Northwestern | INSTITUTE FOR POLICY RESEARCH

RESEARCH EXCELLENCE • POLICY IMPACT

IPR Working Paper Series

WP-25-30

Language Reveals Global Links Between Nature Attitudes and Sustainable Development

Tessa Charlesworth
Northwestern University and IPR

Leland Werden
ETH Zürich

Johan van den Hoogen
ETH Zürich

Madalina Vlasceanu
Stanford University

Thomas Lauber
Agroscope

Thomas Crowther

Version: November 11, 2025

DRAFT

Please do not quote or distribute without permission.

Abstract

Disentangling the cultural drivers of ecological degradation and recovery remains a central challenge for a regenerative future. Here, the authors use language to develop the first systematic record of global variation in nature attitudes and explore the implications for global environmental health. Using natural language processing (multilingual and contextualized word embeddings) they identify nature representations in 120 languages spoken across 189 countries. Starting with English, the authors find moderate associations of nature with importance, although this trend has increased over the last 200 years. Despite being the international standard language of environmental policy discussions, English expresses weaker nature-importance associations than 70% of other languages. In contrast, Afro-Asiatic languages, spoken in Global South nations, tend to express the strongest nature-importance associations. Critically, even after controlling for economic, linguistic, and attitudinal factors, the global variation of nature-importance associations in language robustly correlates with national-level environmental health, especially protection of water and land biodiversity areas.

Acknowledgements. Funding for this project was provided for TESC from the Roberta Buffett Institute; for TWC, TL, LKW, and JvdH through grants from DOB Ecology and the Bernina foundation; and for LKW through a Google carbon removal research award. All data and analyses are provided at the Open Science Framework: https://osf.io/hdxew/?view_only=d2ce89ee81874e9d87dcae201cf177bd (currently an anonymized view-only link that will be made public following peer review).

Language Reveals Global Links Between Nature Attitudes and Sustainable Development

Unsustainable development and environmental degradation are among the most pressing challenges of our time[1,2]. Environmental degradation not only harms natural biodiversity but also drives social inequality and burdens human wellbeing[3,4]. Understanding the complex social, economic, and political drivers of global degradation remains a central challenge in sustainability research. Although contemporary research suggests that economic factors play a prominent role, recent theoretical work suggests that cultural attitudes – how a country, on average, evaluates nature as good, important, worthy of concern, and so on – may present the foundation for anthropogenic impacts[5]. After all, although an individual person's attitudes clearly matter for individual behaviors like environmental donations[6–8], recycling or water usage[9], the sheer magnitude of environmental degradation is more likely to be the emergent property of collective attitudes.

Yet, until now, systematic data on collective attitudes towards nature has remained elusive at a global scale, precluding any exploration of geographic variation. Previous studies have focused largely on Global North countries[10–15] which underrepresent the true global range in natural environments, economic conditions and, presumably, in collective attitudes. Moreover, past research has relied entirely on self-reported survey measures, which are vulnerable to self-presentation and social desirability, rather than considering more objective or data-driven cultural measures of attitudes.[16]

In recent years, the rapid emergence of natural language processing (NLP) tools has opened up new avenues to overcome the practical challenges of collecting cultural attitudes at-scale. Today, NLP tools have already been used to study attitudes about race, gender, social class and many more as revealed through patterns of word co-occurrences in books, Internet text,

newspapers, and so on[16,17]. Indeed, attitudes captured in language correlate robustly with more traditional measures of implicit attitudes[18–20], including across countries[21] and across history[22,23]. To date, no study has extended such NLP applications to generate a systematic record of collective representations of nature across the globe.

Here, we use multilingual word embeddings trained on billions of words from historical text spanning 200 years (in English, French, and German), as well as contemporary text of 120 languages spoken across 189 countries. We first explore the dimensions that characterize nature representations in language, focusing on the association between *Nature* and *Importance,* alongside other dimensions (e.g., *Concern, Protection*). Next, we highlight considerable variation in the strength of these nature-importance attitudes across languages spoken around the world. Finally, we combine these attitudes with indicators of environmental health (vs. degradation), including from both remotely sensed (e.g., from satellite data) and self-reported sources (e.g., from UN Sustainable Development Goal reports). We also explore the relationship of environmental health with numerous other hypothesized factors including economic (e.g., Gini inequality), political (e.g., gender inequality), linguistic (e.g., size of speaker population), and attitudinal variables (e.g., other survey data).

Results and discussion.

Validating language approaches to nature attitudes. We begin by establishing that a language-based approach can be successfully applied to study the concept of *Nature,* as has been previously done for social group attitudes[22,23]. First, we tested face validity in the top 50 nearest neighbors (i.e., the words that had the highest associations or cosine similarities) to words representing *Nature* (e.g., *nature, climate, environment, land*). Nearest neighbors were clearly valid and centered on nature-related concepts, in both contemporary English Internet text using

840 billion words of contemporary English Internet text (from pre-trained GloVe word embeddings[24]), and in the 5 most widely spoken non-English languages of French, Spanish, Arabic, Bengali, and Chinese (from pre-trained *fastText* word embeddings on Wikipedia[25]). Robustness tests using shorter lists to represent *Nature* provided convergent conclusions (*SI Appendix*). Moreover, across all languages, exploratory factor analysis on these top 50 nearest neighbors indicated that the *Nature* concept consistently referred to similar and face-valid latent factors, always including factors referring to: (a) wilderness/ecosystems (e.g., *woodlands, jungle, savanna*); (b) agriculture/farming (e.g., *agriculture, husbandry, cultivation*); and (c) broader culture/human-nature interactions (e.g., *society, culture, policy*).

We then sought to test convergent validity by examining how the *Nature* concept was related to five semantic dimensions discussed in nature attitude research: *(1) positivity* and *(2) negativity*, the central dimensions of all evaluative attitudes[26]; *(3) concern* and *(4) protection,* two of the most widely-used dimensions in sustainability attitudes[27–29]*;* and (5) *importance,* less examined in sustainability attitudes but central in attitude theory, since greater perceived *importance* of a construct (e.g., nature) is a key predictor of attitude strength[30,31]. While we chose these dimensions as the starting point, the flexibility of the current methods will enable future work to test generalization to dozens of other dimensions proposed in the literature on environmental attitudes (e.g., *identification*, *enjoyment, trust*[29]).

