0.184
day nr.	0.571	0.272	4.256	1	0.039
condition:coherence	-0.176	0.022	38.027	1	<0.001

indicated are estimates, together with their standard errors, confidence limits, significance tests, and the range of estimates when excluding individuals one at a time; all covariates were z-transformed, mean and standard of the original variables were 0.338 and 0.320 (coherence), 245.580 and 142.044 (stim. nr.), 1878.953 and 361.553 (stim. dur.), 1.640 and 0.719 (block nr.), and 6.806 and 8.551 (day nr.) respectively; condition was dummy coded with CPRsolo being the reference level; CI and stability not yet included for computational reasons

Table SI 4. Results of the full model with accuracy being the response

term	estimate	SE	CL_{lower}	CL_{upper}	χ^2	df	P	min	max
intercept	1.608	0.045	1.503	1.677				1.503	1.643
condition	0.238	0.028	0.183	0.298				0.213	0.249
coherence	0.578	0.018	0.539	0.605				0.519	0.591
stim. nr.	0.011	0.007	-0.004	0.024	2.291	1	0.130	0.008	0.013
stim. duration	0.024	0.006	0.012	0.036	16.234	1	<0.001	0.022	0.026
block nr.	0.040	0.008	0.024	0.056	22.452	1	<0.001	0.037	0.043
day nr.	0.312	0.071	0.162	0.429	15.371	1	<0.001	0.280	0.345
condition:coherence	-0.297	0.028	-0.343	-0.241	52.575	1	<0.001	-0.308	-0.252

indicated are estimates, together with their standard errors, confidence limits, significance tests, and the range of estimates when excluding individuals one at a time; all covariates were z-transformed, mean and standard of the original variables were 0.338 and 0.321 (coherence), 245.659 and 142.798 (stim. nr.), 1980.406 and 335.149 (stim. dur.), 1.639 and 0.718 (block nr.), and 6.874 and 8.600 (day nr.) respectively; condition was dummy coded with CPRsolo being the reference level

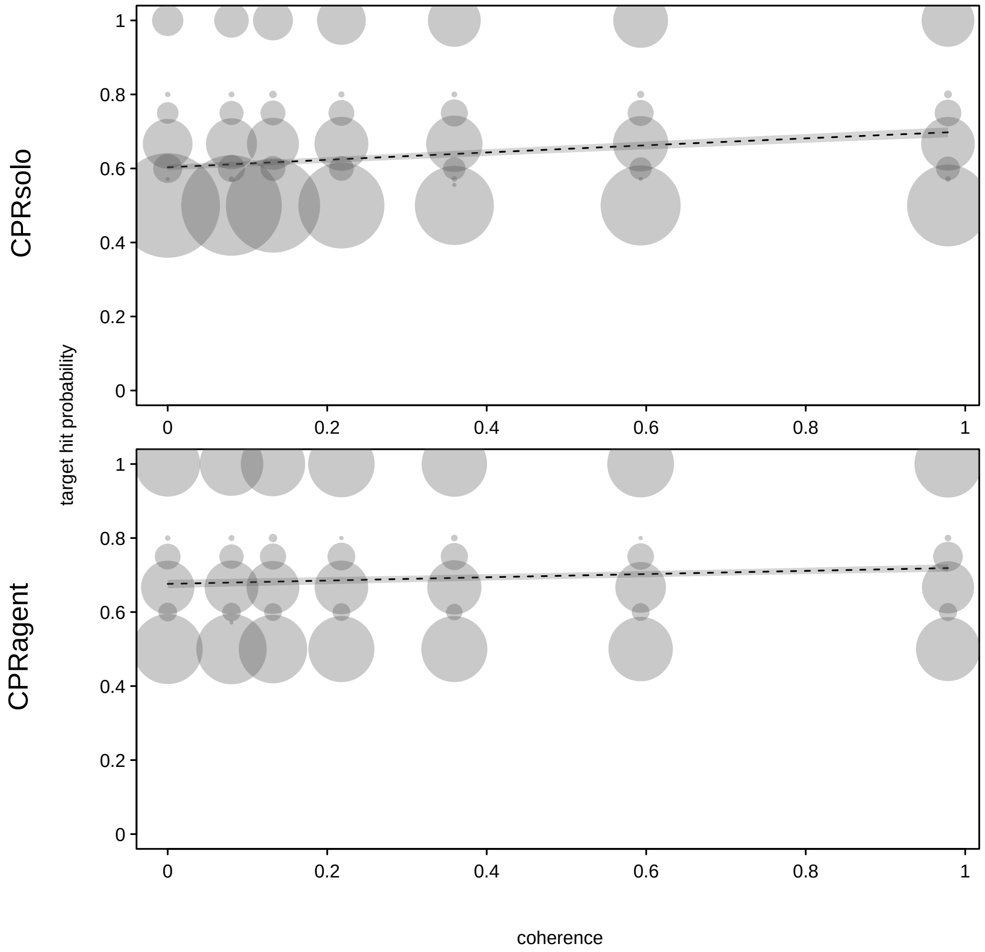

Figure SI 1. Target hit probability as a function of condition and coherence (model 1). Dashed lines and shaded areas depict the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 7 to 15602).

