Supplementary information

Interaction between awareness and feeling

A further model included an awareness × feeling interaction alongside the species × feeling interaction to investigate whether the association between feeling and visitation likelihood depended on awareness after accounting for species-specific variation. The additional interaction was evaluated using a likelihood-ratio test.

To investigate whether awareness modified the association between feeling and visitation likelihood, an awareness × feeling interaction was added to the model containing the species × feeling interaction.

The additional interaction was not statistically significant (β=0.110, SE=0.092, P<0.233). Including this term did not significantly improve model fit (likelihood-ratio $\chi^2_{(1)}$=1.43, P<0.232).

Consequently, the simpler model containing the species × feeling interaction and awareness as a covariate was retained.

These results provided no clear evidence that awareness modified the association between feeling and visitation likelihood after accounting for species-specific differences in that association.

Proportional odds diagnostics and sensitivity analyses

Cumulative link mixed models were fitted using adaptive Gauss–Hermite quadrature with ten quadrature points, and model convergence was assessed using available optimisation diagnostics.

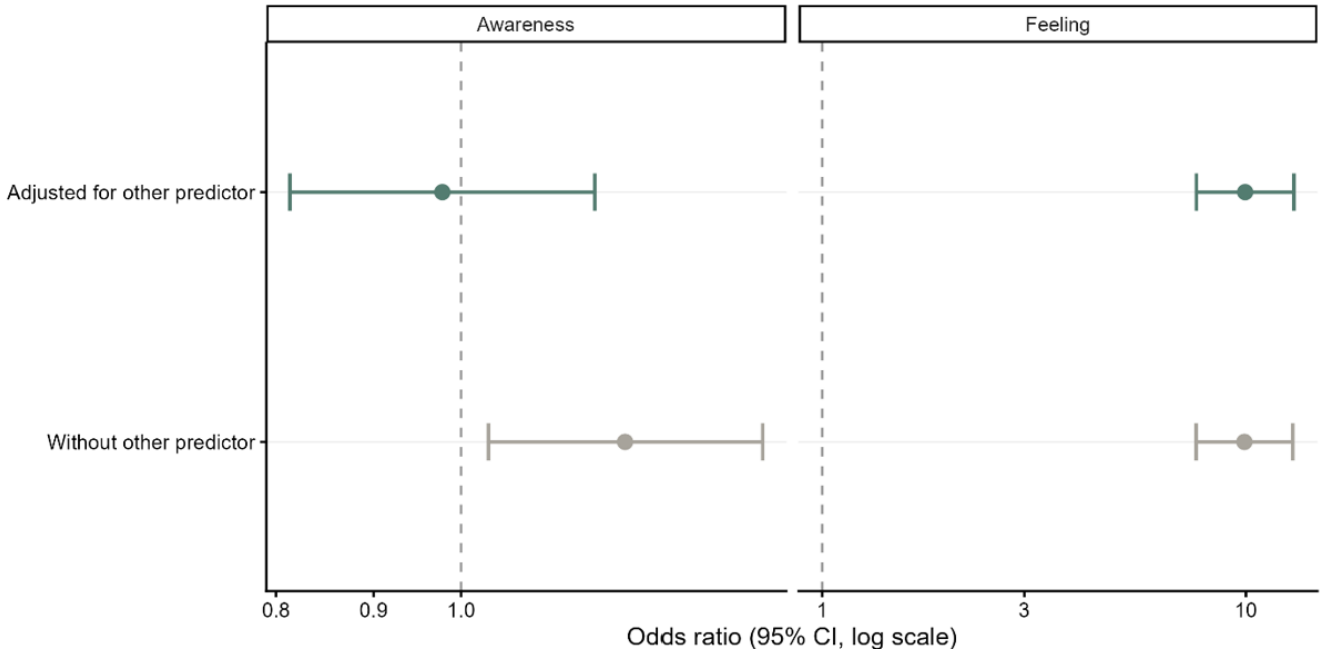
The proportional odds assumption was investigated using exploratory cumulative link models with nominal effects. Because these models did not account for participant-level clustering, their results were interpreted as diagnostic rather than definitive.