As expected, we found that *Nature* (and its five bottom-up discovered factors from the exploratory factor analysis) were positively associated with all five top-down semantic dimensions (Figure 1B). However, the dimension of *Importance* appeared to be slightly stronger and more robustly associated than other dimensions (see also comparisons in *SI Appendix)*. Furthermore, in supplementary tests across 200 years of English books (using pre-trained

word2vec embeddings from Google Books across 1800-1990[32]), we found that the association of *Nature-Importance* has shown the largest increases over time, tripling in magnitude from essentially no association in 1800 $M = 0.04$ to an association of $M = 0.12$ in 1990 (slope of *rho* = .83, $p < .001$; Figure 1C). Although similar increases appeared for *Nature-Concern* and *-Protect*, they were at slower rates (*rho* = .76, .77, respectively, ps < .001), and there were no significant changes for *Nature-Positive* or *-Negative (rho* = -.17, -.36, respectively, ps > .12). Given the robustness, strength, and increasing association of *Nature-Importance* associations, we chose to focus on this dimension to streamline all subsequent reporting. Analyses for all dimensions are reported in the *SI*.

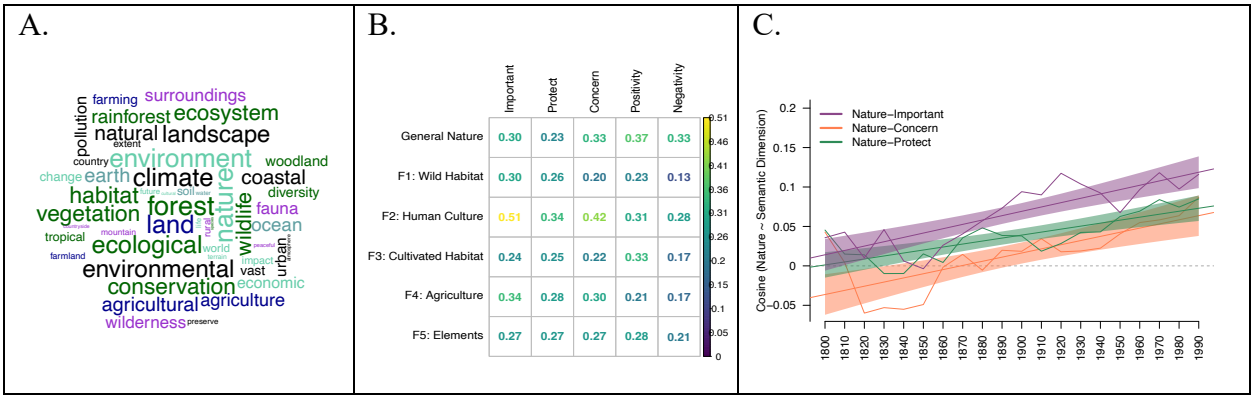

Figure 1. **The representation of "nature" in contemporary and historical English Internet text**. Panel A shows the top 50 words associated with the concept of Nature in contemporary English. Words are sized by the strength of their association, with stronger cosine similarities indicated by larger words; words are also colored by the factor that they uniquely loaded onto, above a factor weight of 0.4: dark green for wild habitat (e.g., habitat, vegetation, woodland); turquoise for culture (e.g., economic, nature, cultural); purple for more cultivated habitat (e.g., surroundings, fauna, countryside); dark blue for agriculture (e.g., agricultural, land); grey for elements (e.g., soil, ocean, earth); and black for those words that did pass a loading threshold of 0.4 (e.g., landscape, environmental). Panel B shows the average cosine similarity between the set of words representing the general Nature concept as well as 5 discovered factors (e.g., wilderness, pollution; see *SI* Appendix). Brighter yellow colors indicate higher relationships (correlations), with the highest relationships seen for the *Importance* dimension across the general concept and all subfactors. Panel C shows the increasing association of Nature-Importance (purple), as well as slightly slower but still increasing associations of Nature-Concern (orange) and Nature-Protect (green).

Variation in Nature-Importance Across 120 languages. To explore the generality of these trends, and search for variation across cultural backgrounds, we generalized these analyses to 120 languages using data from pre-trained multilingual *fastText* embeddings on Wikipedia text [25]. A language was included if it was (1) among the four most prevalent spoken languages in any

country around the world and (2) had available word embeddings data. For each language, we used automated translations from GPT4.0 for the word lists representing *Nature* and each of the five semantic dimensions. To ensure accurate translations, we also (1) collected professional translations from BLEND (https://www.getblend.com/) from native speakers of 84 (of 120) languages; and (2) back-translated all automated and professional translations manually in Google translate to identify any errors. Final word lists are provided in the open data on OSF.

Across all languages, we found that nature was moderately associated with *Importance* ($M = 0.27$, SD = 0.22). Critically, though, we found that the strength of *Nature-Importance* associations varied widely across languages (Figure 2). Compared to Indo-European languages, we found that *Nature-Importance* associations were significantly stronger in the Niger-Congo languages ($M = 0.65$), $\beta = 0.60$, $p < .001$, and Austronesian languages ($M = 0.40$), $\beta = 0.28$, $p = .001$ (full model results in *SI*), both language families that are predominantly spoken in the Global South. In contrast, English had relatively weak *Nature-Importance* associations (M = 0.16), lower than 71% of all other languages across our dataset.

Notably, we also ruled out the concern that stronger associations emerge only because of small sample sizes, whether smaller speaker populations (i.e., more rare languages) or sample sizes of text (i.e., smaller datasets in Wikipedia). In robustness analyses (*SI*) we show that *Nature-Importance* associations are not correlated with either (1) the number of speakers of a language, *rho* = -.04, $p = .69$; nor (2) sample size of Wikipedia text, *rho* = -.17, $p = .09$.

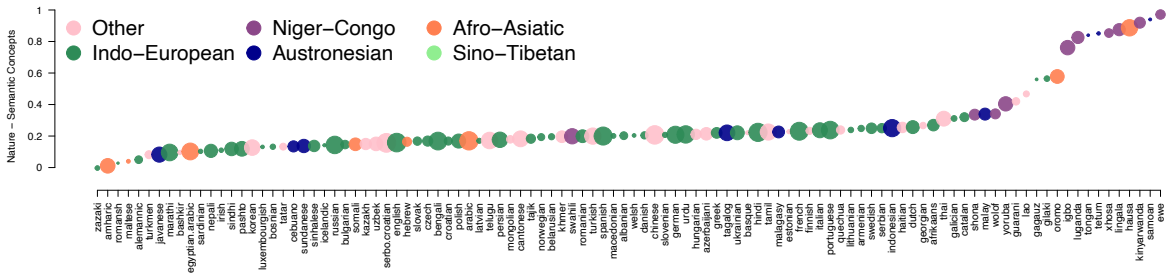

Figure 2. **Variation in Nature-Importance representations across language families.** Y-axis reports the cosine similarities (roughly equivalent to correlation scores) for each language on associations of *Nature-Importance*. Languages are ordered along the x-axis by their average cosine similarity, with languages showing the smallest associations on the left side of the plot and the largest associations on the right side of the plot. Languages are then colored by their language family as indicated in the legend. Sizes are the rank-ordered size of speaker population. Languages spoken by smaller populations (e.g., *Samoan, Romansch, Maori*) are smaller circles, while larger speaker populations (e.g., *English, Chinese, Serbo-Croatian)* are plotted with larger circles.

