Supplementary Materials

E2IATLAS: An Emotional and Environmental Image Atlas of Nature Photographs

These supplementary materials provide extended descriptive, psychometric, validation, and individual-difference analyses that complement the main manuscript. Each section briefly describes the analytic approach, presents the associated figures, reports the corresponding results, and closes with a short interpretive summary. All underlying data are available on the Open Science Framework (https://osf.io/tz74d/).

Contents

S1. Sample Composition and Data Quality

Analytic approach

The rating sample was characterised descriptively. Age, sex, ethnicity, student and employment status, and country of residence were summarised as counts and proportions. Geographic and climatic context was derived from each participant's location: latitude and hemisphere were computed from coordinates, and a Köppen–Geiger climate-zone proxy was assigned. Two data-quality indicators were inspected, study completion time and Prolific approval reputation, and consent-revoked cases were flagged for exclusion. Analyses requiring complete demographic records used the matched subset (N = 1,868).

Figures

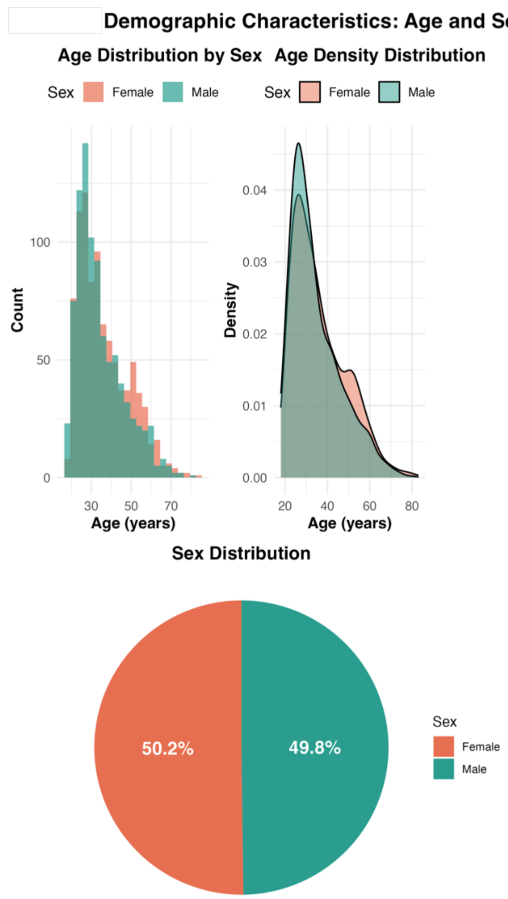

Figure S1. Age and sex composition of the rating sample. Age distribution by sex (histogram and density) and the overall sex split (50.2% female, 49.8% male).

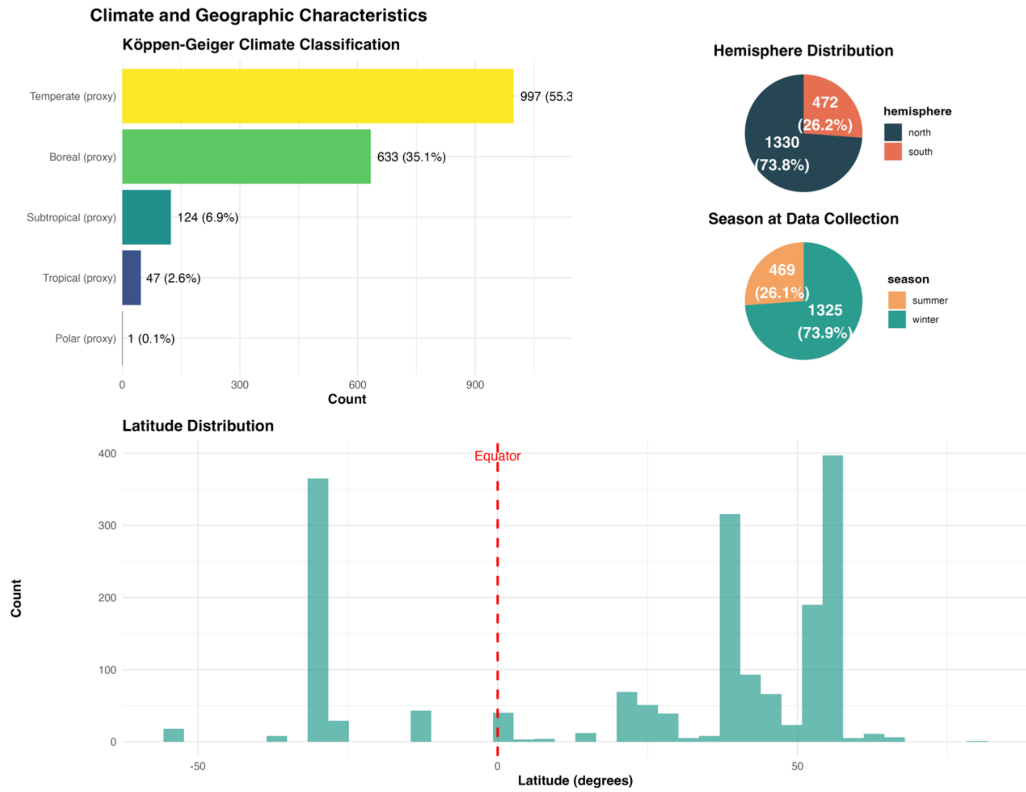

Figure S2. Climate and geographic characteristics of participants. Köppen–Geiger climate-zone classification (proxy), hemisphere distribution (73.8% Northern, 26.2% Southern), season at data collection (73.9% winter, 26.1% summer), and the absolute-latitude distribution of participants.

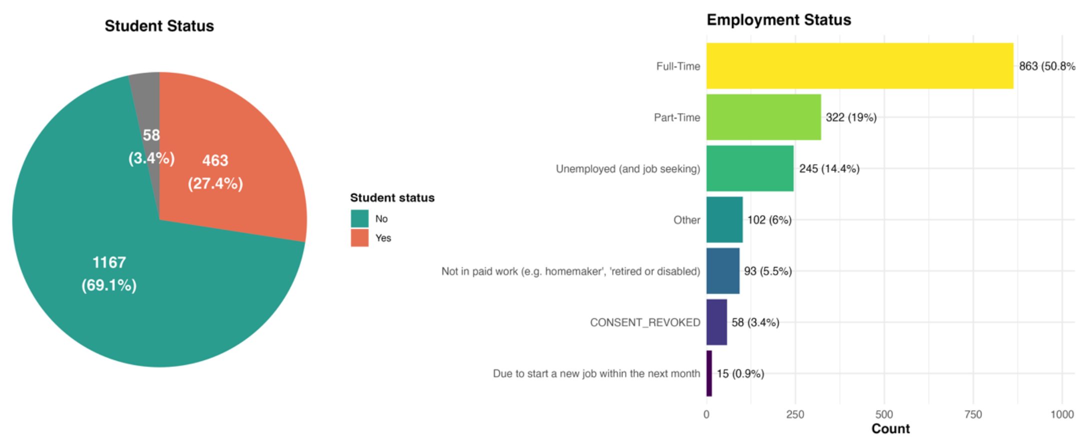

Figure S3. Student and employment status. Student status (69.1% non-students) and employment status; consent-revoked cases are shown separately and were excluded from analysis.

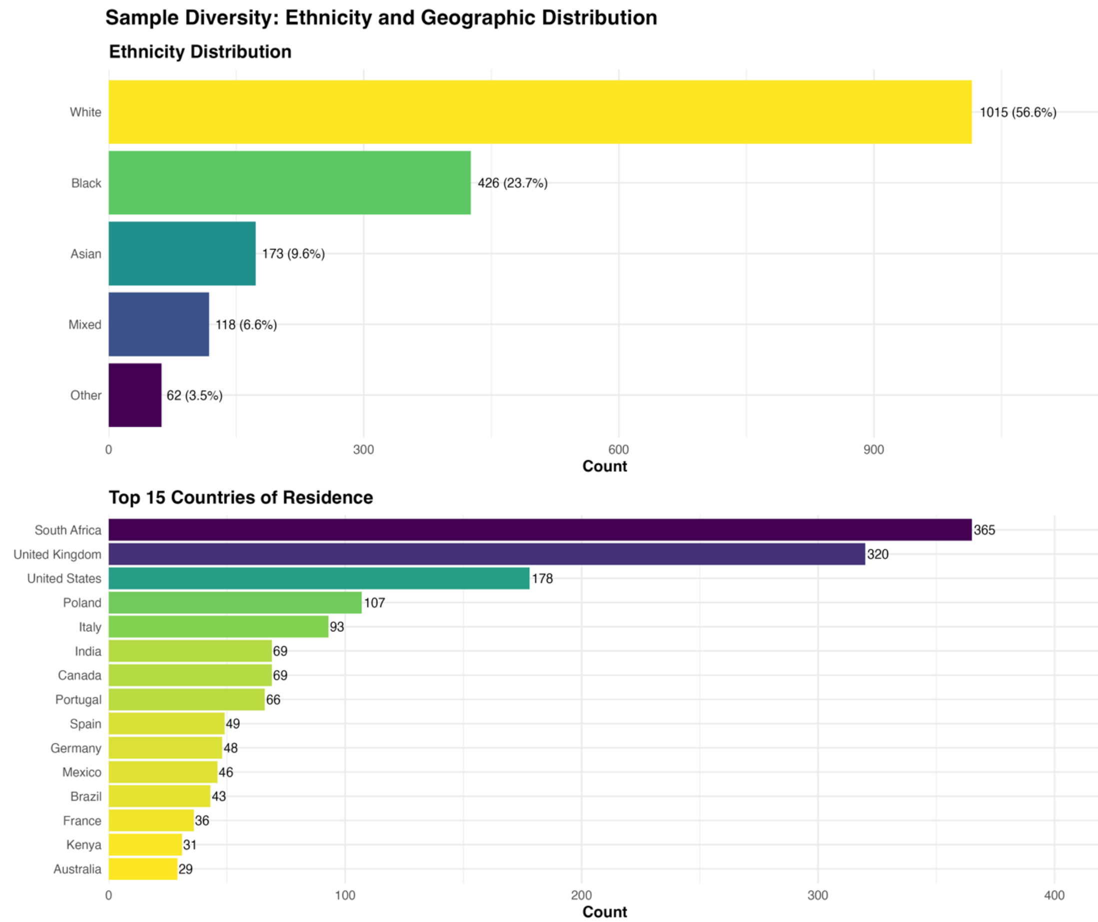

Figure S4. Sample diversity: ethnicity and country of residence. Self-reported ethnicity and the fifteen most frequent countries of residence.