Additional sensitivity analyses used separate binomial generalised linear mixed models for four cumulative visitation-likelihood thresholds (scores ≥2, ≥3, ≥4 and 5). Each model included species, feeling, their interaction, awareness and a participant-level random intercept. These analyses examined whether the direction and magnitude of the feeling–likelihood associations were consistent across outcome thresholds. Model convergence and singularity were assessed, and threshold-specific species slopes were estimated.

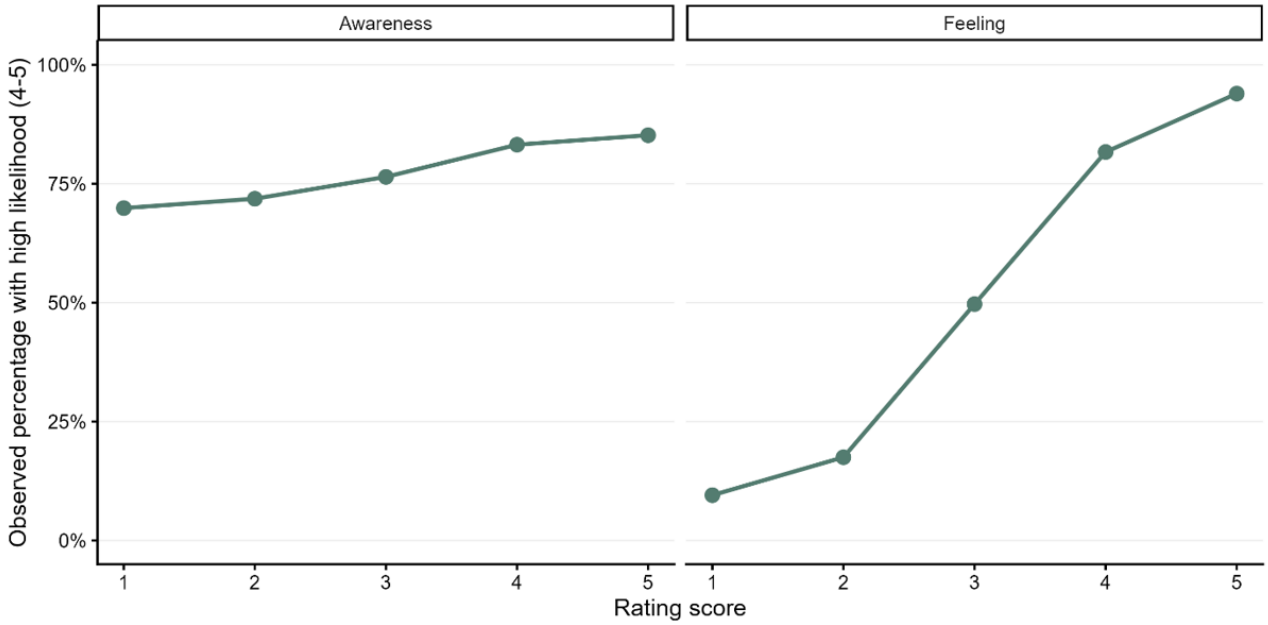
The lowest-threshold model failed to converge and was not interpreted. The remaining threshold-specific models were retained as sensitivity analyses.

Statistical significance was assessed using a nominal alpha level of 0.05, with Holm adjustment applied to the specified family of pairwise species-slope comparisons. Given the observational design, results were interpreted as associations rather than causal relationships.

Exploratory cumulative link models indicated potential violations of the proportional odds assumption. Allowing threshold-dependent effects significantly improved model fit for feeling ($\chi^2_{(3)}$=59.52, P<0.001) and awareness ($\chi^2_{(3)}$=28.09, P<0.001).



Supplementary Figure S1. Associations of awareness and feeling with stated visitation likelihood before and after mutual adjustment. Points represent odds ratios from cumulative link mixed models, with horizontal bars indicating 95% confidence intervals. Estimates are shown for models containing each predictor separately and for a model containing both predictors. All models included species as a fixed effect and participant identity as a random intercept. The dashed vertical line indicates an odds ratio of 1 (no association). The horizontal axis uses a logarithmic scale. Awareness was positively associated with visitation likelihood when examined without feeling but not after adjustment for feeling, whereas the feeling association remained strong in both models.