Variation in Nature-Importance Across 189 Countries. Drawing on this linguistic variation, we next created country-level *Nature-Importance* associations for 189 countries (all countries that spoke one of the 120 languages for which we had data). By averaging *Nature-Importance* associations for the first four primary languages of each country, weighted by the population speaking each of these four languages, we were able to generate weighted national average values for *Nature-Importance* scores (*Methods* and *SI Appendix*). We validated these country-level language-based results by showing that they were significantly and positively correlated with international survey data from nearly 60,000 respondents across 63 countries[33] (*SI Appendix*). Thus, again, language analyses can be used to expand the global map of nature representations.

Mirroring the results for language variation, country-level averages (Figure 3A) showed the strongest *Nature-Importance* associations in the Global South (including African countries like Niger, Nigeria, and Rwanda and some Oceanic countries like Samoa and Tonga). In fact, 53% of African countries in the data were in the top quartile of *Nature-Importance* associations. In contrast, only 18% of European, 15% of Asian, and zero North American countries were in the top quartile of *Nature-Importance* associations.

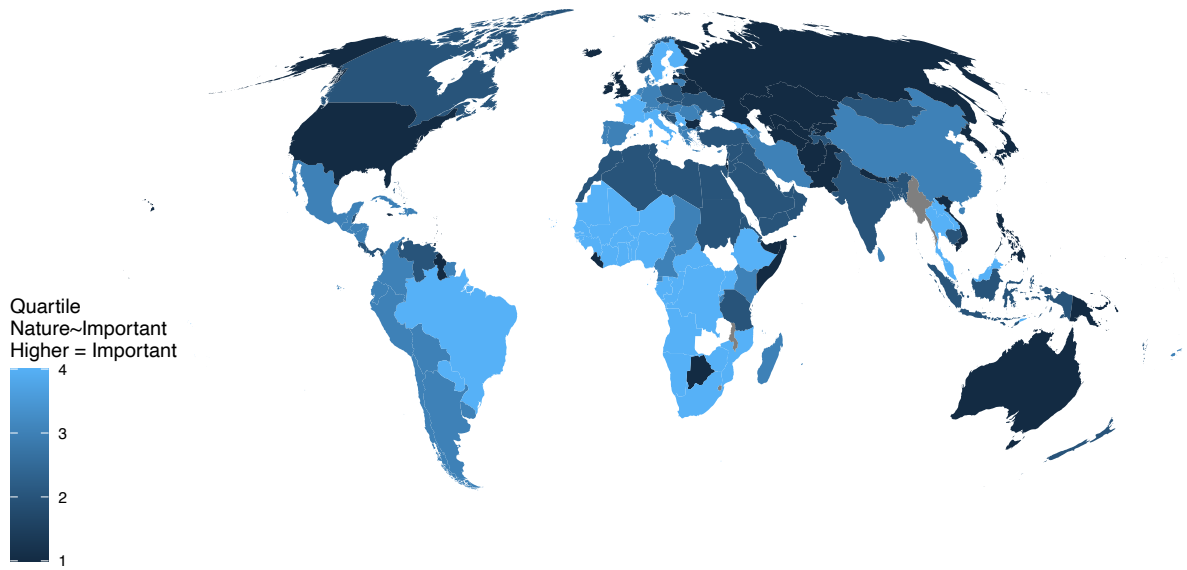

Figure 3. **Variation in Nature representations across countries.** Weighted estimates of *Nature-Importance* associations across 189 countries speaking at least one of 120 languages. Lighter blue indicates that the country is in the top quartile (strongest *Nature-Importance* associations), darker blue is the bottom quartile (weakest *Nature-Importance* associations). White indicates no language data was available for that country (N = 59 countries, largely island nations, out of all possible countries with geographic indicators), gray indicates that language data was available but not for the *Important* dimension (i.e., the importance synonyms were not available in the languages of those countries; N = 3 countries).

Country-level Nature-Importance and Environmental Health vs. Degradation. We used two sources of data on environmental health vs. degradation. First, we used data from the 2020 United Nations Sustainable Development Goals (SDGs) because they are the gold-standard in international policy discussions about the environment. We examined indicators for climate-relevant goals of SDG 14 (Life Below Water) and 15 (Life on Land). For model parsimony, we first identified that indicators of SDG15.1.2 (protection of freshwater and terrestrial biodiversity areas) were the most centrally connected to other indicators (i.e., it was the top-loading indicator on a first factor in Exploratory Factor Analysis). As such, we focus in the main text on the SDG15.1.2 and report all other indicators in the *SI.* To compliment these largely self-reported indicators, we also used a large dataset of 109 indicators compiled from remote-sensing satellite data including tree cover[34], forest biomass[35] and more (see *SI Appendix*). Here, we first identified

only those 13 indicators that showed significant bivariate correlations with *Nature-Importance* and then examined whether the relationships were robust to inclusion of covariates.

To control for other socio-economic factors, we examined six national-scale covariates that have been previously hypothesized to relate to global nature attitudes and nature impacts: (1) Gross Domestic Product (GDP), with theories suggesting both low GDP and very high GDP might explain positive nature attitudes because low GDP countries are before degradation and extraction economies, while high GDP countries have the resources to be post-industrial and divest from extraction[36–38]; (2) Gini inequality[39], for similar reasons as GDP; (3) Gender Inequality Index[40], a proxy indicator for a country's development in social issues, which may trade-off with environmental issues; as well as (4) Internet use[41]; (5) Wikipedia data size[42,43]; and (6) speaker population, with these last three ruling out methodological concerns that higher *Nature-Importance* associations are driven by small samples.