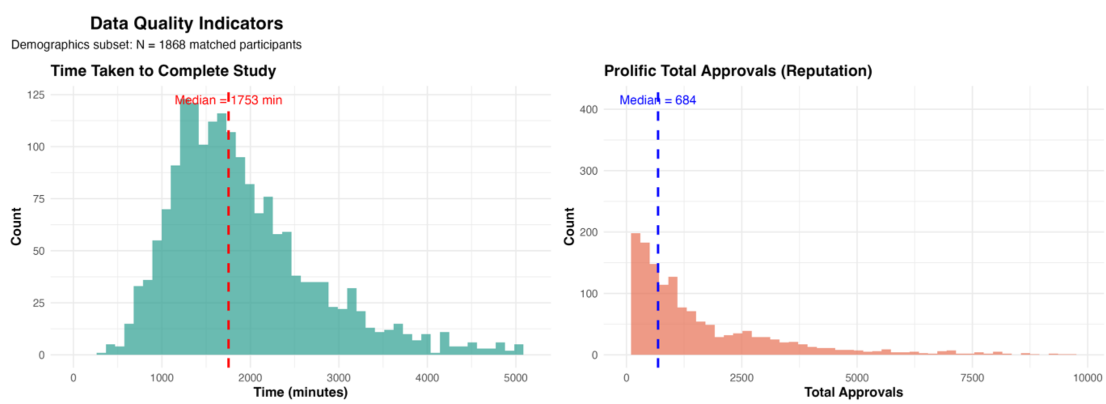

Figure S5. Data-quality indicators. Distribution of study-completion time (median = 1,753 min) and Prolific approval reputation (median = 684) for the matched demographic subset (N = 1,868).

Results

The sample was balanced by sex (50.2% female) and spanned a broad age range with a right-skewed distribution typical of online panels. Participants resided in 50 countries across both hemispheres, though with the expected over-representation of English-speaking and European countries (South Africa, United Kingdom, and United States were most frequent). Most participants were non-students in full- or part-time employment. Both data-quality indicators were consistent with attentive responding, and validation against published norms showed moderate-to-strong convergence within each rater-sex group.

Summary. The rating sample was demographically broad, internationally distributed across 50 countries and both hemispheres, and balanced by sex. Its age structure and residence distribution are typical of online crowdsourced panels, with some over-representation of Western, English-speaking participants that should be borne in mind when generalising cross-culturally. Data-quality indicators and gender-specific normative validation support the reliability of the retained ratings.

S2. Affective Rating Structure and Stimulus Composition

Analytic approach

Image-level mean ratings were computed for each of the six emotional dimensions (valence, arousal, awe, compassion, connectedness, gratitude). Their distributions and pairwise Pearson correlations were examined, and their conditional dependence structure was estimated as a regularised partial-correlation network (Gaussian Graphical Model with EBICglasso regularisation, $\gamma = 0.5$), with node centrality quantified by strength, closeness, betweenness, and expected influence. The composition of the stimulus set was summarised by content category and by source database, and rating profiles were compared across sources to assess heterogeneity.

Figures

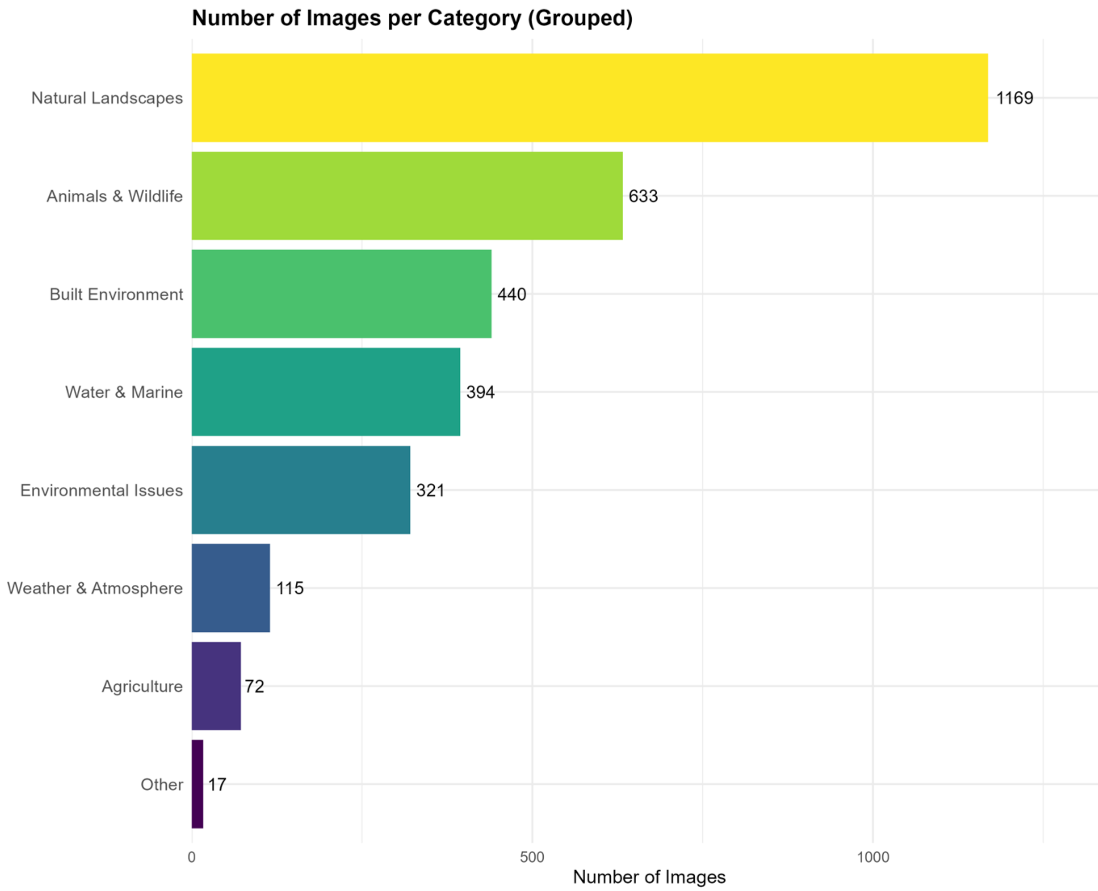

Figure S6. Number of images per grouped content category. Distribution of the 3,161 stimuli across the eight super-ordinate content categories, from Natural Landscapes (n = 1,169) to Other (n = 17).

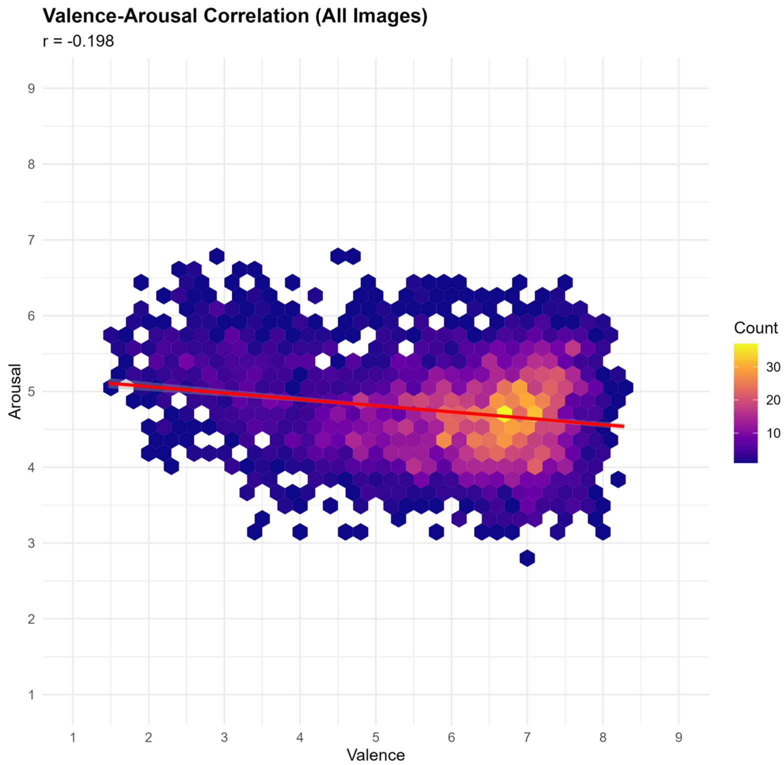

Figure S7. Valence–arousal density across all images. Hexbin density of image-level valence and arousal across all 3,161 stimuli (r = −.198).