Supplementary Figure S2. Observed percentage of high visitation-likelihood responses across awareness and feeling ratings. Points represent the percentage of participant–species observations with a visitation-likelihood score of 4 or 5 at each awareness or feeling score. Data are pooled across species and participants. Lines connect observed percentages for descriptive purposes and do not represent fitted regression relationships. The figure illustrates the stronger descriptive association between feeling ratings and high stated visitation likelihood compared with awareness ratings.

A nominal effect for species alone did not improve fit ($\chi^2_{(27)}$=21.29, P<0.773). Additional diagnostics also suggested threshold-dependent species × feeling effects.

Because these exploratory models did not incorporate participant-level random intercepts, their results were interpreted cautiously.

To assess the robustness of the principal findings, four threshold-specific binomial generalised linear mixed models were fitted, each including a participant-level random intercept.

The model distinguishing likelihood scores of 1 from scores of 2–5 failed to converge, with a degenerate Hessian and extreme coefficient estimates. These results were consistent with possible complete or quasi-complete separation, and estimates from this threshold were not interpreted.

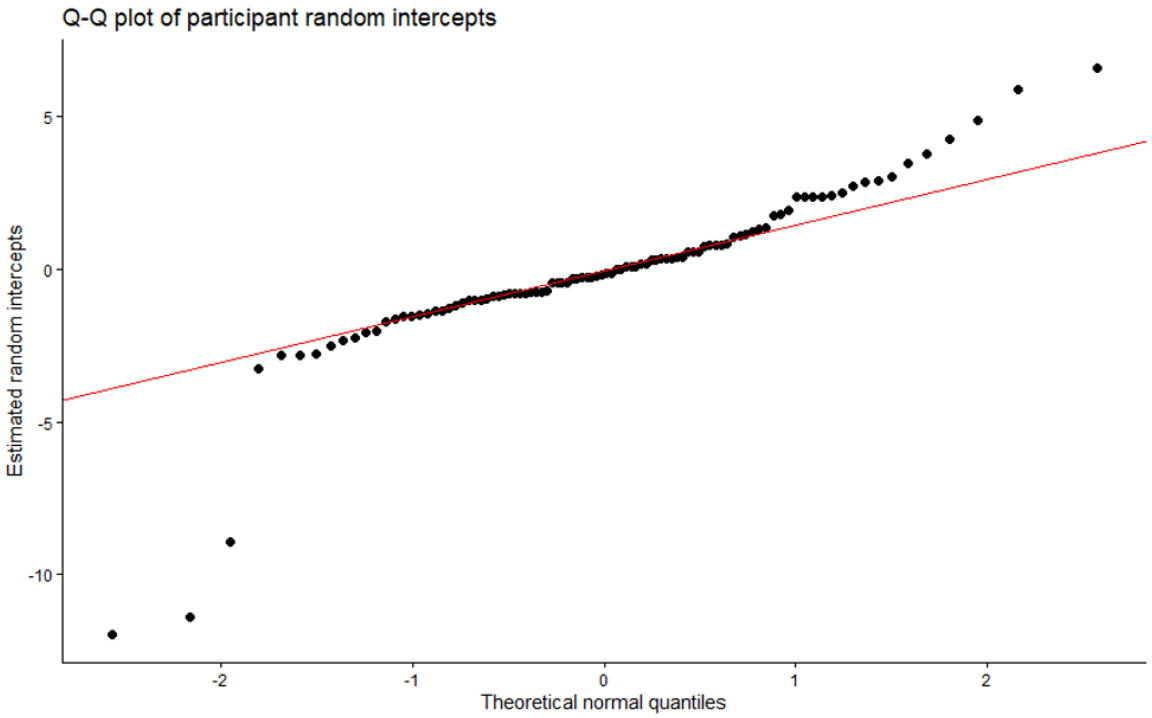
The remaining three models, representing likelihood scores ≥3, ≥4 and 5, converged without reported warnings and did not produce singular fits.