For the self-reported UN SDG data, we found small-moderate relationships between *Nature-Importance* and protection of freshwater biodiversity areas, $\beta = 0.24$, $t(137) = 2.91$, $p = .004$, as well as protection of terrestrial biodiversity areas, $\beta = 0.27$, $t(111) = 3.12$, $p = .002$. However, the direct remote-sensed measurements revealed strong and robust relationships with *Nature-Importance.* Specifically, the strongest correlations were observed with the estimated probability of an area experiencing a forest fire (essentially a proxy both for more forested areas, and more human-forest interaction[44]), with greater *Nature-Importance* scores associated with higher probability of potential fires, $\beta = 0.32$, $t(143) = 4.35$, $p < .001$. The second strongest predictor among remotely sensed indicators was biomass carbon storage (essentially a generalized indicator of ecosystem health[35]), which increased with *Nature-Importance* scores $\beta = 0.28$, $t(144) = 3.78$, $p < .001$. In sum, like the UN SDGs, these two strong relationships appear to

reflect direct indicators of environmental health (i.e., biomass and general forest/potential forest-human interaction). A further 11 remote-sensed indicators also showed persistent and significant relationships after covariates (as reported in *SI* and open data), with those indicators generally characterizing places with high forest cover, vegetation, lower silt soil content, and more stable temperatures (i.e., ecological features that largely characterize sub-Saharan Africa nations).

Comparing against covariate-only models showed that *Nature-Importance* explained an additional 4-9% of country-level variance in these four indicators (full models in *SI*). The magnitude of these effects was on-par with the contribution of Gini inequality, which was one of the few covariates that was usually (but not always) significant. Thus, attitudes of *Nature-Importance* may be at least as critical as economic inequality for our understanding of global variability in indicators of biodiversity protections and existing environmental health (e.g., biomass, burning probability). Of course, there were other environmental indicators and SDGs that did not correlate robustly with *Nature-Importance*; future research may explore whether other dimensions (e.g., *Nature-Protection*) or even specific factors of nature (e.g., *Nature* as *Wilderness vs. Agriculture*) help understand these other aspects of environmental health.

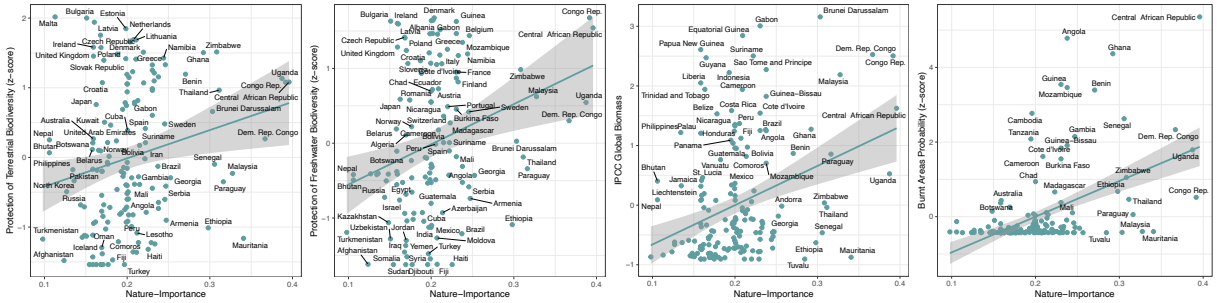

Figure 4. **Correlations of language representation (*Nature-Importance*) and indicators of sustainable development.** All plots have the same x-axis, indicating the *Nature-Importance* associations from language-weighted estimates across countries. Panels A and B show correlations with progress towards UN SDG15.1.2, normalized as z-scores, with higher scores indicating more progress towards the goal (i.e., more protected freshwater and terrestrial Key Biodiversity Areas). Panel C reports the correlation with the International Panel on Climate Change (IPCC) biomass carbon stored in above and belowground vegetation/soil (z-scored). Panel D reports the correlation with z-scored probability of burned areas in a region (larger in places with more forest cover and more human-nature interaction). Further details on the remote-sensed variables scoring, interpretation, and data sources are provided in the SI Appendix.

Conclusions, limitations, and future directions.

Here, we used a language-based approach to understand how humans represent nature, not only in the Western world today (as most previous studies have done[14,15]), but across global cultures and even back 200 years in history. In English, the concept of *Nature* appears robustly associated with *Importance*, alongside other dimensions that have been more widely studied in attitude and sustainability research (e.g., positivity/negativity, concern[28,29]). Yet, comparisons across 120 languages showed that English has one of the weakest *Nature-Importance* associations, especially lower than Niger-Congo and Austronesian languages spoken in the Global South. Given that most past research emphasizes English-speaking and WEIRD (Western, Educated, Industrialized, Rich, Democratic) nations[45–47], these findings suggest that past work has been missing the unique intrinsic value placed on nature in other countries and languages.

Indeed, global variation in *Nature-Importance* representations correlated significantly and positively with indicators of national environmental health, especially protections of biodiversity in land and water, and more biomass-rich and forested landscapes at risk of fire. In fact, although much theorizing has already considered the roles of economic and social factors (e.g., GDP, Gini, gender inequality) in environmental health[36,38], we show that collective attitudes of *Nature-Importance* add significant explanatory value beyond those past variables and at least on-par with economic inequality. Thus, even though we emphasize that the raw correlations and effect sizes may be small, even small relationships aggregated to a global scale across billions of people can help answer the puzzle of environmental degradation.

The present research faces limitations to be addressed in future work. Almost all non-English text data are currently available for a single year[25] limiting our ability to make inferences about the temporal ordering between cultural attitudes and environmental health. In all

likelihood, relationships are bidirectional – associating nature with importance may spur a country to introduce policies that protect nature, and protection of nature may feedback to increase the perceived of importance of nature. Efforts to collect, digitize, and train language models on historical text from multiple languages and countries will be necessary to tease apart these feedback loops.

Relatedly, we relied on the only large-scale cross-cultural texts available, which are static embeddings (i.e., one embedding per word) that do not capture nuanced contexts and polysemy of words. Although there are initial efforts at generalizing English contextualized embeddings (*BERT*[48]) and/or generative language models (e.g., *GPT*), there are concerns about their validity and representativeness of other languages[49]. Development and validation of non-English large language models will help generalize the current findings to new methods.

Finally, although we validate translations and explore bottom-up concepts across widely spoken languages (e.g., Arabic, Chinese), our choices of word lists and semantic dimensions may still miss emic perspectives from other cultures. Collaborative work with local communities and NGOs could benefit culture-specific understandings of nature representations globally.

By capitalizing on an unprecedented scale of standardized language data, the current work unlocks new insights into (1) how nature is represented in language, (2) how such representations vary across the globe, and the (3) the implications of such variation for environmental health and protection. Given the prominence of English within international environmental policy frameworks[1], the relatively low *Nature-Importance* associations in English and other "developed" nation languages is perhaps alarming. As such, addressing our global environmental challenges may benefit from integrating perspectives and collective attitudes of those regions where the importance of nature is already culturally embedded.

Materials and Methods.

Data sources.

Contemporary English Language. GloVe word embeddings[24] were trained on Common Crawl text, capturing a broad sweep of English Internet language (predominantly from Western countries). The underlying data comprise 840-billion-word tokens, with a vocabulary of approximately 2.2 million unique words. Note that, across all analyses, we show robust conclusions across not only the *GloVe* embedding algorithm but also *word2vec* and *fastText* algorithms (for historical and cross-cultural texts, respectively).

