

Figure S1. Novel objects used for the belugas to select across phases. All items were made out of various materials, including rubber, PVC, high-density plastic, hollow thin plastic, car wash material, duct tape, and durable water-proof cloth. Some items were solid while others were hollow.

Model Fits and Assumptions

For our GLM modeling trial-by-trial selections as a function of subject and trial number, we checked for overdispersion by calculating the Pearson dispersion parameter. We found no evidence of overdispersion (dispersion parameter = 1.28). We further assessed influential observations using standardized measures including leverage, Cook's distance, and dfbetas. While a few points had moderately high leverage, no single observation substantially influenced the model estimates or predicted values, and Cook's distances were all small (< 0.06), indicating no problems of influence. Collinearity among predictors was low, with all VIF values < 1.1.

For the GLMM modeling trial-by-trial selections as a function of time of day and rewards received on the previous trial, residuals were assessed using simulated residuals from the DHARMa package. QQ plots of the residuals closely followed the diagonal, indicating that the residuals were well-behaved and the model adequately captured the data distribution. The DHARMa dispersion test showed no significant overdispersion or underdispersion (dispersion = 0.88, p = .35). Random effect variance for whale was substantial (SD = 1.18), indicating meaningful variation among subjects. Comparison of the full model to a null model including only the random effect showed a significant improvement in fit ($\chi^2(5) = 36.84$, $p < .001$). Collinearity among fixed effects was also low (VIF < 1.1).

Table S1. Binomial GLM predictor estimates

Predictor	Estimate logit scale	SE	χ^2	df	*p*-value [b]
Intercept	-0.987	0.364			
Maple [a]	1.55	0.539			
Shila [a]	1.07	0.600			
Whisper [a]	1.89	0.555			
Trial Number	0.004	0.004			
Maple x Trial Number	-0.04	0.008			
Shila x Trial Number	0.04	0.012	57.632	3	**< .001**
Whisper x Trial Number	0.008	0.007			

χ^2values calculated from Wald Type III Chi Square Tests
[a] compared to Qinu [referent]
[b] Significant p-values are shown in bold, but some p-values are not indicated due to their limited interpretation

Table S2. Comparison of the predicted probability of selecting the risky option in the final trial for each subject.

Contrast	Odds Ratio	SE	*p* [a]
Qinu – Maple	35.62	27.92	**< .001**
Qinu – Shila	0.001	0.002	**< .001**
Qinu – Whisper	0.047	0.031	**< .001**
Maple – Shila	< 0.001	< 0.001	**< .001**
Maple – Whisper	0.001	0.001	**< .001**
Shila – Whisper	34.59	50.53	.0915

[a] Significant p-values are shown in bold
Adjusted with Bonferroni correction for multiple comparisons

Table S3. Null model comparison of Binomial GLMM

Model	AIC	BIC	$\chi^2(5)$	p
GLMM [a]	541.08	571.45	36.836	< .001
Null Model 1	567.91	576.59		

[a] Binomial GLMM predicting risk vs safe selections from time of day, number of fish received in previous trial, and their two way interaction

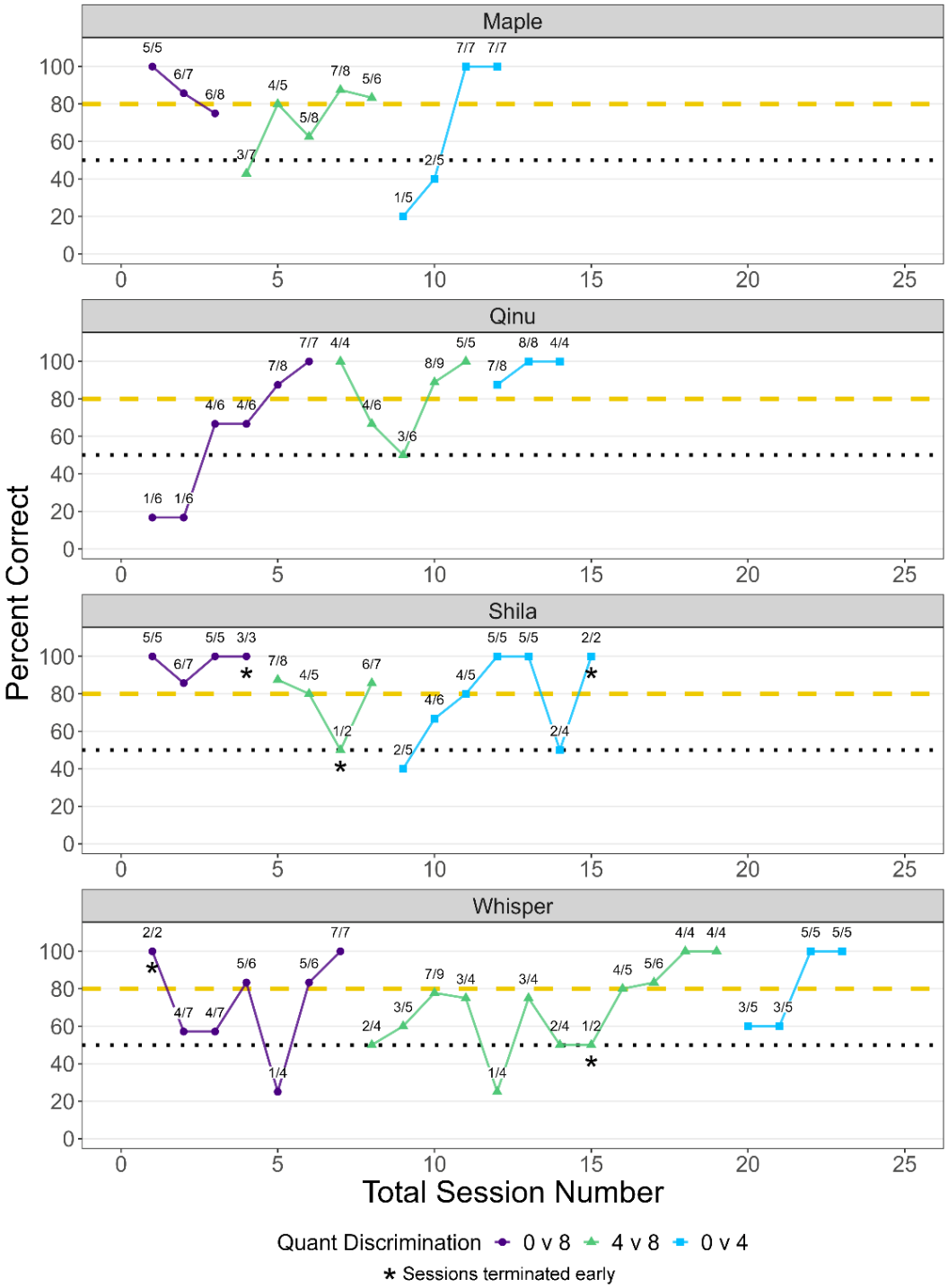

Figure S2. Proportion of higher-valued item selected across the quantity discrimination phase. Graphs are divided by 0 vs 8 trials (purple circles), 4 vs 8 trials (green triangles), and 0 vs 4 trials (blue squares). The black dotted line represents accuracy expected by chance (0.50), and the gold dashed line represents our accuracy criteria (0.80). The fractions above each point indicate the total number of trials the subject correctly chose the higher-valued item divided by the total number of trials in that session. Once a subject chose the object representing the larger quantity in 80% of 20 consecutive trials (16/20 trials), regardless of how those trials were distributed across sessions, they reached criterion on that discrimination. Asterisks indicate sessions that were terminated earlier than expected due to husbandry, logistical, or public-viewing constraints.