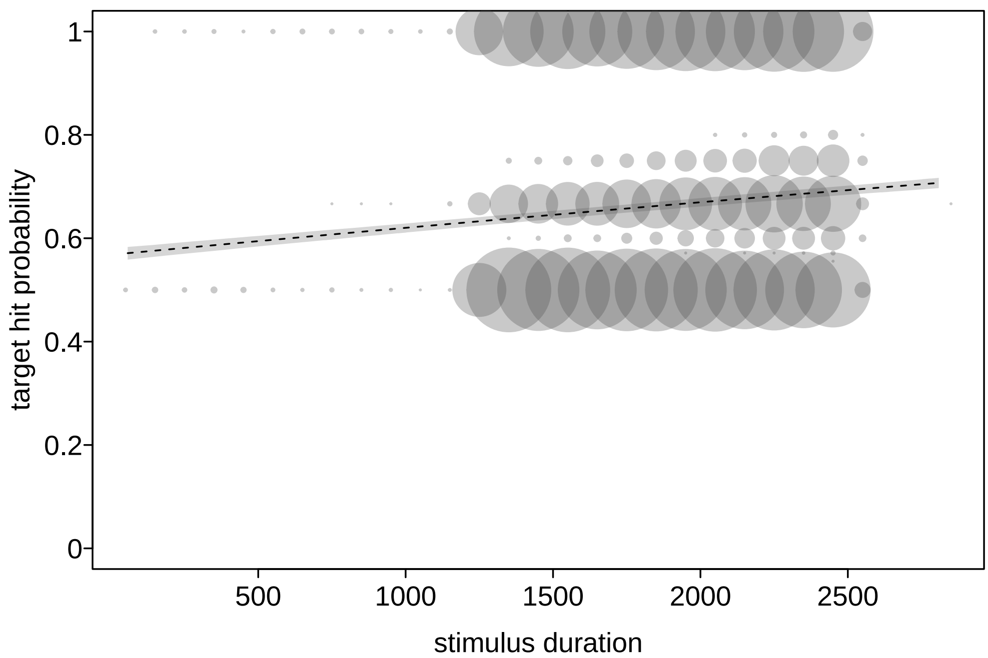

Figure SI 2. Target hit probability as a function of stimulus duration (model 1). The dashed line and shaded area depicts the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 1 to 4435).

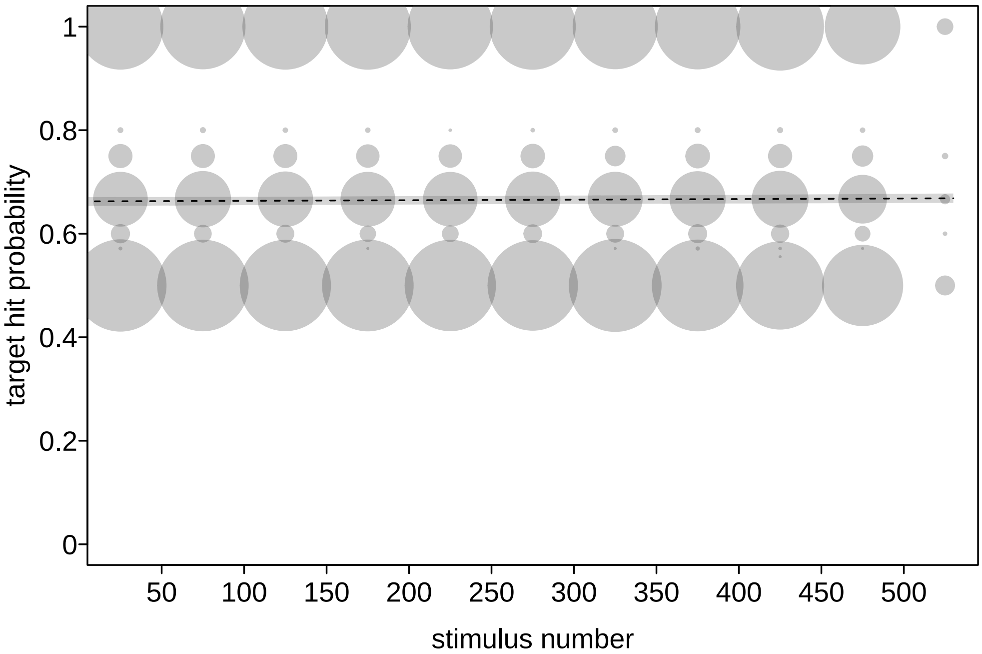

Figure SI 3. Target hit probability as a function of stimulus number (model 1). The dashed line and shaded area depicts the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 1 to 5338).