Across all ten species, feeling was positively associated with visitation likelihood at each of these three thresholds. Species-specific coefficients ranged from 1.83 to 3.31 for likelihood

Supplementary Table S1. Species-specific associations between feeling and stated visitation likelihood across three cumulative response thresholds. The table presents the estimated feeling coefficients (β), standard errors (SE) and P values from the three converged binomial generalised linear mixed models. Each model included species, the species × feeling interaction, awareness and a participant-level random intercept.

Species	Likelihood >=3	Likelihood >=4	Likelihood=5
Aldabra giant tortoise	2.92 (0.93)	2.49 (0.61)	3.40 (0.62)
Annam leaf turtle	3.31 (0.84)	2.55 (0.48)	2.75 (0.55)
Axolotl	1.94 (0.84)	1.49 (0.53)	1.20 (0.49)
Dyeing poison dart frog	2.64 (0.70)	2.29 (0.59)	2.54 (0.65)
Golden mantella	1.83 (0.66)	1.71 (0.43)	1.29 (0.50)
Green anaconda	2.42 (0.60)	2.24 (0.44)	2.44 (0.45)
King cobra	2.21 (0.52)	2.48 (0.49)	2.97 (0.54)
Komodo dragon	2.58 (0.73)	2.53 (0.58)	3.07 (0.55)
Panther chameleon	2.82 (1.29)	3.24 (0.83)	3.28 (0.59)
West African dwarf crocodile	2.49 (0.71)	2.53 (0.55)	2.73 (0.53)

Journal of Zoo and Aquarium Research 14(4) 2026 supplement
https://doi.org/10.19227/66fw4378



Supplementary Figure S3. Normal Q-Q plot of participant-level random intercepts from the preferred cumulative link mixed model.

≥3, from 1.49 to 3.24 for likelihood ≥4, and from 1.20 to 3.40 for likelihood=5. All species-specific feeling associations were statistically significant at each threshold (all P<0.05).

Although coefficient magnitudes varied between thresholds and species, the consistently positive associations supported the principal finding that more favourable feeling ratings were associated with greater stated likelihood of visiting.

Nevertheless, the variation in estimated coefficients and the exploratory proportional odds diagnostics indicated that the common-slope assumption may be restrictive. Accordingly, the primary cumulative link mixed-model results were interpreted alongside the threshold-specific sensitivity analyses, with particular caution regarding the magnitude of species-specific effects.