Although static embedding models are often seen to be "outdated" compared to contextualized transformer models (e.g., BERT[48]) and generative language models (e.g., GPT, Llama), static embeddings remain the best approach for our current needs of data across histories and languages. This is because static embeddings have already been pre-trained and validated on historical and cross-lingual databases (outlined below), while no such pre-trained databases are available for transformer models across history, languages, or geographics. Thus, by using static embeddings for all analyses, we do not expend the substantial environmental and energy costs [50] for scraping and preparing new data and fine-tuning/training new transformer models.

Historical English Language. Pre-trained historical word embeddings were obtained from *Histwords*[51], which are *word2vec* embeddings trained on English Google Books data from 1800-1990. These historical data have already shown robust internal validity (e.g., they capture semantic shifts in words like *gay* or *broadcast*) as well as external validity, aligning with real-world events (e.g., the women's movement altered gender stereotypes in the language[22,23]).

Non-English Languages. *FastText* word embeddings were pre-trained on over 200 different languages with available text on Wikipedia[25,52]. All text was scraped around 2014, providing an international snapshot of relatively contemporary language representations. We used word embeddings from a subset of 120 languages, since many of the >200 available languages (e.g., *Aragonese, Assamese*) do not have a geographic match linked to country-level ecological health (see below on the approach of language-country matching).

Ecological Health indicators. Measures of ecological health were operationalized using two approaches. First, the UN Sustainable Development Goals (SDGs), which are used for international benchmarking and policy making but have the disadvantage of providing data that is normalized (i.e., not the raw results) and sometimes self-reported by the nation itself. Second, we addressed these concerns using a compilation of >100 geospatial indicators including tree cover, forest biomass, terrain ruggedness, and more (see *SI Appendix,* open data on OSF).

For the UN SDGs, we extracted data from the 194 UN member countries covered in the UN SDG 2020 report. Our focus was on environment-related SDGs, specifically SDG14 (Life Below Water); and SDG15 (Life On Land). All indicators are outlined in the open data codebook provided on OSF. Notably, sustainable development encompasses not only the environment, but also political, social, and economic outcomes as well. As such, in exploratory analyses reported in the *SI Appendix*, we also consider all other available UN SDG indicators, encompassing 83 indicators across the 17 UN SDGs.

Data preparation.

Extracting representations of nature from language. To identify the representations of nature in contemporary and historical English, and non-English languages, we first created lists of target words related to the concept of *Nature*: *nature, climate, environment, land, forest; SI*

Appendix reports robustness checks using other word lists. We also created word lists of five key semantic dimensions that have been identified as meaningful in other studies of nature-related attitudes: *Positivity, Negativity, Concern, Importance,* and *Protection*, each represented by a set of synonyms (e.g., *Importance = important, importance, significant, meaningful;* see *SI Appendix* for all word lists).

For non-English analyses, these target words for *Nature* and the semantic dimensions were translated into each of the 120 languages using GPT-4.0 as an automated dictionary (following similar approaches validated elsewhere[19]). Here, we also validated these automated translations with professional translations from native speakers on BLEND.com. Any inconsistencies were flagged and then checked using back-translations in Google translate and consulting native speakers. Professional translations and automated translations showed generally high agreement, especially on the central concepts of *Nature* and *Importance*, with most inconsistencies involving synonyms of the *Positivity/Negativity* concepts. We therefore kept the automated word lists as the primary source. This choice helps ensure a generalizable and automated pipeline (including for researchers without resources to pay for professional translations) and to ensure comparability across all languages (including those without native-speaker translations).

Next, using these lists, we computed the mean average cosine (MAC) similarity[22] between the *Nature* words and the *Importance* words and separately also for all other dimensions. Cosine similarity is essentially a measure of how correlated two word-vectors are in the word embedding space. Higher cosine similarities indicate that *Nature* and *Importance* are highly semantically related and frequently used in similar contexts; low cosine similarities indicate that the concepts are unrelated and rarely used in similar contexts.

Note that, in pilot analyses, we also assessed the bottom-up associated concepts with *Nature* in English and the five most widely-spoken non-English languages, without any top-down researcher constraints about the expected dimensions: we computed the nearest neighbor words (from all possible words) that had the highest average cosine similarities to the *Nature* concept. These bottom-up explorations generally revealed face-valid and consistent neighborhoods of words with meaningful latent factors, implying that the representation of *Nature* can indeed be validly extracted from both English and non-English language corpora.

Geo-locating and weighting non-English language representations. Linking languages to geographic locations is non-trivial: many countries have multiple official and non-official languages (e.g., Switzerland has four languages), and many languages are spoken across multiple countries (e.g., English is spoken around the world). Although there are some databases (e.g., Ethnologue) that provide geo-tagged language codes, these codes are limited and largely focused on Indigenous languages (e.g., English is not coded as spoken in Canada in Ethnologue), and do not provide information on the number of speakers. Such data is essential because we are interested in the collective representation of a country, which will inevitably be weighted not towards rare languages but towards a country's more commonly spoken languages that pervade cultural products.

As such, to match languages with countries, we introduce a new three step approach. First, we began with the list of 204 countries with available outcome data from the UN SDGs and remote-sensed variables and created a database of the primary 1-4 languages spoken in each country as well as the proportion of the population speaking each language (collected through GPT4.0 and checked against country-level censuses). Note that the proportion of a population speaking a fifth or further language was usually <0.01, implying that the top 1-4 languages

provide a near-complete linguistic landscape of each country. This approach yielded 306 languages total, of which 134 had available pretrained word embeddings from Wikipedia text (Grave et al., 2018), and a final 120 languages had sufficient vocabulary (i.e., included the necessary *Nature* synonyms). Critically, these 120 languages were generally the most common languages spoken worldwide and usually represented the first and second languages spoken in most countries.

Next, for each of these final 120 languages we calculated the average cosine similarity of *Nature* with *Importance* (and other semantic dimensions, as described above). Third, for each country representation, we computed a weighted mean across the primary languages spoken in the country, using weights from the population proportion speaking each of the primary languages. For example, in Switzerland, the final language representation is a weighted average of the *Nature-Importance* association from German (weighted by 0.63, reflecting that 63% of the population speaks German), French (0.23), Italian (0.08), and Romansch (0.01). Thus, Switzerland representations end up being different from those of other German-speaking countries (e.g., Austria, where 98% of the population speaks German, and 2% speaks Turkish).