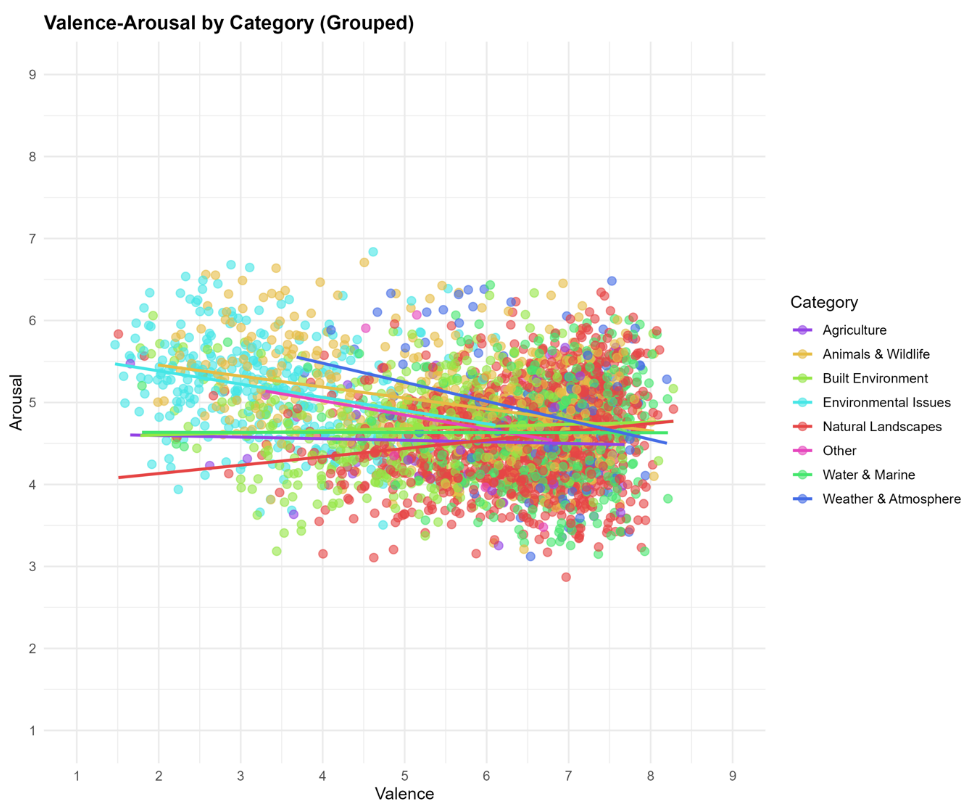

Figure S8. Valence–arousal associations by content category. Image-level valence and arousal coloured by content category, with category-specific regression lines.

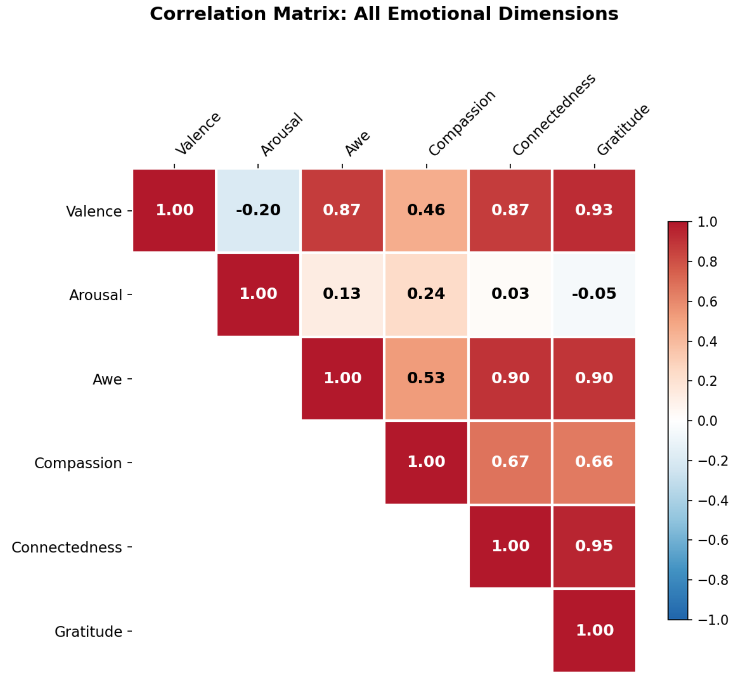

Figure S9. Correlation matrix of all emotional dimensions. Pearson correlations among image-level mean ratings for the six emotional dimensions.

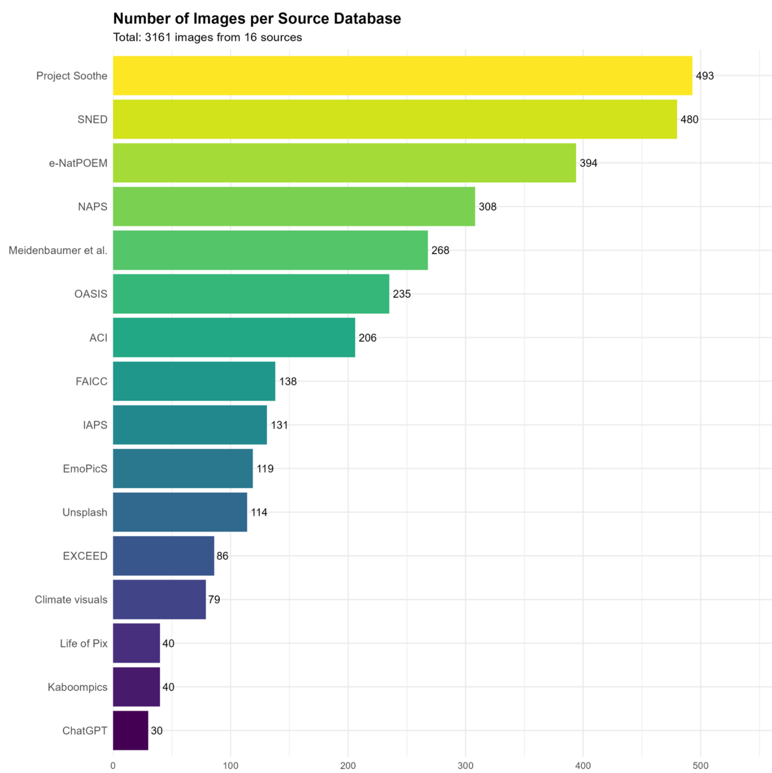

Figure S10. Number of images per source database. Composition of the atlas by source (16 databases, 3,161 images), from Project Soothe (n = 493) to ChatGPT-generated stimuli (n = 30).

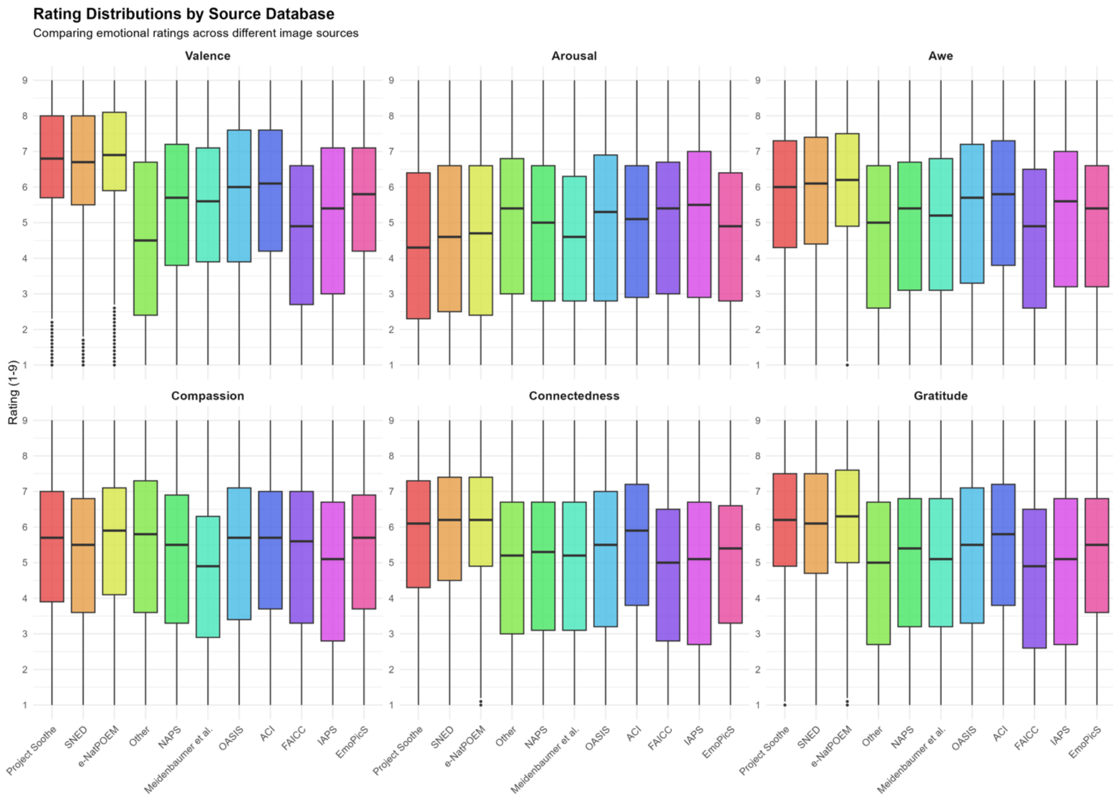

Figure S11. Rating distributions by source database. Boxplots of image-level mean ratings for each dimension, separated by source.

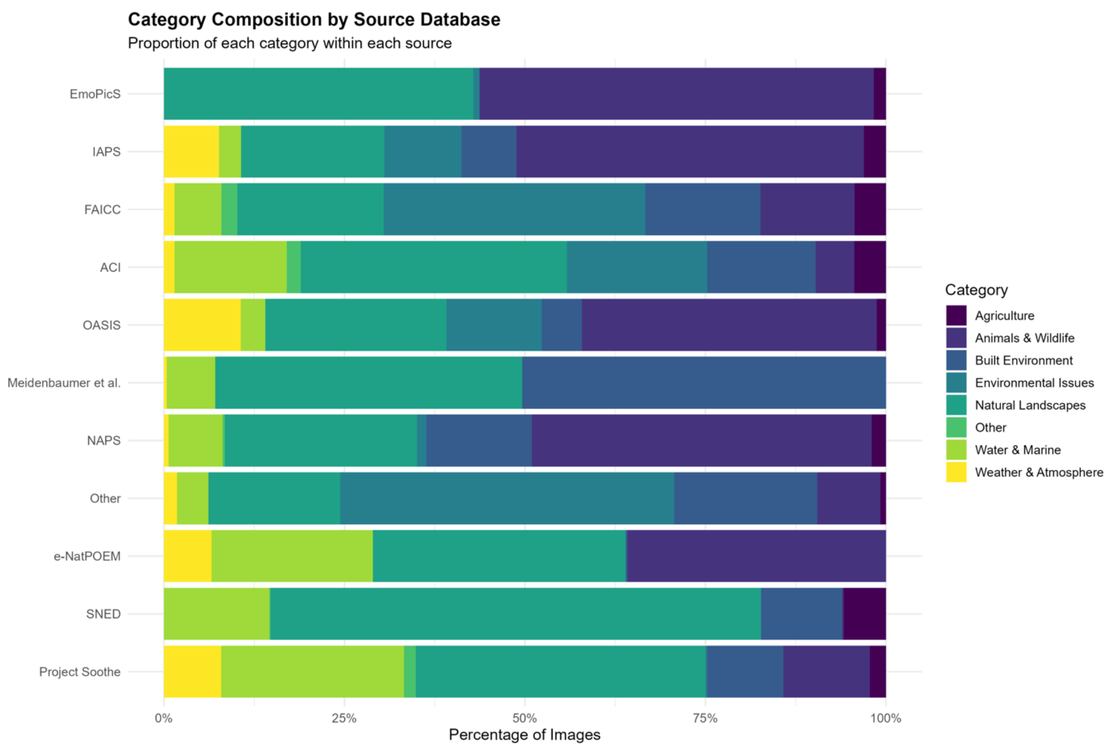

Figure S12. Category composition by source database. Proportion of each content category within each source database.