Model validation and sensitivity analyses

Supplementary Methods

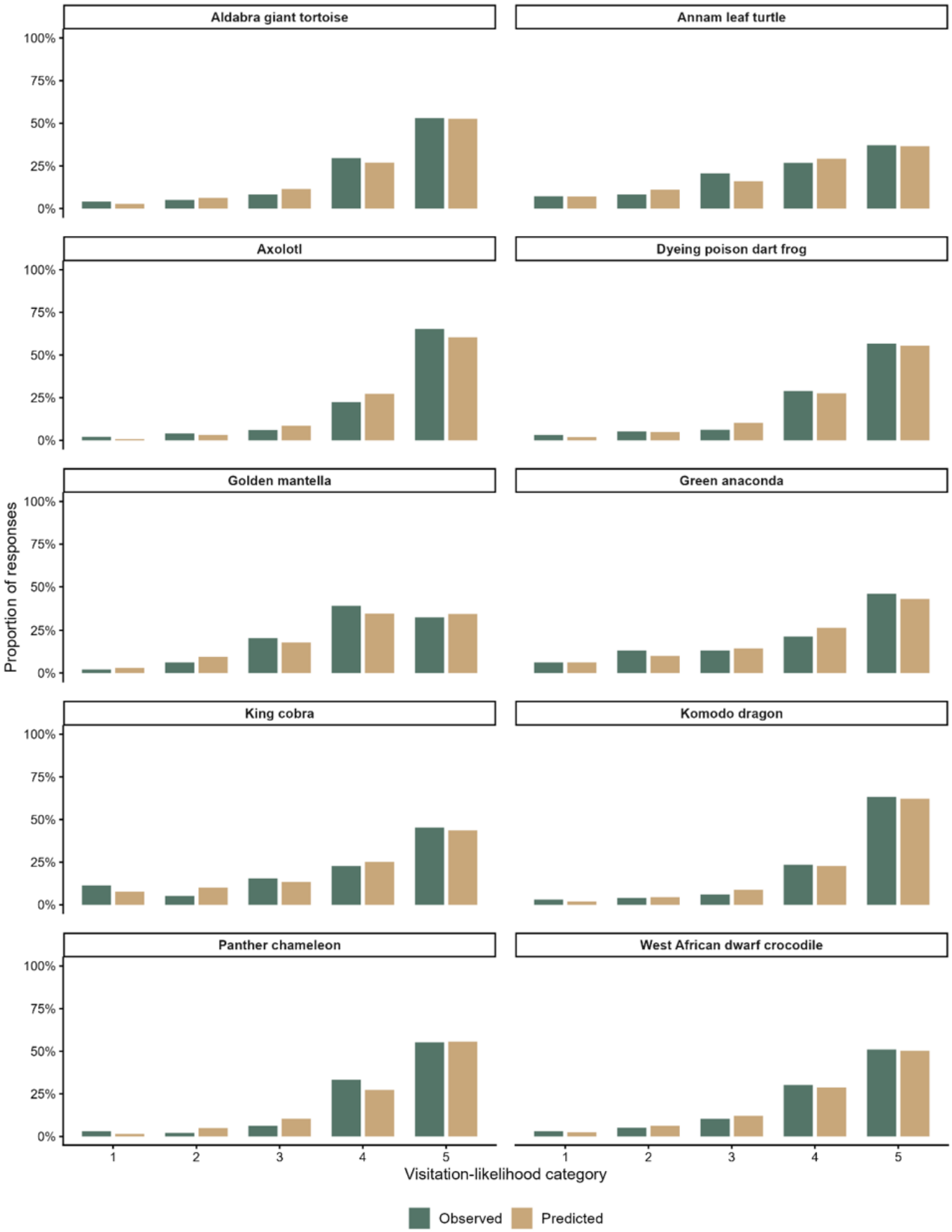
Model convergence and numerical stability

The preferred cumulative link mixed model (CLMM; M4),

Supplementary Table S2. Species-specific associations between feeling and visitation likelihood across cumulative response thresholds. Estimated log-odds coefficients (β) from separate threshold-specific binomial generalised linear mixed models.

Species	Likelihood >=3	Likelihood >=4	Likelihood=5
Aldabra giant tortoise	2.92 (0.93)	2.49 (0.61)	3.40 (0.62)
Annam leaf turtle	3.31 (0.84)	2.55 (0.48)	2.75 (0.55)
Axolotl	1.94 (0.84)	1.49 (0.53)	1.20 (0.49)
Dyeing poison dart frog	2.64 (0.70)	2.29 (0.59)	2.54 (0.65)
Golden mantella	1.83 (0.66)	1.71 (0.43)	1.29 (0.50)
Green anaconda	2.42 (0.60)	2.24 (0.44)	2.44 (0.45)
King cobra	2.21 (0.52)	2.48 (0.49)	2.97 (0.54)
Komodo dragon	2.58 (0.73)	2.53 (0.58)	3.07 (0.55)
Panther chameleon	2.82 (1.29)	3.24 (0.83)	3.28 (0.59)
West African dwarf crocodile	2.49 (0.71)	2.53 (0.55)	2.73 (0.53)

Note. Each model included species × centred feeling, centred awareness and a participant-level random intercept. Coefficients represent the association between a one-unit increase in centred feeling and the log odds of meeting the indicated visitation-likelihood threshold, within each species. All reported species-specific feeling coefficients were positive and statistically significant (P<0.05) in the three converged models. The model for likelihood ≥2 was excluded because of convergence difficulties.



Supplementary Figure S4. Observed and model-predicted visitation-likelihood distributions across ten reptile and amphibian species.

Journal of Zoo and Aquarium Research 14(4) 2026 supplement
https://doi.org/10.19227/66fw4378

Supplementary Table S3. Overall cross-validated predictive performance of the preferred cumulative link mixed model.