This weighting approach still faces limitations. For example, the estimated weighted language representation for Chile (99% Spanish-speaking) is very similar to the weighted language representation for Spain (99% Spanish-speaking), even though these countries are on different continents with different dialects, histories, and climates. Nevertheless, the proof-of-concept results reported here can help motivate future work to conduct more granular geo-tagged analyses (e.g., by training newspapers, social media data) that may better approximate regional within-country differences.

Code and Data availability statement
All data and code to reproduce the results presented in the paper are present in the supplementary materials and/or on the Open Science Framework (https://osf.io/hdxew/?view_only=d2ce89ee81874e9d87dcae201cf177bd (currently an anonymized view-only link that will be made public following peer review).

Competing interest statement
The authors declare that they have no competing interests.

Author contributions.
TESC, LKW, and TWC conceptualized the research; TESC, JvdH, and TL collected and prepared data for analysis; TESC performed analyses and drafted the initial manuscript; all authors revised and approved the final manuscript.

References

1. Lee, H., Romero, J. & IPCC. *IPCC 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. (2023) doi:10.59327/IPCC/AR6-9789291691647.
2. Mo, L. *et al.* Integrated global assessment of the natural forest carbon potential. *Nature 2023 624:7990* **624**, 92–101 (2023).
3. Cushing, L., Morello-Frosch, R., Wander, M. & Pastor, M. The Haves, the Have-Nots, and the Health of Everyone: The Relationship Between Social Inequality and Environmental Quality. *Annu. Rev. Public Health* **17**, 32 (2024).
4. Morello-Frosch, R. & Lopez, R. The riskscape and the color line: Examining the role of segregation in environmental health disparities. *Environ Res* **102**, 181–196 (2006).
5. Climate change and human behaviour. *Nature Human Behaviour 2022 6:11* **6**, 1441–1442 (2022).
6. Klöckner, C. A. A comprehensive model of the psychology of environmental behaviour—A meta-analysis. *Global Environmental Change* **23**, 1028–1038 (2013).
7. Busch, J. & Ferretti-Gallon, K. What drives deforestation and what stops it? A meta-analysis. *Rev Environ Econ Policy* **11**, 3–23 (2017).
8. Bamberg, S. & Möser, G. Twenty years after Hines, Hungerford, and Tomera: A new meta-analysis of psycho-social determinants of pro-environmental behaviour. *J Environ Psychol* **27**, 14–25 (2007).
9. Kormos, C. & Gifford, R. The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *J Environ Psychol* **40**, 359–371 (2014).
10. Schultz, P. W. Environmental Attitudes and Behaviors Across Cultures. *Online Readings in Psychology and Culture* **8**, 4 (2002).
11. Ihemezie, E. J., Nawrath, M., Straußß, L., Stringer, L. C. & Dallimer, M. The influence of human values on attitudes and behaviours towards forest conservation. *J Environ Manage* **292**, 112857 (2021).

12. Freymeyer, R. H. & Johnson, B. E. A cross-cultural investigation of factors influencing environmental actions. *Sociological Spectrum* **30**, 184–195 (2010).
13. Mede, N. G. *et al.* Perceptions of Science, Science Communication, and Climate Change Attitudes in 68 Countries: The TISP Dataset. *PsyArXiv Preprints* https://doi.org/10.31234/OSF.IO/JKTSY (2024) doi:10.31234/OSF.IO/JKTSY.
14. Bauer, N., Wallner, A. & Hunziker, M. The change of European landscapes: Human-nature relationships, public attitudes towards rewilding, and the implications for landscape management in Switzerland. *J Environ Manage* **90**, 2910–2920 (2009).
15. Howe, P. D., Mildenberger, M., Marlon, J. R. & Leiserowitz, A. Geographic variation in opinions on climate change at state and local scales in the USA. *Nature Climate Change 2014 5:6* **5**, 596–603 (2015).
16. Charlesworth, T. E. S., Morehouse, K., Rouduri, V. & Cunningham, W. A. Echoes of Culture: Relationships of implicit and explicit attitudes with contemporary English, historical English, and 53 non-English languages. *Soc Psychol Personal Sci* **15**, (2024).
17. Charlesworth, T. E. S. & Banaji, M. R. Word embeddings reveal social group attitudes and stereotypes in large language corpora. in *Handbook of Language Analysis in Psychology* (eds. Dehghani, M. & Boyd, R. L.) 508–594 (Guilford Publications Inc., New York, 2022).
18. Jackson, J. C. *et al.* From Text to Thought: How Analyzing Language Can Advance Psychological Science. *Perspectives on Psychological Science* **17**, 805–826 (2022).
19. Caliskan, A., Bryson, J. J. & Narayanan, A. Semantics derived automatically from language corpora necessarily contain human biases. *Science (1979)* **356**, 183–186 (2016).
20. Bhatia, S. & Walasek, L. Predicting implicit attitudes with natural language data. *Proceedings of the National Academy of Sciences* **120**, (2023).
21. Lewis, M. & Lupyan, G. Gender stereotypes are reflected in the distributional structure of 25 languages. *Nat Hum Behav* **4**, 1021–1028 (2020).
22. Charlesworth, T. E. S., Caliskan, A. & Banaji, M. R. Historical Representations of Social Groups Across 200 Years of Word Embeddings from Google Books. *Proceedings of the National Academy of Sciences* **119**, (2022).
23. Garg, N., Schiebinger, L., Jurafsky, D. & Zou, J. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proc Natl Acad Sci U S A* **115**, E3635–E3644 (2018).
24. Pennington, J., Socher, R. & Manning, C. D. GloVe: Global vectors for word representation. in *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference* 1532–1543 (2014). doi:10.3115/v1/d14-1162.
25. Grave, E., Bojanowski, P., Gupta, P., Joulin, A. & Mikolov, T. Learning Word Vectors for 157 Languages. in (2018).
26. Eagly, A. H. & Chaiken, S. Attitude Structure and Function. in *The handbook of social psychology* (eds. Gilbert, D., Fiske, S. T. & Lindzey, G.) 269–323 (Oxford University Press, New York, 1998). doi:10.2307/2072868.
27. Dunlap, R. E. & Jones, R. E. Environmental Concern: Conceptual and Measurement Issues. https://www.researchgate.net/publication/285810112 (2002).
28. Cruz, S. M. & Manata, B. Measurement of Environmental Concern: A Review and Analysis. *Front Psychol* **11**, 493793 (2020).