Figure S13. Mean ratings by source database and dimension. Heatmap of mean ratings across source databases for each emotional dimension.

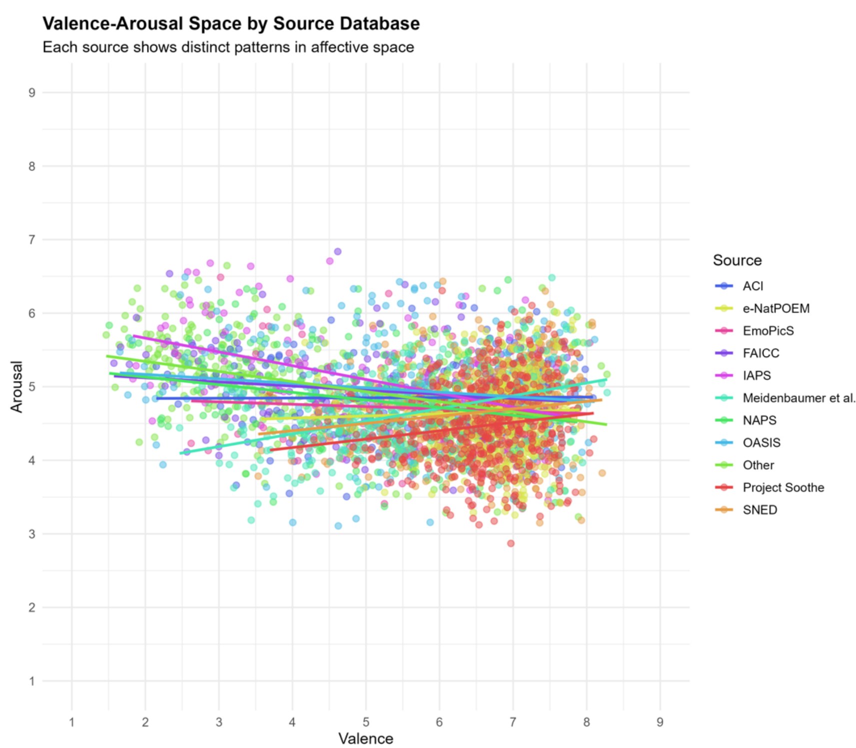

Figure S14. Valence–arousal space by source database. Image-level valence and arousal coloured by source, with source-specific regression lines.

Results

The six dimensions formed a coherent structure. Valence, awe, connectedness, and gratitude were strongly inter-correlated (r ≈ .87–.95), arousal was largely independent of this cluster (|r| ≤ .20), and compassion was only moderately related to it ($r \approx .46$–$.67$). In the partial-correlation network, gratitude and valence were the most central nodes (highest strength), whereas arousal and connectedness were peripheral; valence showed the highest betweenness, bridging the transcendence cluster and arousal (see Table S4). The atlas spanned eight content categories and sixteen source databases; rating ranges were broadly comparable across sources, with nature-focused sources (e.g., e-NatPOEM, SNED, Project Soothe) eliciting somewhat higher valence than conventional affective-image sets.

Summary. The affective ratings exhibit a robust and interpretable structure in which positive self-transcendent emotions cohere strongly, arousal operates as a largely independent axis, and compassion occupies a partially distinct position. This pattern is consistent across the correlation matrix, the partial-correlation network, and the dimensional-structure analyses reported in the main text. The stimulus set is broad and heterogeneous but internally comparable, with content and rating profiles that differ systematically yet predictably across contributing sources.

Table S3. Source databases

The 3,161 stimuli were compiled from sixteen source databases and image repositories. The table lists each source, the number of contributing images, and the corresponding citation where a peer-reviewed reference is available. Full download links and per-image provenance are provided in the E2IATLAS folder on the Open Science Framework.

Source	N	Citation
ChatGPT	30	AI-generated stimuli (see AI-generated images statement).
Unsplash	115	Unsplash. Royalty-free image platform (https://unsplash.com).
Kaboompics	40	Kaboompics. Royalty-free image platform (https://kaboompics.com).
Life of Pix	40	Life of Pix. Royalty-free image platform (https://www.lifeofpix.com).
Climate Visuals	79	Climate Visuals. Climate Outreach (https://www.climatevisuals.org).
EXCEED	86	Magalhães, S. d. S., Miranda, D. K., de Miranda, D. M., Malloy-Diniz, L. F., & Romano-Silva, M. A. (2018). The Extreme Climate Event Database (EXCEED): Development of a picture database composed of drought and flood stimuli. PLOS ONE, 13(9), e0204093.
ACI	207	Lehman, B., Thompson, J., Davis, S., & Carlson, J. M. (2019). Affective Images of Climate Change. Frontiers in Psychology, 10, 960.
FAICC	139	Ottavi, S., Roussel, S., & Syssau, A. (2021). The French Affective Images of Climate Change (FAICC): A dataset with relevance and affective ratings. Frontiers in Psychology, 12, 650650.
EmoPicS	119	Wessa, M., Kanske, P., Neumeister, P., Bode, K., Heissler, J., & Schönfelder, S. (2010). EmoPicS: Subjektive und psychophysiologische Evaluation neuen Bildmaterials für die klinisch-bio-psychologische Forschung. Zeitschrift für Klinische Psychologie und Psychotherapie, 39(Suppl. 1/11), 77.
IAPS	131	Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual (Technical Report A-8). University of Florida.
NAPS	308	Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. Behavior Research Methods, 46(2), 596–610.
OASIS	235	Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). Behavior Research Methods, 49(2), 457–470.
SNED	480	Bendall, R. C. A., Royle, S., Dodds, J., Watmough, H., Gillman, J. C., Beevers, D., … Gregory, S. E. A. (2024). The Salford Nature Environments Database (SNED): An open-access database of standardized high-quality pictures from natural environments. Behavior Research Methods, 57(1), 21.
e-NatPOEM	395	Dal Fabbro, D., Catissi, G., Borba, G., Lima, L., Hingst-Zaher, E., Rosa, J., … Leão, E. (2021). e-Nature Positive Emotions Photography Database (e-NatPOEM): Affectively rated nature images promoting positive emotions. Scientific Reports, 11(1), 11696.
Meidenbauer et al.	270	Meidenbauer, K. L., Stenfors, C. U. D., Bratman, G. N., Gross, J. J., Schertz, K. E., Choe, K. W., & Berman, M. G. (2020). The affective benefits of nature exposure: What's nature got to do with it? Journal of Environmental Psychology, 72, 101498.
Project Soothe	493	MacLennan, K., Schwannauer, M., McLaughlin, A. L., Allan, S., Blackwell, S. E., Ashworth, F., & Chan, S. W. (2024). Project Soothe: A pilot study evaluating the mood effects of soothing images collected using a citizen science approach. Wellcome Open Research, 8, 218.

Note. Standard references are given for databases with a peer-reviewed publication; image repositories (Unsplash, Kaboompics, Life of Pix, Climate Visuals) are cited by platform. IAPS is cited by its technical manual, as no open download source is currently maintained.

AI-generated images

Thirty stimuli were generated using GPT-4o + the OpenAI image generation model (03/2025). Images were manually screened for photorealism and thematic fit. AI-generated images were rated using the identical protocol and showed comparable rating profiles relative to photographic stimuli.

Table S4. Network centrality of emotional dimensions

Centrality indices from the regularised partial-correlation network (EBICglasso, $\gamma = 0.5$) over the six emotional dimensions. Strength is the sum of absolute edge weights connected to each node; closeness and betweenness index a node's position within the network; expected influence is the sum of signed edge weights, which is informative when negative edges are present.

Dimension	Strength	Closeness	Betweenness	Expected Influence
Gratitude	1.573	0.064	2	1.461
Valence	1.568	0.067	4	0.253
Awe	1.487	0.057	2	1.095
Compassion	1.292	0.050	0	0.395
Arousal	1.154	0.055	0	0.230
Connectedness	1.118	0.057	0	1.118

Note. Gratitude and valence are the most central (highest strength) dimensions, whereas arousal and connectedness are the most peripheral. Valence shows the highest betweenness, consistent with its bridging role between the positive-transcendence cluster and arousal.

S3. Convergence with Established Normative Databases

Analytic approach

For the subset of stimuli drawn from previously published normative image sets (N = 1,468 with available norms), E2IATLAS mean ratings were compared with the original normative values. Convergence was quantified with Pearson correlations, computed overall, separately for each source with at least 30 rated images, and separately for female and male raters within selected sources.