Predictive performance measure	Preferred CLMM (M4)	Baseline model	Relative reduction (%)
Normalised ranked probability score (RPS)	0.1031	0.1403	26.53
Multiclass log loss	1.0888	1.2681	14.14

Note. Predictions for held-out participants were marginalised over the fitted participant-level random-intercept distribution using 30-point Gauss–Hermite quadrature. The baseline used response-category frequencies estimated in each training fold. Lower RPS and log-loss values indicate better predictive performance. Relative reduction=100 × (baseline score − M4 score) / baseline score

incorporating species, feeling, the species × feeling interaction, awareness and a participant-level random intercept, was evaluated for numerical convergence and stability. The model was fitted using a logit link and adaptive Gauss–Hermite quadrature with ten quadrature points. Convergence was assessed using the optimiser convergence code, maximum absolute gradient, Hessian eigenvalues and condition number, and the ordering of the estimated response thresholds.

Assessment of random effects and sensitivity analysis

The distribution of participant-level random-intercept conditional modes was examined using descriptive statistics and a normal quantile–quantile (Q-Q) plot. Participants with extreme conditional modes were investigated by examining their visitation-likelihood response patterns across the ten species.

To evaluate the influence of extreme response patterns, a sensitivity analysis was conducted excluding five participants who provided consistently very low or high visitation-likelihood ratings. The preferred CLMM was refitted to the reduced dataset using the same model structure and estimation procedure. Model convergence, participant-level random-intercept variance and species-specific feeling coefficients were compared descriptively between the original and sensitivity models.

Assessment of the proportional odds assumption

The proportional odds assumption was investigated using exploratory nominal-effects tests in a cumulative link model containing the same fixed-effects structure as the preferred CLMM, but without participant-level random effects. These tests assessed whether the effects of species, feeling, awareness and the species × feeling interaction varied across cumulative response thresholds.

As a sensitivity analysis, separate binomial generalised linear mixed models were fitted for four cumulative dichotomisations of visitation likelihood (scores ≥2, ≥3, ≥4 and 5), retaining the species × feeling interaction, awareness and participant-level random intercept. The direction and magnitude of species-specific feeling associations were examined across thresholds. These analyses were used to assess the robustness of the direction of associations rather than to replace the primary ordinal model.

Model calibration

Model calibration was assessed by comparing observed response proportions with population-averaged predicted probabilities across the five ordered visitation-likelihood categories, both overall and separately for each species. Predicted probabilities were calculated from the preferred CLMM by integrating over the estimated normal distribution of participant-level random intercepts using 30-point Gauss–Hermite quadrature.

Agreement between observed and predicted proportions was quantified using the mean absolute difference and maximum absolute difference across response categories, expressed in percentage points. These comparisons assessed in-sample aggregate calibration.

Cross-validated predictive performance

Out-of-sample predictive performance was evaluated using five-fold participant-grouped cross-validation. All observations from each participant were retained within the same fold to prevent information leakage between training and test datasets. The preferred CLMM was refitted to each training dataset, and population-averaged predicted probabilities were calculated for held-out participants.

Predictive accuracy was assessed using the normalised ranked probability score (RPS), which accounts for the ordered nature of the response categories, and multiclass log loss. Lower values indicated better predictive performance. Results were compared with a baseline model assigning response-category probabilities according to their frequencies in the corresponding training dataset. Predictive performance was evaluated overall and separately for each species.

Supplementary Results

Model convergence and numerical stability

The preferred CLMM demonstrated satisfactory numerical convergence, with an optimiser convergence code of 0 and a maximum absolute gradient of 5.85×10^{-4}. All Hessian eigenvalues were positive (range: 0.765–251.684), with a condition number of approximately 329, indicating a positive-definite Hessian. All four estimated response thresholds were correctly ordered. These findings provided no evidence of numerical convergence failure.

Assessment of random effects and sensitivity analysis

Examination of the 98 participant-level random-intercept conditional modes revealed substantial between-participant variation (median=−0.16; range=−11.95 to 6.56). The normal Q-Q plot (Figure S3) showed approximate agreement with the theoretical normal distribution in the central portion, but pronounced deviations in both tails, particularly among participants with strongly negative conditional modes.

Examination of extreme response patterns identified two participants who assigned the minimum visitation-likelihood rating (1) to all ten species, two who consistently assigned the maximum rating (5), and one who provided predominantly low ratings (mean=1.3; range=1–2). These patterns provided a plausible explanation for the extreme conditional modes, although the Q-Q plot did not establish whether the underlying random-effects distribution departed from normality.