29. Milfont, T. L. & Duckitt, J. The environmental attitudes inventory: A valid and reliable measure to assess the structure of environmental attitudes. *J Environ Psychol* **30**, 80–94 (2010).
30. Eaton, A. A. & Visser, P. S. Attitude Importance: Understanding the Causes and Consequences of Passionately Held Views. *Soc Personal Psychol Compass* **2**, 1719–1736 (2008).
31. Krosnick, J. A. Attitude Importance and Attitude Accessibility. *Pers Soc Psychol Bull* **15**, 297–308 (1989).
32. Hamilton, W. L., Leskovec, J. & Jurafsky, D. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 1489–1501 (2016).
33. Doell, K. C. *et al.* The International Climate Psychology Collaboration: Climate change-related data collected from 63 countries. *Scientific Data 2024 11:1* **11**, 1–17 (2024).
34. Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science (1979)* **342**, 850–853 (2013).
35. Gibbs, H. K. & Ruesch, A. New IPCC Tier-1 Global Biomass Carbon Map for the Year 2000. https://doi.org/10.15485/1463800 (2008) doi:10.15485/1463800.
36. Inglehart, R. Public Support for Environmental Protection: Objective Problems and Subjective Values in 43 Societies. **28**, 57–72 (1995).
37. Panayotou, T. Demystifying the environmental Kuznets curve: turning a black box into a policy tool. **2**, 465–484 (1997).
38. Wang, Q., Wang, X., Li, R. & Jiang, X. Reinvestigating the environmental Kuznets curve (EKC) of carbon emissions and ecological footprint in 147 countries: a matter of trade protectionism. *Humanities and Social Sciences Communications 2024 11:1* **11**, 1–17 (2024).
39. World Bank Poverty and Inequality Platform. Income inequality: Gini coefficient, 2024. https://ourworldindata.org/grapher/economic-inequality-gini-index?time=latest (2025).
40. UNDP Human Development Report. Gender Inequality Index, 2023. https://ourworldindata.org/grapher/gender-inequality-index-from-the-human-development-report (2025).
41. International Telecommunication Union & World Bank. Share of the population using the Internet, 2023. https://ourworldindata.org/grapher/share-of-individuals-using-the-internet (2025).
42. Wikipedia. List of Wikipedias. https://meta.wikimedia.org/wiki/List_of_Wikipedias (2025).
43. Wikipedia. Wikimedia Statistics - All wikis - Page views by country. https://stats.wikimedia.org/#/all-projects/reading/page-views-by-country/normal%7Cmap%7Clast-month%7C~total%7Cmonthly (2025).
44. Dataset Record: ESA Land Cover Climate Change Initiative (Land_Cover_cci): Land Surface Seasonality Products. https://catalogue.ceda.ac.uk/uuid/7c114fc6e2884c1f9ca107e7a502fdbf/.
45. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behavioral and Brain Sciences* **33**, 61–83 (2010).
46. Muthukrishna, M. *et al.* Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) Psychology: Measuring and Mapping Scales of Cultural and Psychological Distance. *Psychol Sci* **31**, 678 (2020).

Supplementary Information Appendix for:

Language Reveals Global Links Between Nature Attitudes and Sustainable Development

Tessa E. S. Charlesworth*[1], Leland K. Werden[2], Johan van den Hoogen[2], Madalina Vlasceanu[3], Thomas Lauber[2], Thomas W. Crowther[4,5]

[1] Kellogg School of Management, Northwestern University
[2] Department of Environmental Systems Science, ETH Zürich
[3] Department of Environmental Social Sciences, Stanford Doerr School of Sustainability
[4] Branch Institute, Zurich, Switzerland
[5]Environmental Sciences and Engineering, King Abdullah University of Science and Technology

Author note. *Correspondence concerning this article should be directed to Tessa Charlesworth, Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston, IL, 60208, email: tessa.charlesworth@kellogg.northwestern.edu.

Funding for this project was provided for TESC from the Roberta Buffett Institute; for TWC, TL, LKW, and JvdH through grants from DOB Ecology and the Bernina foundation; and for LKW through a Google carbon removal research award.

All data and analyses are provided at the Open Science Framework: https://osf.io/hdxew/?view_only=d2ce89ee81874e9d87dcae201cf177bd (currently an anonymized view-only link that will be made public following peer review).

Table of Contents

1. Word lists: Representing nature and semantic dimensions in English

To represent concepts of nature and semantic dimensions of importance, positivity, negativity, concern, and protection, we sought to balance comprehensiveness and specificity. Comprehensiveness meant that we wanted more than one word to capture the concept. Indeed, averaging results across a *list* of target words helps to guard against idiosyncratic associations or polysemous meanings that may arise from any single word (e.g., *land* might have polysemous associations with the verb "to land" but, when averaged with other nature-related words, results will converge towards the intended nature-related meaning). On the other hand, we did not want to be so comprehensive as to lose the focus of the concept. As such, we used only words that were clearly direct synonyms of the concept of interest.

For the "nature" concept, we started by generating a longer list that also focused on words more related to current climate and biodiversity concerns, using the words *nature, climate, environment, land, forest, forests, biodiversity, restoration, reforestation, ecology*. However, within the English vocabulary, only the words of *nature, climate, environment, land, forest, restoration* were available. Furthermore, tests of the average cosine similarities between these words showed that *restoration* was less related to the other terms ($M = 0.30$, all others $M > .40$), and so we removed the term *restoration.* Because of a potential concern that results could be driven by the specific focus on the term *forest* (i.e., that this would lead to more correlations with forest-related outcomes rather than, say, water-related environmental outcomes) we also performed robustness tests with only the words *nature, climate, environment, land* and ensured that similar results emerged for this shorter list.

Table S1.
Word lists to represent *Nature* and semantic dimensions in English

Language	Nature	Importance	Positive	Negative	Concern	Protect
English	Nature, climate, environment, land, forest	Important, importance, significant, meaningful	Good, nice, wonderful, excellent, exceptional, beautiful, pleasant	Bad, ugly, horrible, gross, unpleasant, horrific, horrendous	Concern, concerned, worry, worried	Protect, defend, preserve, save