Figures

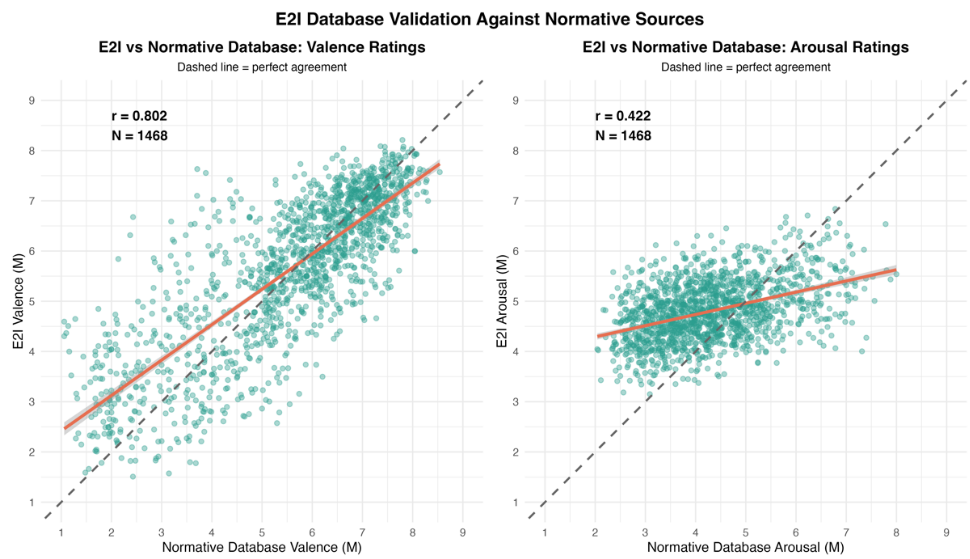

Figure S15. E2IATLAS validation against normative sources. E2IATLAS mean valence (left; r = .802) and arousal (right; r = .422) plotted against normative database values (N = 1,468). The dashed line indicates perfect agreement.

Figure S16. Arousal convergence with normative ratings by source. E2IATLAS versus normative arousal for each source with N ≥ 30 rated images.

Figure S17. Valence convergence with normative ratings by source. E2IATLAS versus normative valence for each source with N ≥ 30 rated images.

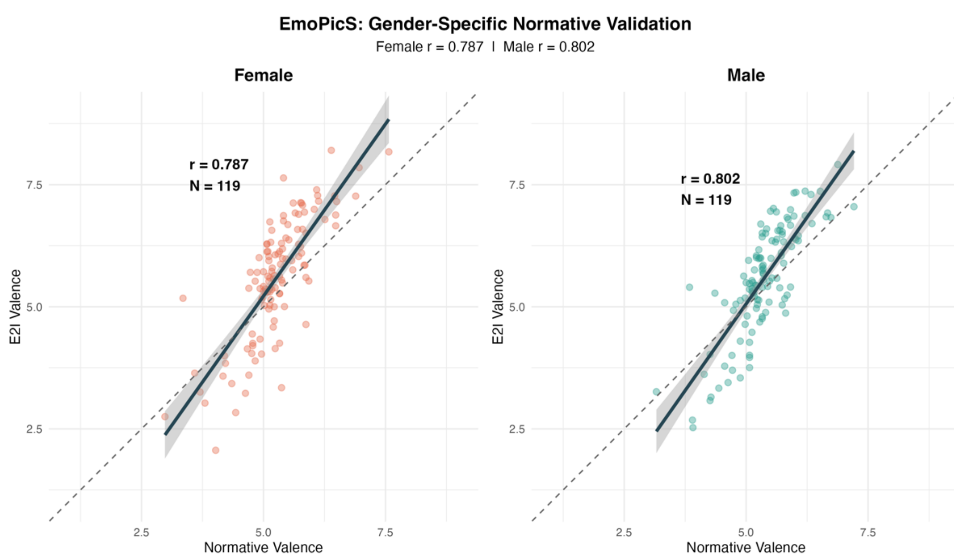

Figure S18. Gender-specific normative validation (EmoPicS). E2IATLAS versus normative valence for the EmoPicS subset, separately for female (r = .787) and male (r = .802) raters (N = 119).

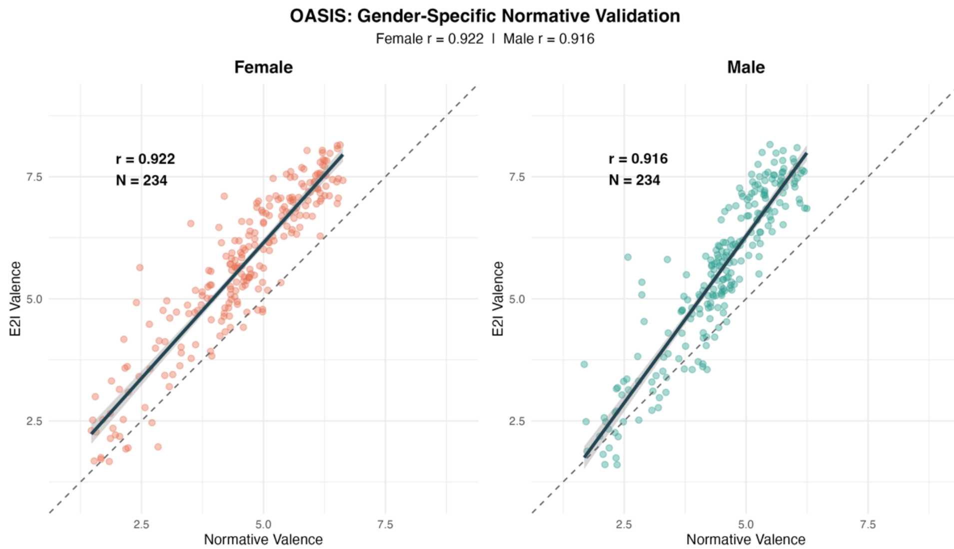

Figure S19. Gender-specific normative validation (OASIS). E2IATLAS versus normative valence for the OASIS subset, separately for female (r = .922) and male (r = .916) raters (N = 234).

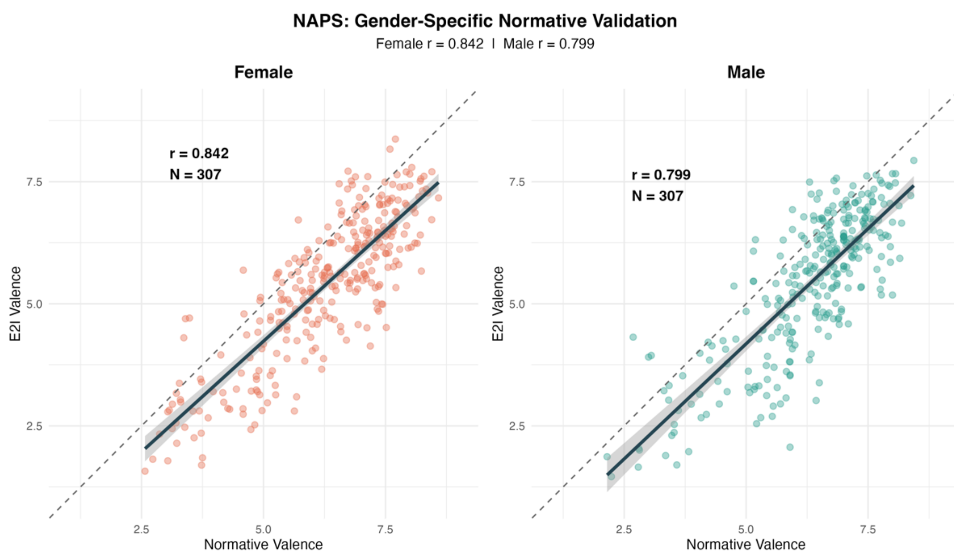

Figure S20. Gender-specific normative validation (NAPS). E2IATLAS versus normative valence for the NAPS subset, separately for female (r = .842) and male (r = .799) raters (N = 307).

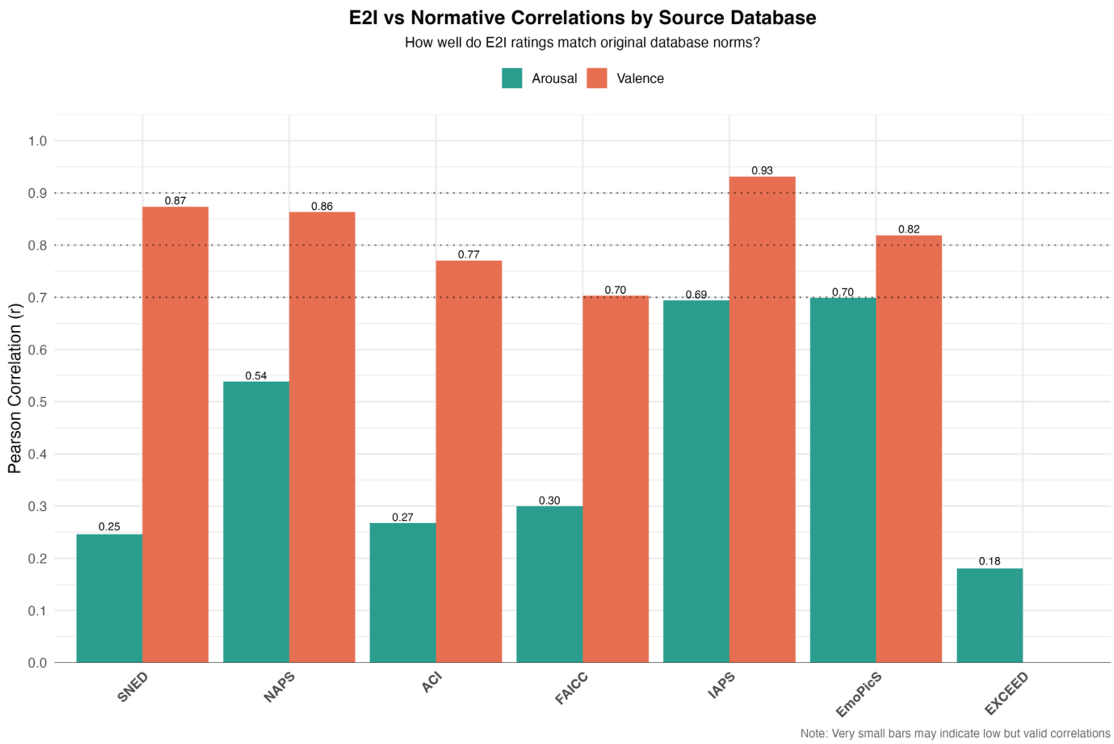

Figure S21. E2IATLAS–normative correlations by source database. Pearson correlations between E2IATLAS and original normative ratings for valence and arousal across source databases.