Exclusion of these five participants reduced the dataset from 98 participants and 973 observations to 93 participants and 923 observations. The refitted model demonstrated satisfactory numerical convergence (maximum absolute gradient=2.18×10^{-4}). The estimated participant-level random-intercept variance decreased from approximately 8.59 to 2.26, representing a reduction of approximately 73.7%.

Despite this reduction in estimated between-participant heterogeneity, all ten species-specific feeling coefficients remained

positive, with changes in magnitude ranging from 0.72% to 6.45% relative to the original model. The largest proportional change was observed for Annam leaf turtle and the smallest for Panther chameleon. Aldabra giant tortoise retained the largest estimated feeling coefficient, while Golden mantella retained the smallest.

These findings indicate that participants with extreme response patterns contributed substantially to estimated between-participant heterogeneity but had relatively little influence on the direction or magnitude of the species-specific feeling associations.

Normal quantile–quantile (Q-Q) plot showing the estimated conditional modes of participant-level random intercepts (n=98) from the cumulative link mixed model examining the associations between species, feeling, awareness and stated visitation likelihood. Points represent individual participant random-intercept estimates plotted against theoretical quantiles of a normal distribution. The red reference line indicates the expected pattern under approximate normality. Most estimates followed the reference line in the central portion of the distribution, although pronounced deviations were observed in both tails, particularly among participants with strongly negative random intercepts. These deviations were associated with participants reporting consistently low or high visitation-likelihood ratings across species. As conditional modes are estimated quantities subject to shrinkage, the plot provides an exploratory assessment rather than a definitive test of the normal random-effects assumption.

Proportional odds assumption

Exploratory nominal-effects tests identified evidence of departures from the proportional odds assumption for feeling ($\chi^2_{(3)}$=59.52, P<0.001), awareness ($\chi^2_{(3)}$=28.09, P<0.001) and the species × feeling interaction ($\chi^2_{(30)}$=138.86, P<0.001). No corresponding evidence was identified for the species main effect ($\chi^2_{(27)}$=21.29, P<0.773).

Of the four threshold-specific binomial mixed models, the model distinguishing likelihood scores ≥2 from scores below 2 exhibited convergence difficulties and was not interpreted. The remaining three models converged satisfactorily.

Across the three converged thresholds (scores ≥3, ≥4 and 5), species-specific feeling coefficients were consistently positive and statistically significant for all ten species. Although coefficient magnitudes varied across thresholds, the direction of the associations was preserved (Table S2).

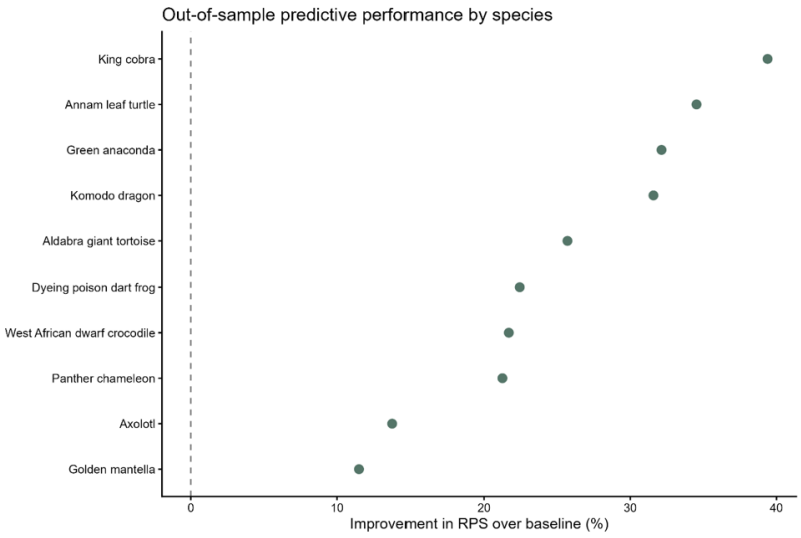
Table S2. Species-specific associations between feeling and visitation likelihood across cumulative response thresholds.