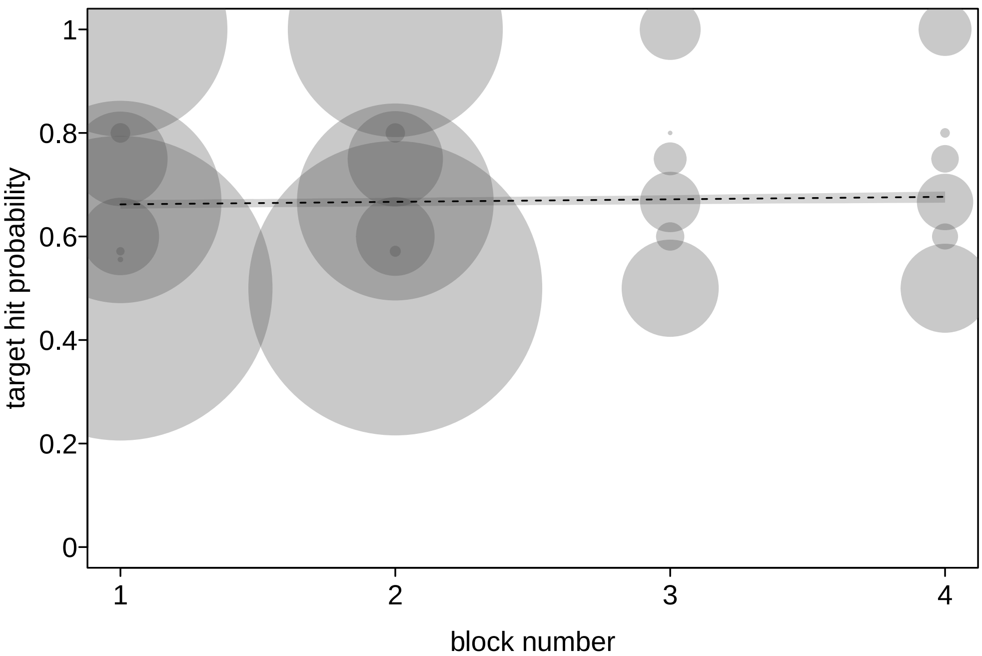

Figure SI 4. Target hit probability as a function of block number (model 1). The dashed line and shaded area depicts the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 5 to 58886).

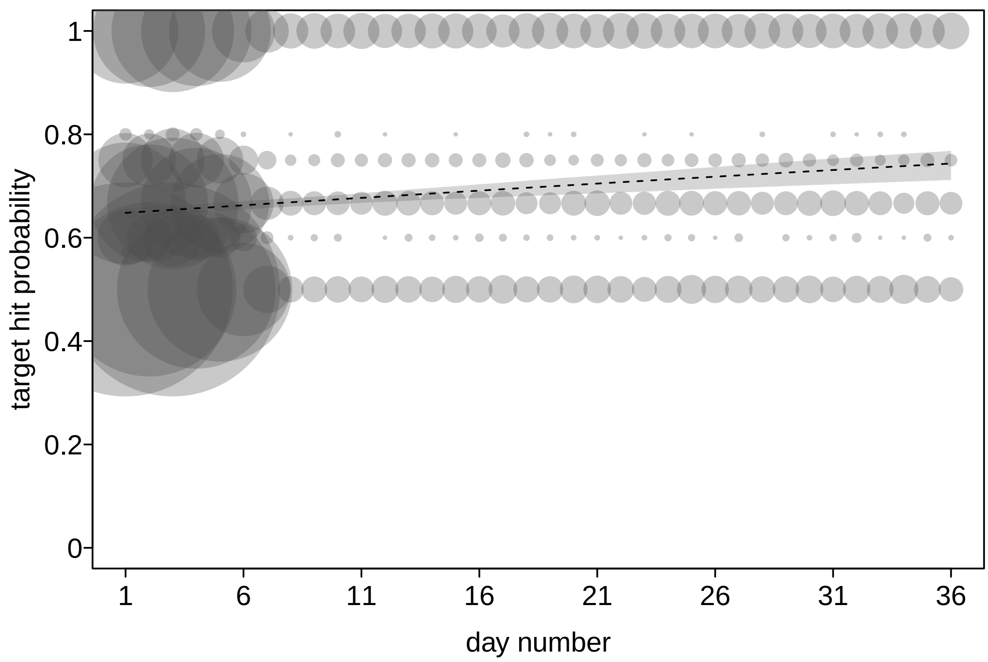

Figure SI 5. Target hit probability as a function of day number (model 1). The dashed line and shaded area depicts the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 5 to 29022).