Results

E2IATLAS ratings converged strongly with the original norms for valence (overall r = .80; up to r ≈ .92 within individual sources such as OASIS) and consistently across female and male raters (all within-source r ≥ .79). Arousal convergence was lower and more variable across sources (overall r = .42), in line with the generally lower cross-sample reliability of arousal ratings. Convergence was strongest for conventional affective-image sets (e.g., IAPS, NAPS) and weaker for smaller or more specialised collections.

Summary. E2IATLAS reproduces the valence norms of its constituent databases with high fidelity and does so equivalently for male and female raters, supporting the validity of the re-collected ratings. The weaker and more variable arousal convergence mirrors a well-documented property of arousal ratings rather than a limitation specific to this atlas, and is consistent with the low image-level reliability of arousal reported in the psychometric analyses.

S4. Associations Among Extracted Image Features

Analytic approach

Low-level visual features (RGB channels, contrast, edge complexity, spectral power, saturation, colour temperature, symmetry) were extracted automatically for every image, alongside AI-derived semantic scene proportions. Pairwise Pearson correlations among these features, and between the features and the affective outcome dimensions, were computed to assess redundancy and their zero-order relationships with affect.

Figures

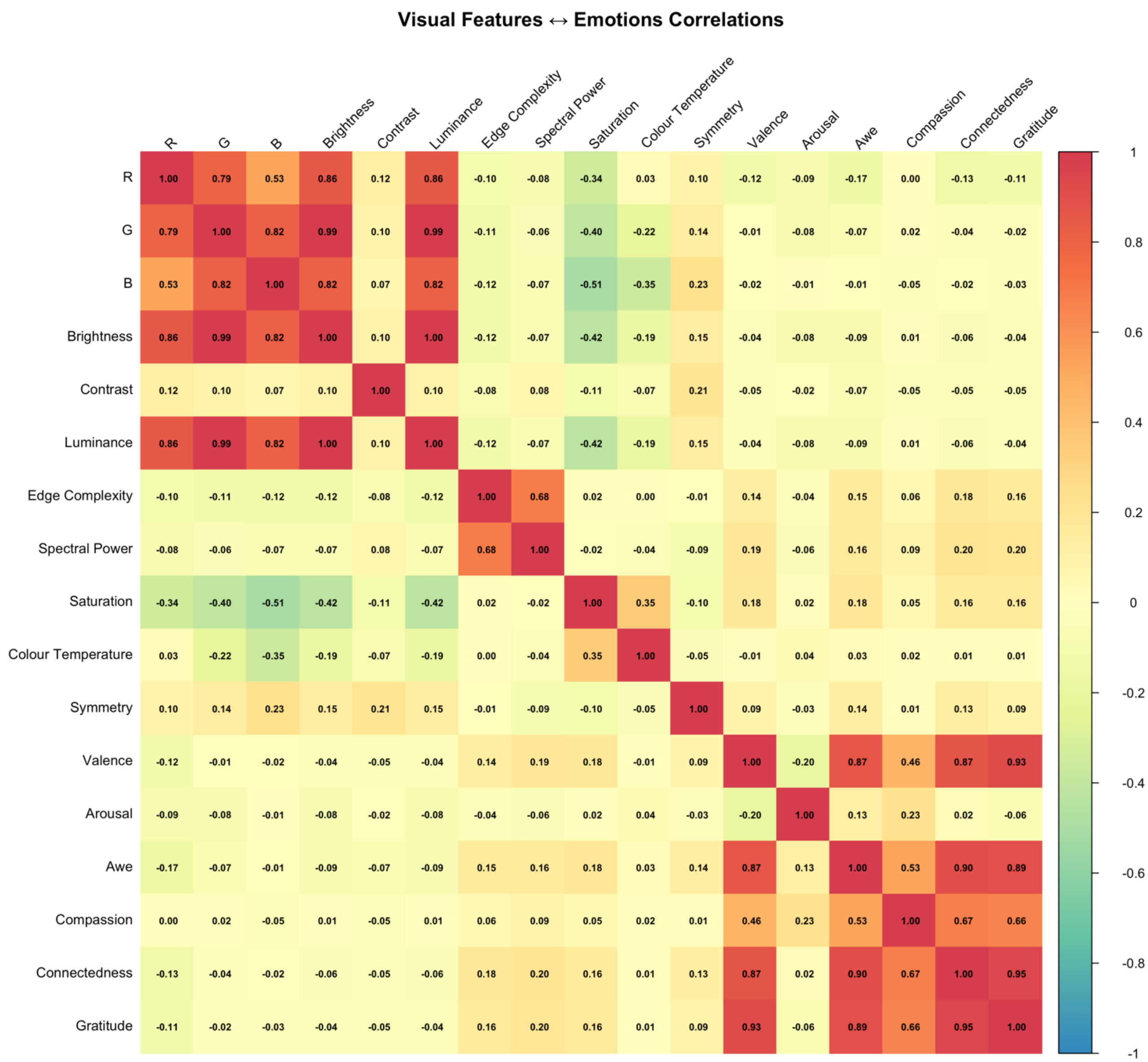

Figure S22. Correlations among visual features and affective outcomes. Pearson correlations of the low-level visual features with each other and with all six emotions.

Results

Feature inter-correlations were generally modest, with the notable exception of the expected collinearity among luminance-based descriptors (brightness, luminance, and the RGB channels). Individual features showed only weak zero-order associations with the affective outcomes, the strongest involving saturation and spectral power for valence. This collinearity among luminance-based features is the reason two of the eleven low-level features are dropped as rank-deficient in the dominance analysis reported in the main text.

> **Summary.** The extracted image features are largely non-redundant apart from expected collinearity among luminance-based descriptors, and individually explain little affective variance at the zero-order level. This reinforces the main-text conclusion that low-level image statistics alone are poor predictors of affective responses to nature imagery, and that semantic content carries most of the predictive signal.

S5. Predictive Modelling and Reliability

Analytic approach

Affective outcomes (valence, arousal, compassion, and the ST-Index) were predicted from the combined image-feature set using three algorithms: ordinary linear regression, elastic-net regression, and random forests. Models were evaluated with cross-validation and on a held-out 20% test set (R^2, RMSE, MAE); overfitting was indexed by the cross-validation-minus-test difference. Effect sizes of the human-coded scene features were estimated with one-way ANOVAs (η^2). Split-half reliability was computed by randomly partitioning raters into two halves, correlating image-level means across halves, and applying the Spearman-Brown correction.

Figures and tables

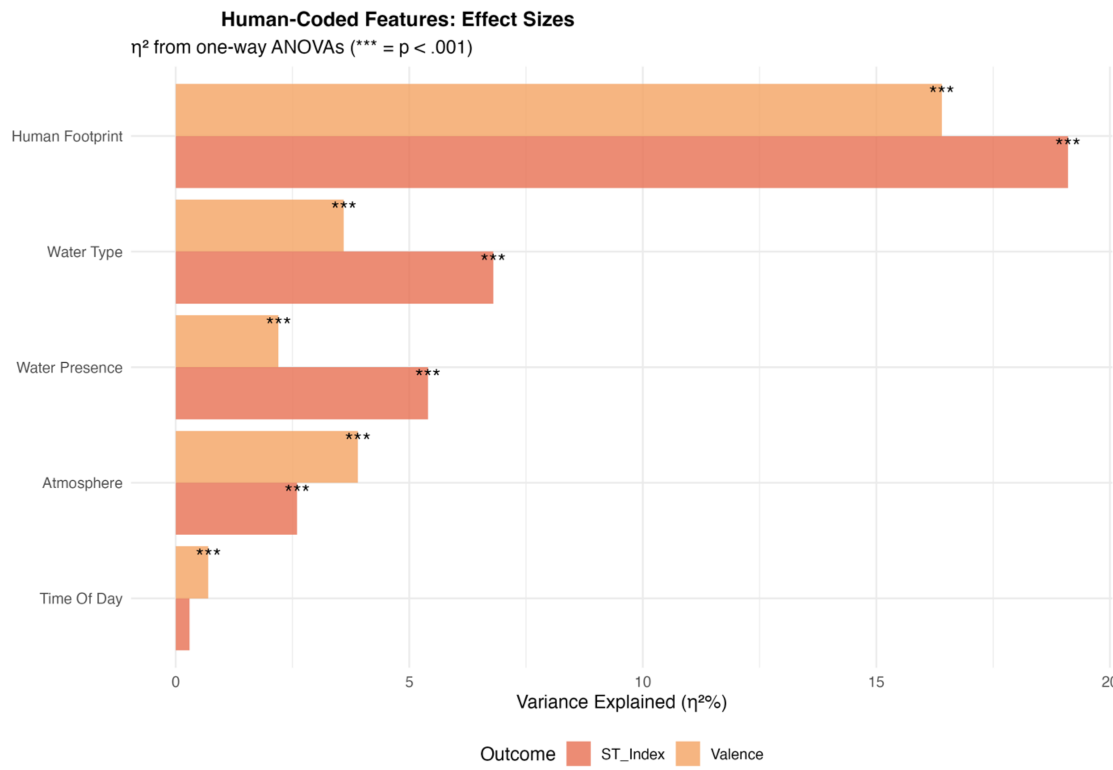

Figure S23. Effect sizes of human-coded scene features. Variance explained (η^2) by one-way ANOVAs of each human-coded scene feature on valence and the ST-Index (all $p < .001$).

Table S1. Split-half reliability

Split-half reliability by dimension: correlation of image-level means across two random rater halves (r_half) and its Spearman-Brown-corrected value (r_sb). N_A and N_B are the number of raters per half.