Estimated log-odds coefficients (β) from separate threshold-specific binomial generalised linear mixed models.

These findings indicate that the proportional odds assumption was not fully supported, particularly for feeling-related effects. Nevertheless, the threshold-specific analyses provided evidence that the positive association between feeling and stated visitation likelihood was robust across the converged cumulative response thresholds. Accordingly, the preferred CLMM was retained as a parsimonious summary of the overall associations, while recognising that the magnitudes of these associations may differ between response thresholds.

Model calibration

The preferred CLMM demonstrated close agreement between observed and predicted visitation-likelihood distributions across the complete dataset (Figure S4). The mean absolute difference across the five response categories was 0.91 percentage points, with a maximum difference of 1.23 percentage points. The model slightly underestimated responses in the lowest and highest



Supplementary Figure S5. Species-specific improvement in out-of-sample predictive performance of the preferred cumulative link mixed model.

Journal of Zoo and Aquarium Research 14(4) 2026 supplement
https://doi.org/10.19227/66fw4378

likelihood categories and slightly overestimated responses in categories 2 and 3.

Species-specific calibration analyses also indicated relatively small discrepancies between observed and predicted response proportions. Mean absolute differences ranged from 1.17 percentage points for West African dwarf crocodile to 3.03 percentage points for Panther chameleon. The largest individual species-category discrepancy was 6.02 percentage points, observed for Panther chameleon.

These results suggest that the preferred model reproduced the observed response distributions reasonably well, both overall and across individual species. However, these analyses represent in-sample aggregate calibration and do not independently establish individual-level predictive accuracy.

Cross-validated predictive performance

Five-fold participant-grouped cross-validation included all 98 participants and 973 observations, with 19–20 participants allocated to each fold.

The preferred CLMM demonstrated improved out-of-sample predictive performance relative to the baseline model. The mean normalised RPS was 0.1031 for the preferred model compared with 0.1403 for the baseline, representing a 26.5% reduction in prediction error. Multiclass log loss was also lower for the preferred model (1.0888) than for the baseline (1.2681), representing a 14.1% reduction (Table S3).

Species-specific analyses indicated improved predictive performance for all ten species (Supplementary Figure S5). Relative reductions in RPS ranged from 11.5% for Golden mantella to 39.4% for King cobra. The greatest improvements were observed for King cobra (39.4%), Annam leaf turtle (34.5%) and Green anaconda (32.1%). The lowest absolute RPS was observed for Komodo dragon (0.0809). Multiclass log loss was also lower for the preferred model than for the baseline across all ten species (Figure S5).

Percentage reduction in the normalised ranked probability score (RPS) for the preferred cumulative link mixed model (M4) relative to a baseline model based on training-set visitation-likelihood category frequencies. Predictive performance was assessed using five-fold participant-grouped cross-validation (98 participants; 973 observations), ensuring that all observations from each participant were retained within the same fold. Predicted probabilities for held-out participants were marginalised over the estimated normal distribution of participant-level random intercepts using 30-point Gauss–Hermite quadrature. Each point represents the percentage reduction in RPS for an individual species, with larger positive values indicating greater improvement in predictive accuracy. The dashed vertical line at 0% represents equivalent predictive performance between models. Improvements were observed for all ten species, ranging from 11.5% for Golden mantella to 39.4% for King cobra.

These findings indicate that the preferred model provided more accurate predictions for held-out participants than a baseline model based solely on response-category frequencies. The predictive improvements were consistent across species, although their magnitudes varied.

Overall interpretation

Collectively, the validation analyses indicated satisfactory numerical convergence, limited influence of extreme participant response patterns on species-specific feeling coefficients, close agreement between observed and predicted response distributions, and improved out-of-sample predictive performance relative to a category-frequency baseline.

Although the proportional odds assumption was not fully supported, threshold-specific sensitivity analyses indicated that the positive associations between feeling and stated visitation likelihood were maintained across the three converged cumulative response thresholds.

The preferred CLMM was therefore considered a useful and reasonably robust summary of the associations between species, feeling, awareness and stated visitation likelihood, while acknowledging uncertainty regarding the proportional odds and random-effects distributional assumptions.