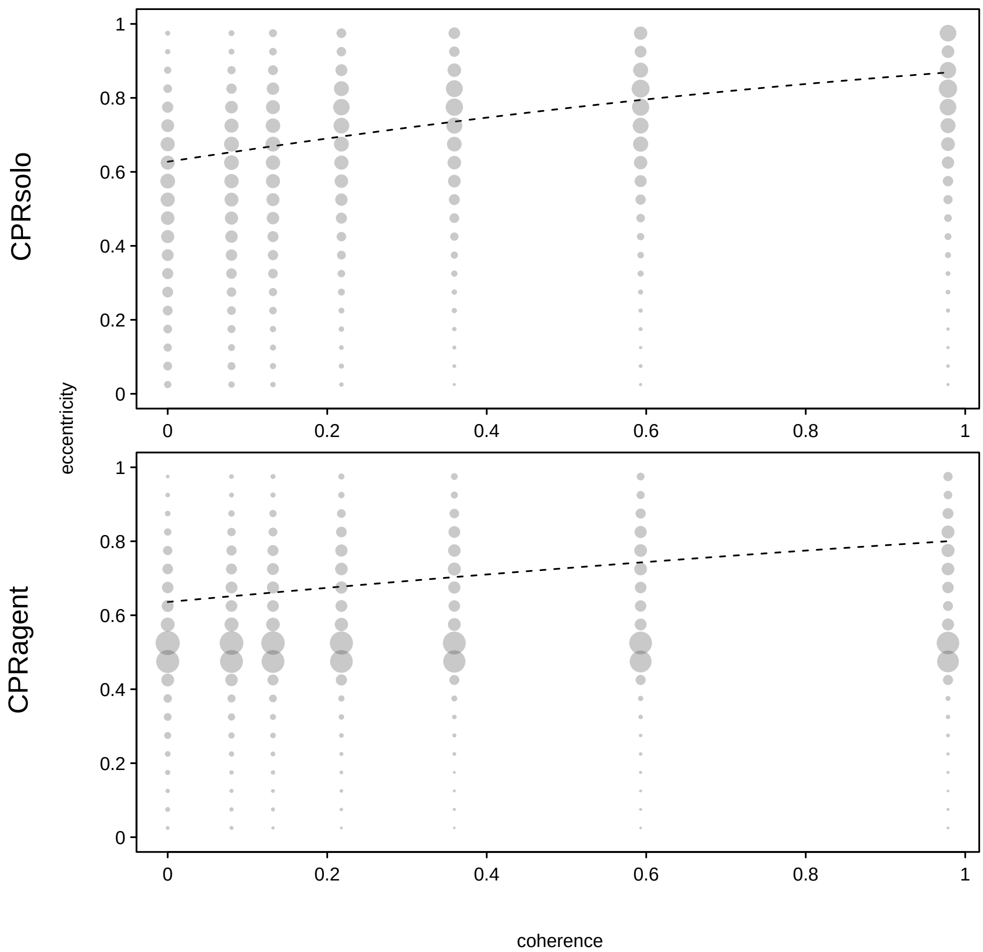

Figure SI 6. Eccentricity as a function of condition and coherence (model 1b). Dashed lines and shaded areas depict the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 4 to 2958).

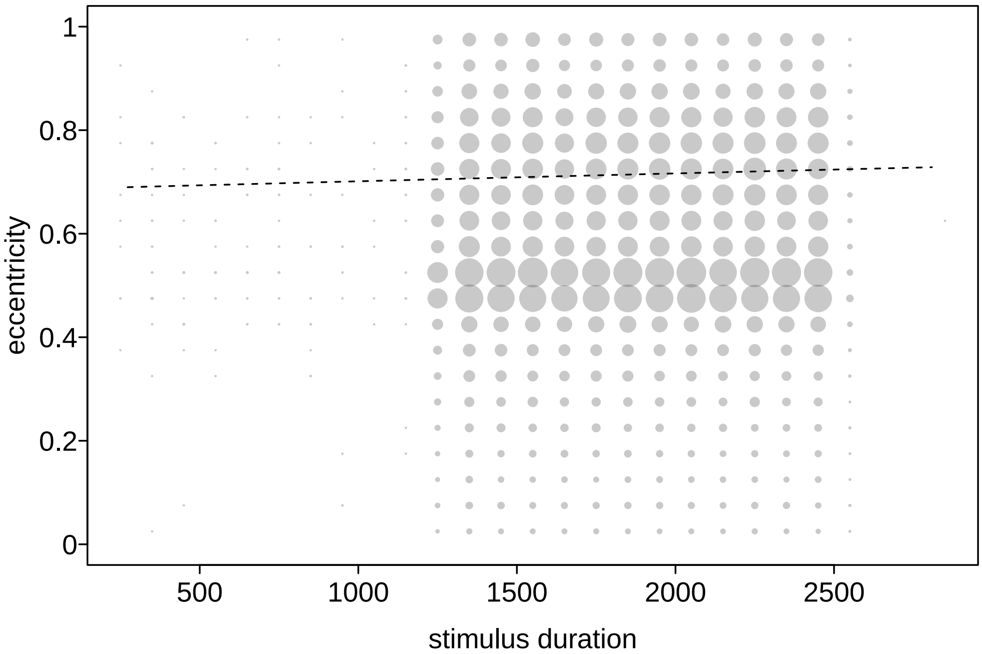

Figure SI 7. Eccentricity as a function of stimulus duration (model 1b). The dashed line and shaded area depicts the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 1 to 2022).