Outcome	r_half	r_sb	Split A M	Split A SD	Split B M	Split B SD	N_A	N_B
Valence	.965	.982	5.841	1.478	5.858	1.427	1583	1575
Arousal	.776	.874	4.754	0.625	4.742	0.634	1583	1575
Compassion	.903	.949	5.235	0.857	5.242	0.846	1583	1575
ST_Index	.936	.967	5.372	1.033	5.379	1.009	1583	1575

Table S2. Predictive model comparison

Cross-validated (CV_R²) and held-out test-set performance for each algorithm and outcome. Overfitting is the CV-minus-test difference.

Outcome	Model	CV_R²	Test_R²	Test_RMSE	Test_MAE	Overfit
Valence	Linear	.176	.189	1.336	1.047	−.013
Valence	Elastic Net	.181	.190	1.336	1.048	−.009
Valence	Random Forest	.268	.281	1.262	0.967	−.013
Arousal	Linear	.027	.024	0.629	0.501	.003
Arousal	Elastic Net	.033	.026	0.628	0.500	.007
Arousal	Random Forest	.058	.085	0.609	0.487	−.027
Compassion	Linear	.108	.092	0.813	0.636	.016
Compassion	Elastic Net	.109	.088	0.815	0.637	.021
Compassion	Random Forest	.137	.127	0.798	0.614	.010
ST_Index	Linear	.220	.248	0.906	0.720	−.028
ST_Index	Elastic Net	.224	.249	0.906	0.720	−.025
ST_Index	Random Forest	.293	.326	0.863	0.679	−.033

Results

Random forests consistently outperformed linear and elastic-net models, with held-out test R^2 of .28 (valence), .33 (ST-Index), .13 (compassion), and .09 (arousal), and negligible overfitting throughout. Split-half reliability was high to excellent for all outcomes (r_{sb} = .87–.98), with arousal the least reliable, mirroring its low image-level ICC. Among the human-coded scene features, human footprint produced the largest effect on self-transcendent responding.

Summary. The predictive analyses show that affective responses to nature imagery are systematically but only moderately predictable from image content, with non-linear models (random forests) capturing meaningfully more variance than linear approaches. Valence and self-transcendent responding are the most predictable outcomes and arousal the least, consistent with its low reliability. The prominence of human footprint among the human-coded features foreshadows the central role of human presence in shaping self-transcendent responses reported in the main text.

S6. Demographic, Geographic, and Climatic Moderators

Analytic approach

Moderators of affective responding were examined at the participant level. Associations with age were estimated as Pearson correlations, and hemisphere differences were tested as mean comparisons across dimensions. Cross-cultural divergence was summarised as the between-climate-zone variance in valence for each content category. In addition, exploratory mixed-effects models nested by image category examined standardised daylength as a predictor of valence, both as a main effect and in interaction with content category. Daylength was not measured directly but derived from each participant's latitude and data-collection season, so these daylength analyses are exploratory and should be interpreted with caution.

Figures

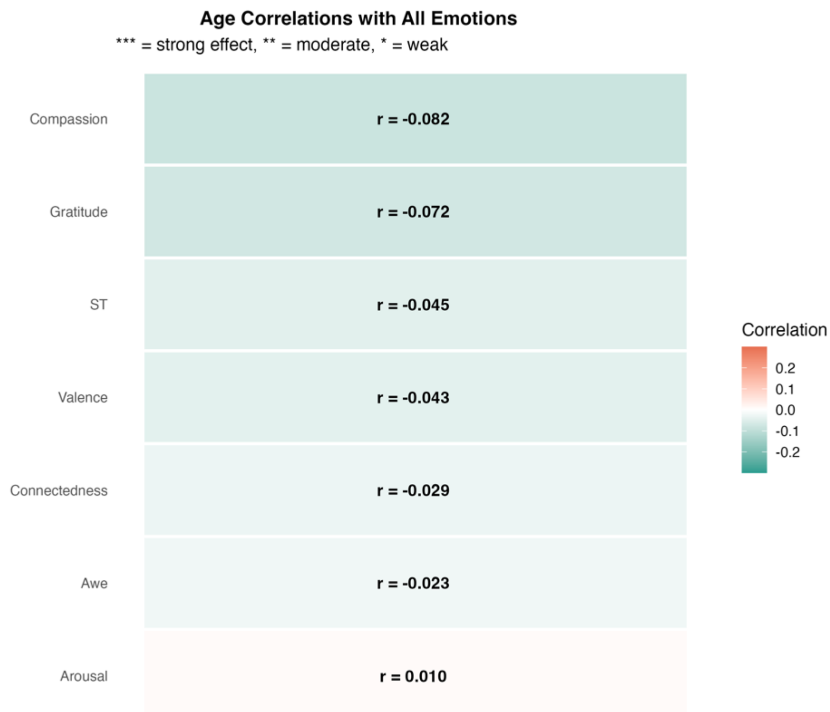

Figure S24. Age correlations with affective responding. Pearson correlations between participant age and mean ratings on each dimension; associations are weak (largest for compassion, r = −.082).

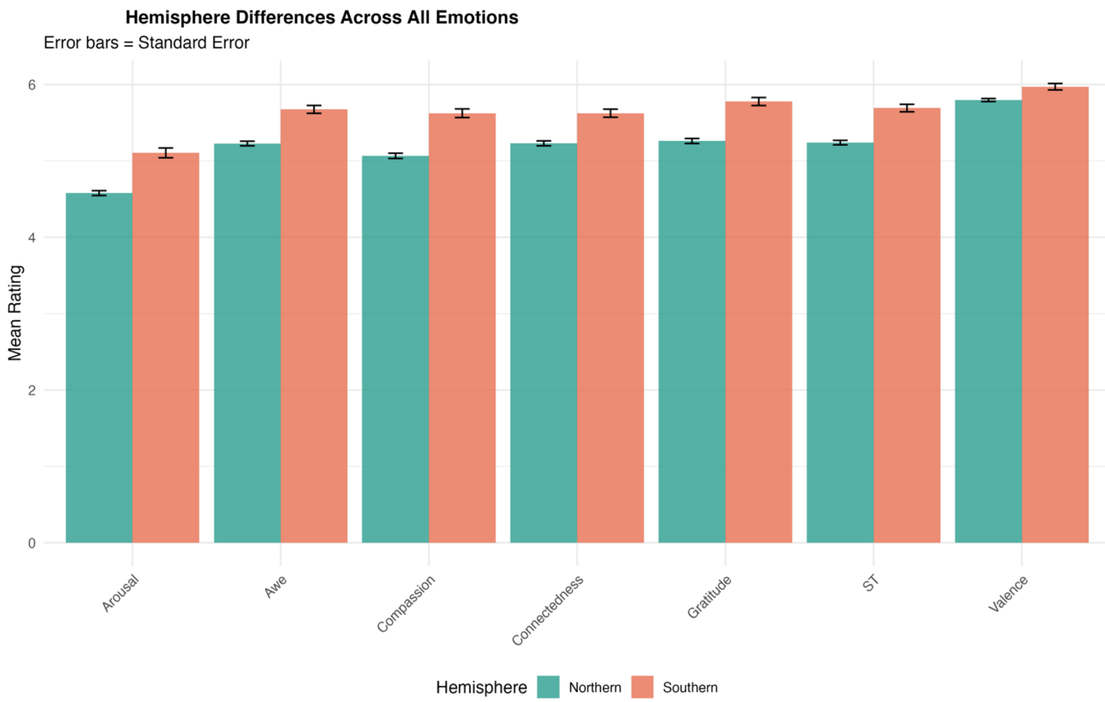

Figure S25. Hemisphere differences across dimensions. Mean ratings (± SE) for Northern versus Southern Hemisphere raters across all six dimensions.

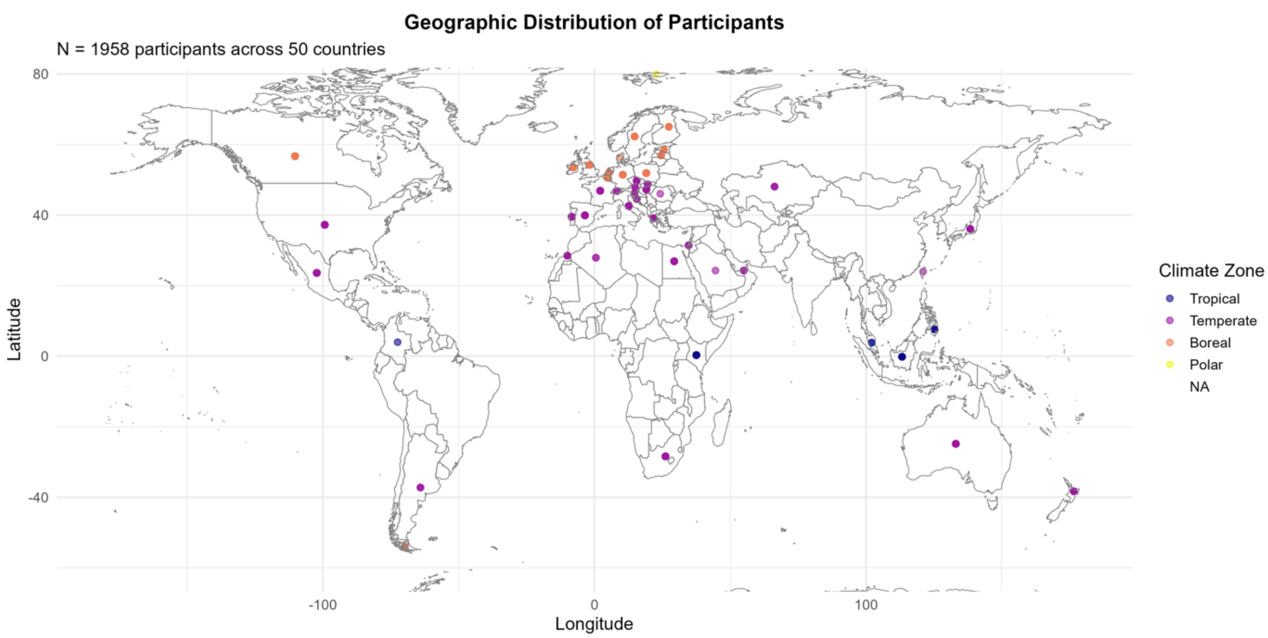

Figure S26. Geographic distribution of participants. World map of participant locations (N = 1,958 across 52 countries), coloured by climate zone.