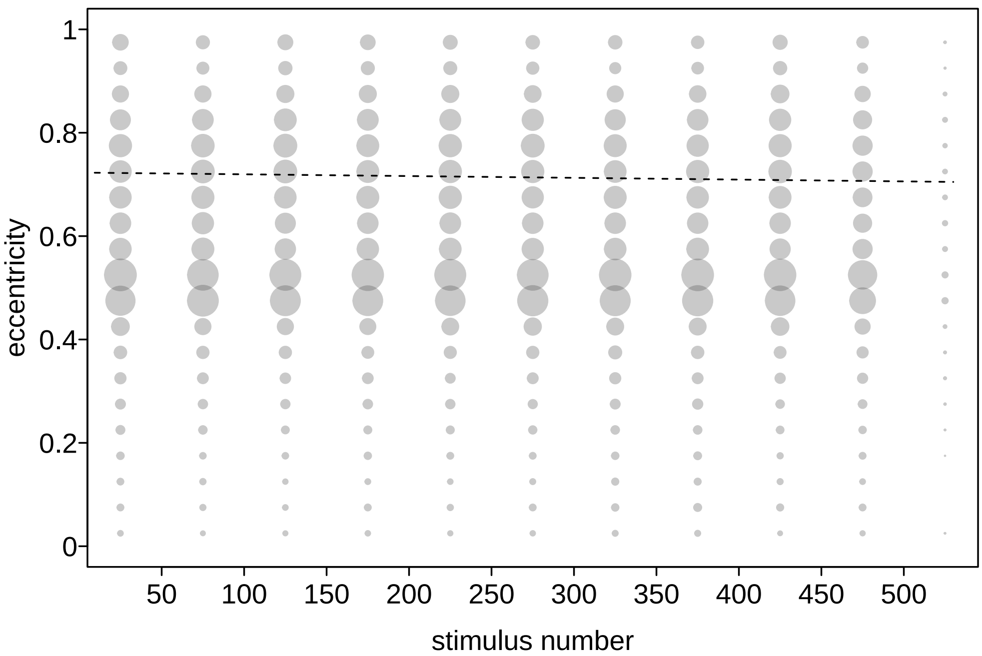

Figure SI 8. Eccentricity as a function of stimulus number (model 1b). The dashed line and shaded area depicts the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 1 to 2479).

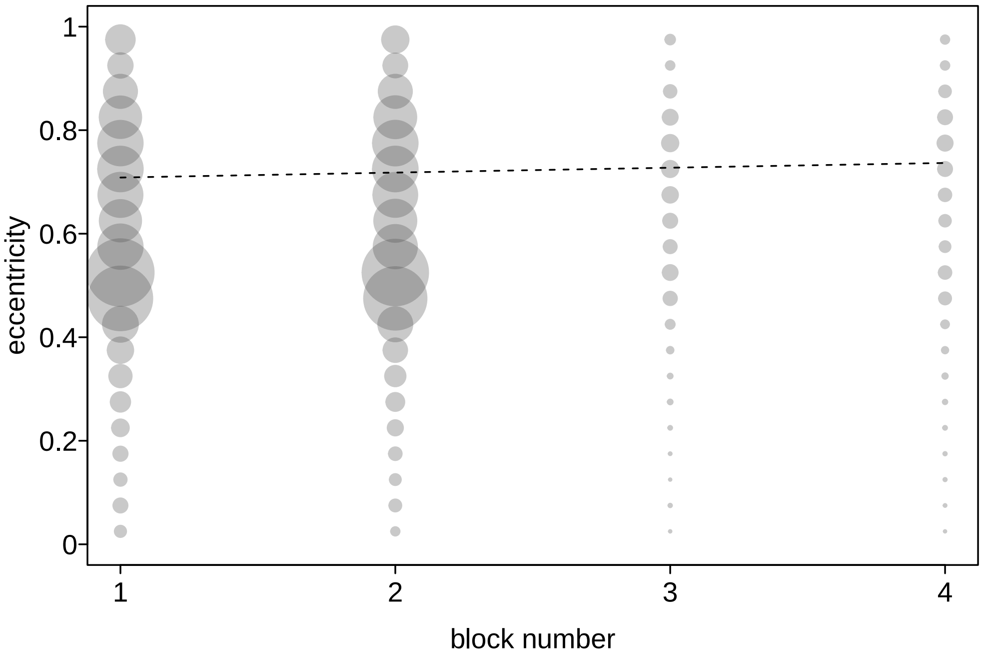

Figure SI 9. Eccentricity as a function of block number (model 1b). The dashed line and shaded area depicts the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 18 to 11411). Note that block number was not significant in this model.