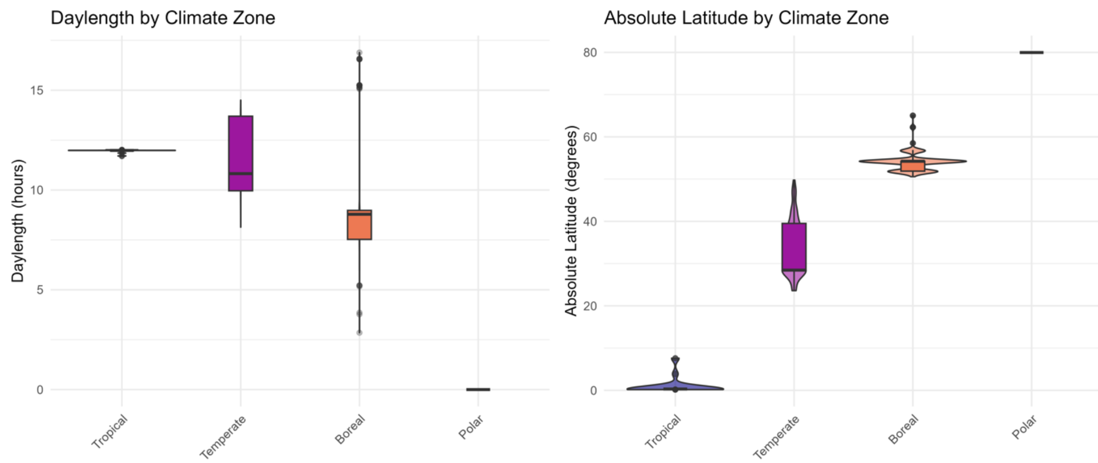

Figure S27. Daylength and latitude by climate zone. Distribution of daylength (left) and absolute latitude (right) across Köppen–Geiger climate zones. Exploratory: daylength was derived from participant latitude and data-collection season rather than measured directly.

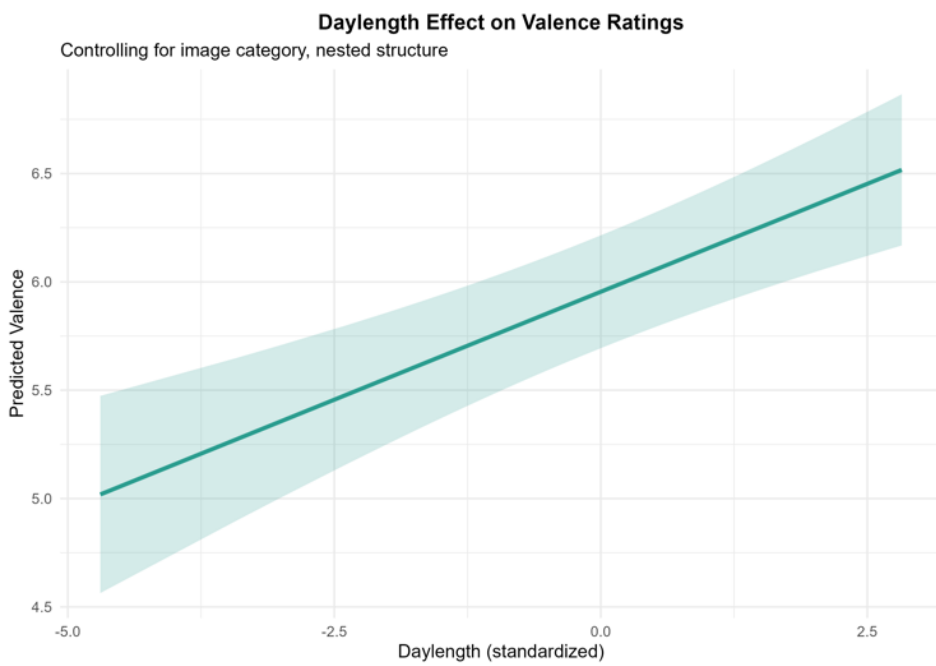

Figure S28. Daylength effect on valence (exploratory). Predicted valence as a function of standardised daylength, controlling for image category in a nested mixed-effects model. Daylength is a derived, post-hoc variable; this analysis is exploratory.

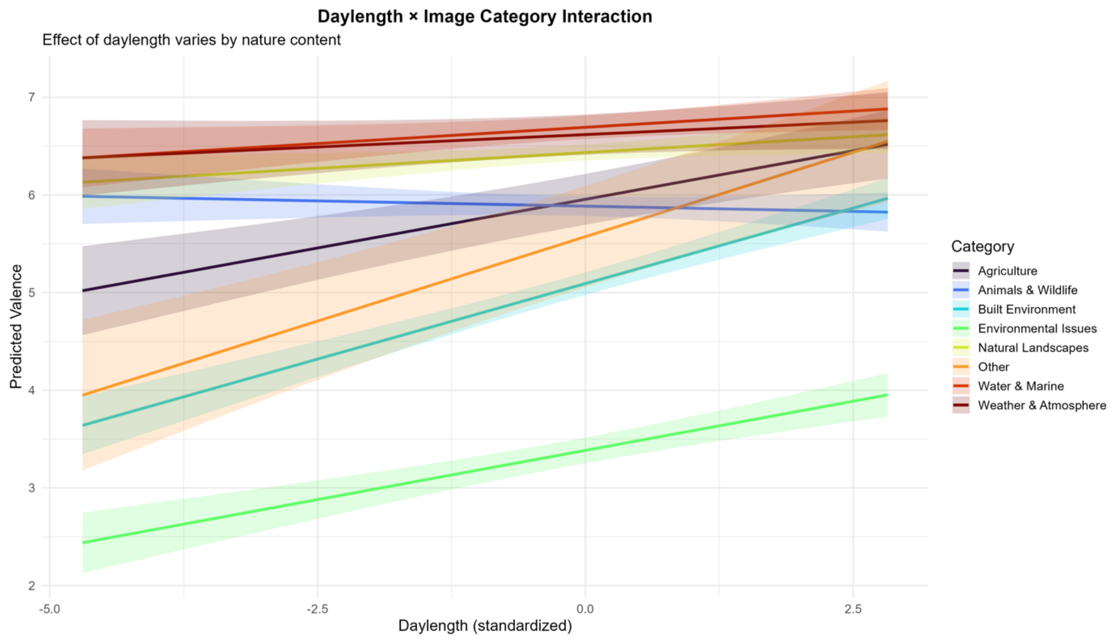

Figure S29. Daylength × content-category interaction (exploratory). Predicted valence as a function of daylength, separately by content category. Daylength is a derived, post-hoc variable; this analysis is exploratory.

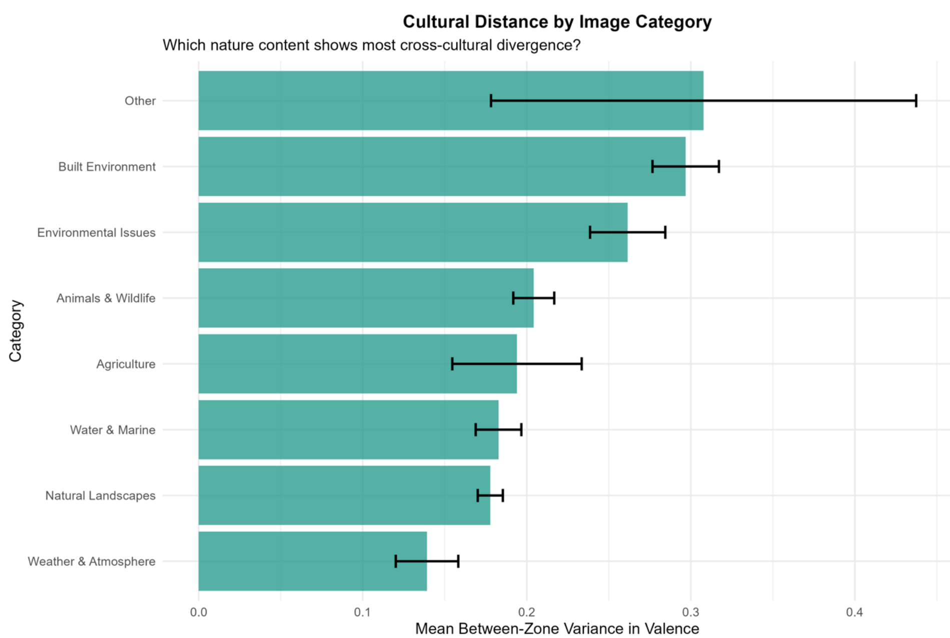

Figure S30. Cross-cultural divergence by content category. Mean between-zone variance in valence for each content category (± SE).

Results

Demographic moderators were weak: age correlated only trivially with each dimension ($|r| \leq .08$). Geographic context showed more systematic effects: Southern Hemisphere raters rated all six dimensions somewhat higher than Northern Hemisphere raters, with the largest gap for arousal. Cross-cultural divergence in valence was greatest for human-made and environmental-issue content and smallest for prototypical nature scenes. In the exploratory daylength analyses, longer daylength was associated with higher valence, and this association varied by content, being strongest for negatively valenced content

(Environmental Issues) and weakest for Animals & Wildlife; because daylength is a derived, post-hoc variable, these patterns are reported as exploratory.

Summary. Individual affective responses to nature imagery are only weakly shaped by demographic characteristics such as age, but are more systematically modulated by geographic context. Hemisphere and cross-cultural divergence account for reliable if modest variation, and the effects concentrate on content involving human presence or environmental threat rather than prototypical natural scenes. Exploratory analyses further suggest that derived daylength is associated with valence in a content-dependent way, though these post-hoc patterns require confirmatory testing. Together these findings motivate the region-stratified, cross-cultural experimental work outlined in the main discussion and underline the value of the atlas's broad geographic coverage.