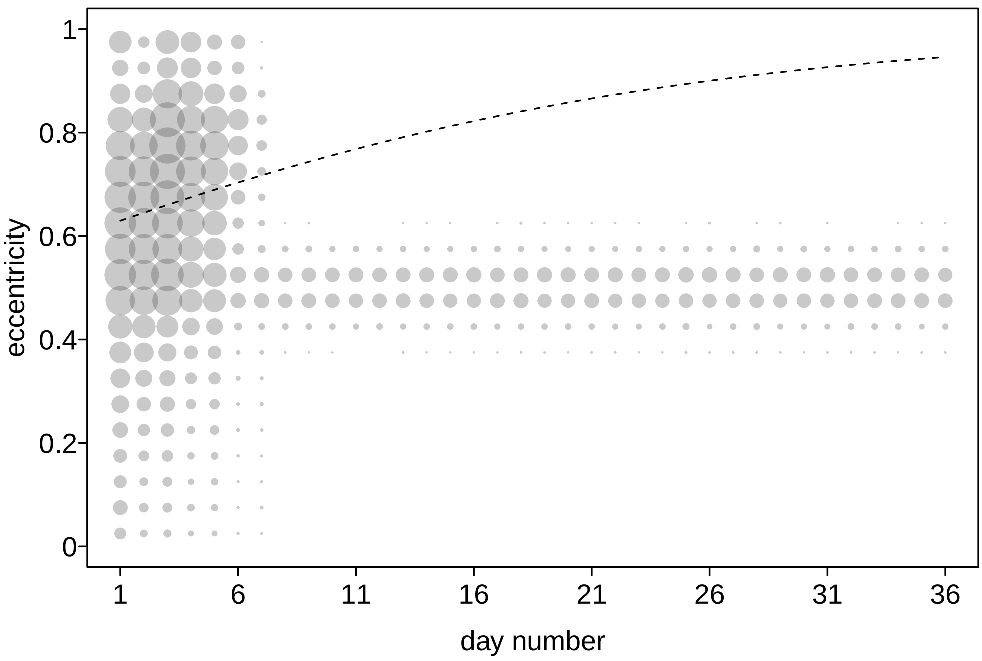

Figure SI 10. Eccentricity as a function of day number (model 1b). The dashed line and shaded area depicts the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 1 to 3050).

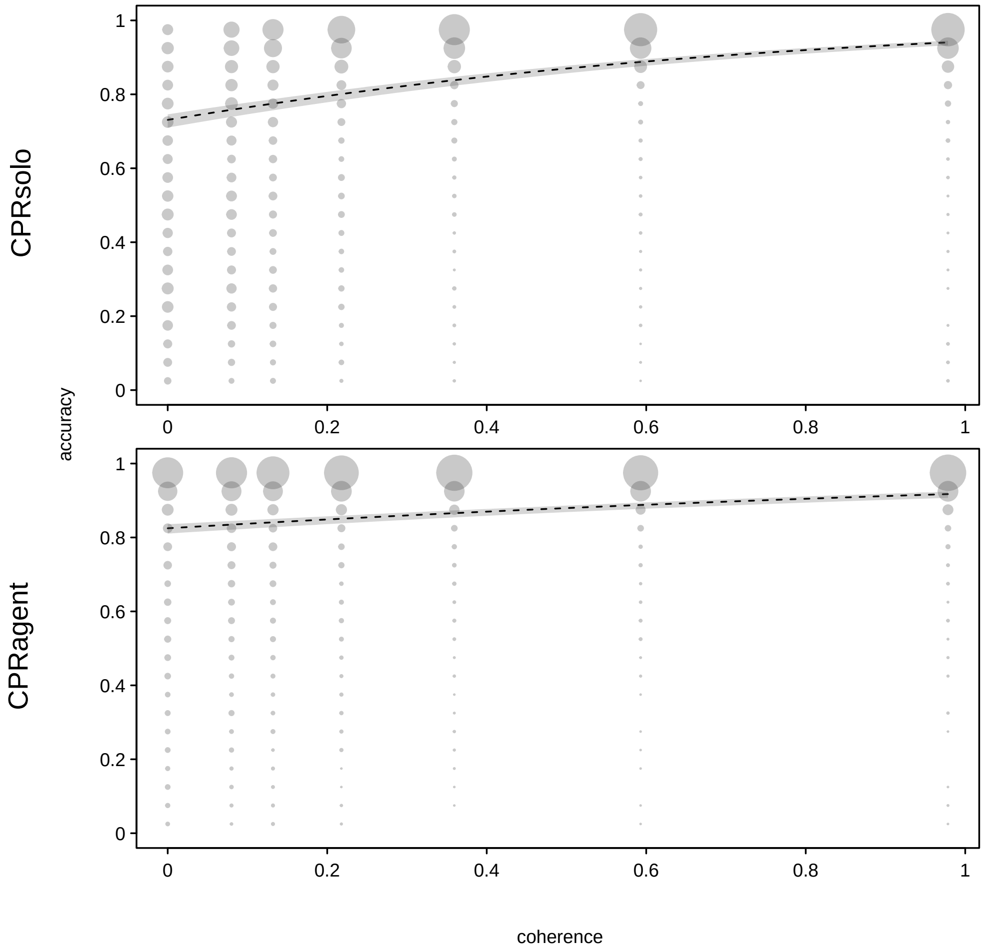

Figure SI 11. Accuracy as a function of condition and coherence (model 1b). Dashed lines and shaded areas depict the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 1 to 1781).

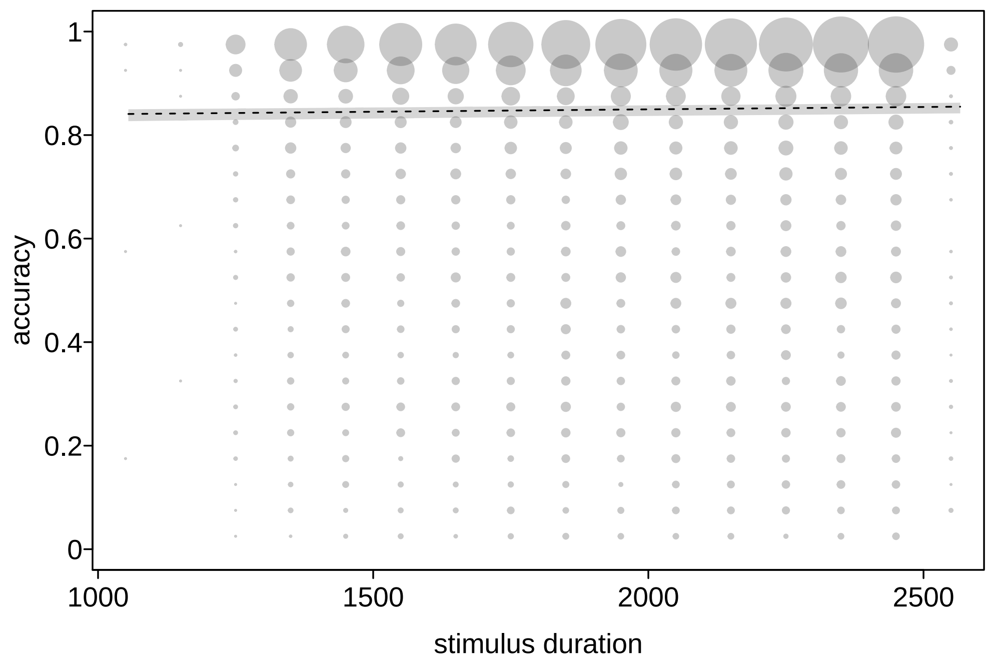

Figure SI 12. Accuracy as a function of stimulus duration (model 1c). The dashed line and shaded area depicts the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 1 to 1914).

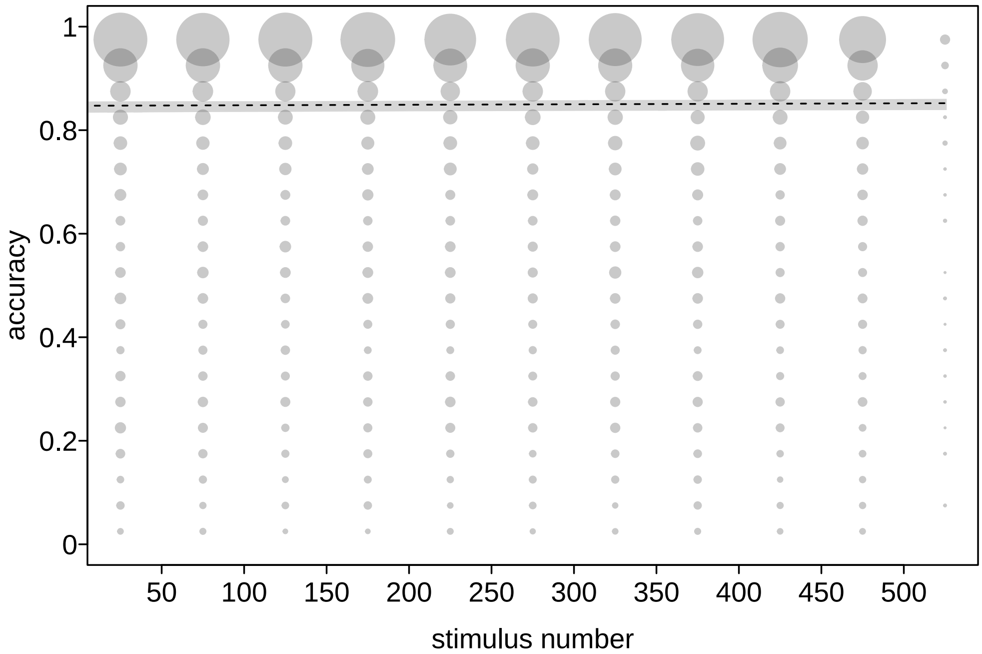

Figure SI 13. Accuracy as a function of stimulus number (model 1c). The dashed line and shaded area depicts the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 1 to 1844). Note that stimulus number was not significant in this model.

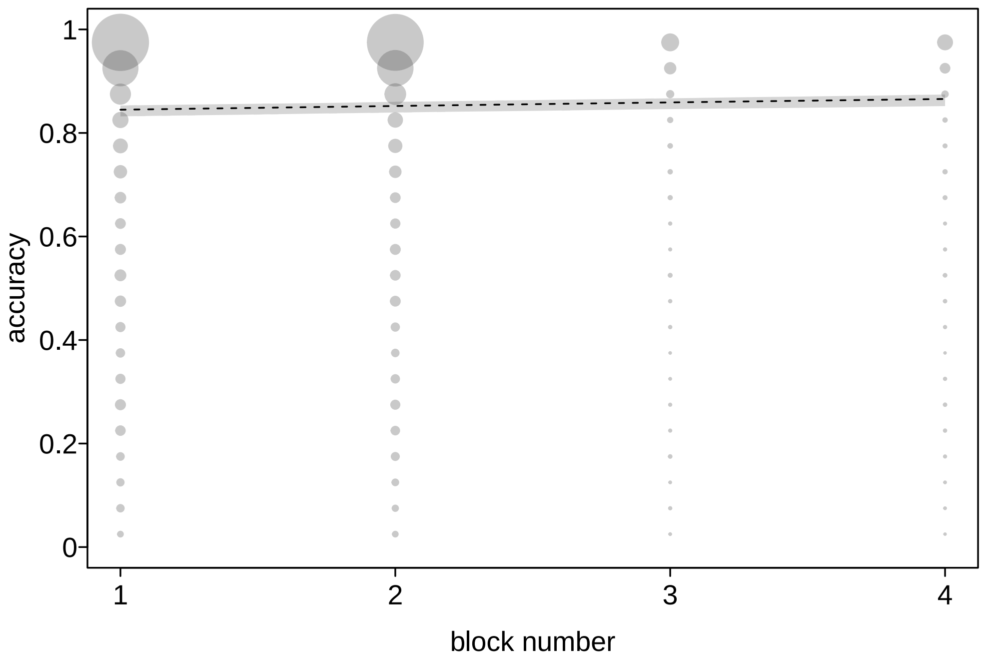

Figure SI 14. Accuracy as a function of block number (model 1c). The dashed line and shaded area depicts the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 8 to 7915).

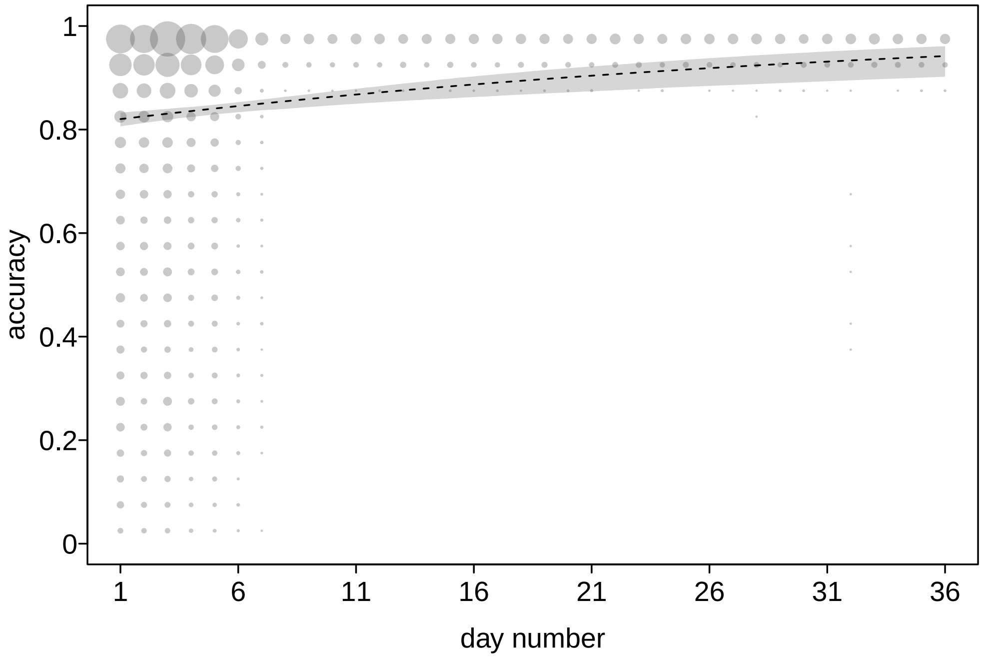

Figure SI 15. Accuracy as a function of day number (model 1c). The dashed line and shaded area depicts the fitted model and its 95% confidence limits for all other predictors in the model being at their average. Dots depict the raw data, whereby their area corresponds to the number of tied observations (range: 1 to 2878).

References

Baayen, R. H. (2008). *Analyzing Linguistic Data.* Cambridge University Press, Cambridge.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68:255–278.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48.

Bolker, B. M. (2008). *Ecological Models and Data in R.* Princeton University Press, Princeton, New Jersey.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., and Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9:378–400.

Dobson, A. J. (2002). *An Introduction to Generalized Linear Models.* Chapman & Hall/CRC, Boca Raton.

Forstmeier, W. and Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behavioural Ecology and Sociobiology*, 65:47–55.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing type I error and power in linear mixed models. *Journal of Memory and Language*, 94:305–315.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models.* Chapman and Hall, London.

R Core Team (2023). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Schielzeth, H. and Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecololgy*, 20:416–420.

Smithson, M. and Verkuilen, J. (2006). A better lemon squeezer? maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11:54–71.