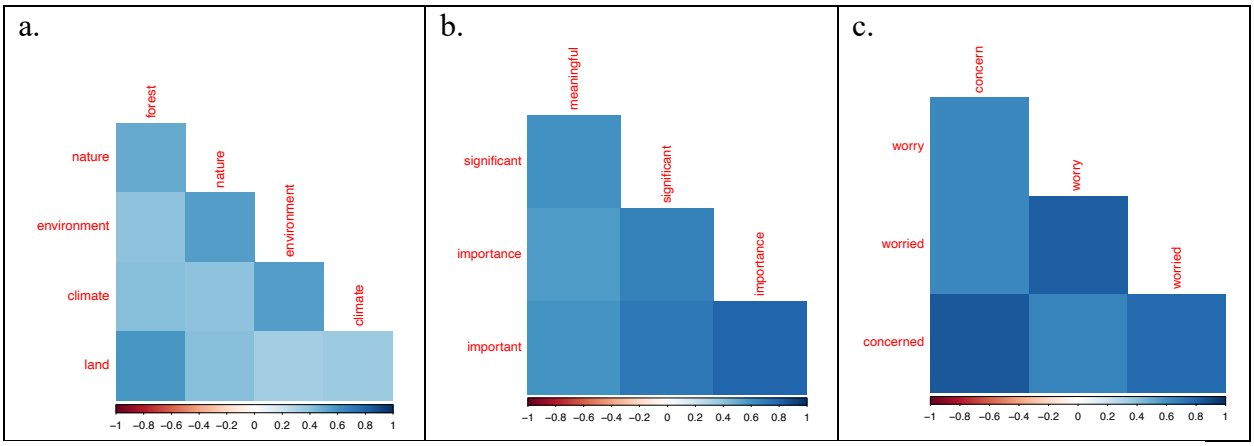

Fig S1. Inter-correlations between words representing nature (a), importance (b), and concern (c). All concepts show significant inter-correlations, indicating that they are unified concepts in language.

Notably, the inter-correlations between nature (or semantic dimension) concepts were significantly higher than the empirical null of all possible pairwise inter-correlations (or inter-cosine similarities) between word vectors. To construct this empirical null, we took all pairwise correlations (and, for robustness, also all pairwise cosine similarities) of all ~14,000 words in the vocabulary. On average, the inter-correlation among all words was $r = .11$ (SD = .10) which, although significantly different from zero itself ($p < .001$), was also significantly lower than the inter-correlations among the nature and semantic dimension concepts which were always $r > .30$. Said simply, the inter-correlations among the nature and semantic dimension concepts are not merely an artifact of random word similarities but are significantly more related than would be expected from any random set of words.

2. Semantic exploration of contemporary English: How is nature represented in English Internet text?

This project is the first to explore the bottom-up representation of *Nature* in English text. As such, we first ran a pilot study to understand the signatures of the top-associated words with a basic *Nature* concept across various sources of English Internet text. The main result of the pilot study is that it establishes the *Importance* dimension (rather than, say, *Positivity, Negativity*) as the most consistently-associated dimension with *Nature,* across both the general concept and the discovered latent factors. As such, in the main text, we focus primarily on the *Nature-Importance* associations, although results for other dimensions are reported in this supplementary appendix.

Bottom-up discovery of the associated words with Nature in English. We began by discovering, bottom-up, the top-50 words (out of a possible 14,000-word vocabulary; Warriner et al., 2014) that were most associated with *Nature* (represented by synonyms including *nature, climate, environment, land;* Table S1) within 840 billion words of contemporary English Internet text (using the pre-trained GloVe word embeddings[1]). The top-50 words were discovered by calculating the average cosine similarity of each word in the vocabulary to the list of *Nature* synonyms, and then ranking all words by their average cosine similarity. The final list of top-50 words is provided in Table S2 below.

Note that we also replicated this result using a shorter list of just 4 *Nature* words (nature, climate, environment, land) that omitted the additional words focusing on restoration and biodiversity. Critically, we found robust and similar face-valid results regardless of whether we used the longer or shorter word lists.

Table S2.
Bottom-up discovery of top-50 words associated with *Nature* in English Internet text

Overall list	forest, nature, environment, climate, land, ecological, environmental, landscape, habitat, vegetation, conservation, ecosystem, wildlife, natural, coastal, earth, agricultural, rainforest, ocean, agriculture, surroundings, wilderness, fauna, pollution, economic, urban, woodland, diversity, farming, vast, change, tropical, soil, world, impact, country, extent, rural, farmland, mountain, future, preserve, life, terrain, water, peaceful, countryside, atmosphere, cultural, species
Replication with 4-word list	environment, nature, climate, environmental, land, ecological, landscape, forest, ecosystem, natural, economic, conservation, earth, habitat, change, impact, surroundings, agricultural, wildlife, future, diversity, agriculture, world, understanding, ocean, pollution, extent, coastal, vegetation, life, society, sense, existence, cultural, vast, concern, urban, atmosphere, matter, situation, way, importance, concerned, human, perspective, affect, development, country, exist, global

Discovered factors in the *Nature* concept. To summarize the latent meanings and factors in these top-50 words, we then performed an Exploratory Factor Analysis (EFA) on the inter-correlations among the top-50 words (Figure S2), with oblimin rotation and parallel analysis as implemented in the *fa()* function in R. Visual inspection of the scree plot suggested five factors and, indeed, five factors were sufficient to explain 59% of variance in the inter-correlations among words. The five factors that emerged were also distinct and interpretable, centering on concepts of: habitats (either more wild habitats in Factor 1, or more cultivated pastoral habitats in Factor 3); human-nature interactions and culture; agriculture; and broad, elemental and life concepts (Table S3).

These factors aligned with known discussions around nature, including highlighting a distinction between agricultural land versus wilderness areas, a concept that pervades contemporary English-speaking perspectives. The fact that data-driven linguistic approach uncovers face-valid factors, even without top-down researcher control, reinforces confidence in the method to expand horizons of how nature is represented across history and languages.

Table S3.
Bottom-up discovery of 5 factors explaining the top-50 words associated with *Nature* in English Internet text. Top words (loadings > 0.30) associated with each factor.

F1: Wild habitats	**F2: Human-nature interaction**	**F3: Cultivated habitats**	**F4: Agriculture**	**F5: Elements**
habitat, fauna, species, wildlife, conservation, vegetation, rainforest, ecological, ecosystem, woodland, diversity, forest, tropical, coastal, environmental, preserve, wilderness, nature	cultural, future, life, impact, change, economic, world, diversity, nature, extent, environment, country, environmental, atmosphere, vast, surroundings, climate, ecological, urban	countryside, mountain, surroundings, peaceful, woodland, rural, wilderness, country, landscape, farmland, terrain	agricultural, agriculture, farming, farmland, rural, economic, land, pollution, soil, urban, environmental, country, countryside	water, ocean, earth, soil, tropical, pollution, atmosphere, land, natural, mountain, world, climate

Relationship of *Nature* concept (and subfactors) to other semantic dimensions.

Finally, our primary interest in the pilot study was Nature concepts (and its subfactors, such as habitats and culture) are linked to commonly-studied semantic dimensions in nature attitudes. To aid interpretability of this bottom-up *Nature* representation, we also examined how the overall concept and the bottom-up factors are associated with five other well-studied semantic dimensions: *Positivity*, *Negativity, Concern*, *Protection*, and *Importance,* each represented by a set of synonyms (Table S1).

Specifically, we examined the average cosine similarity between the words representing each bottom-up factors (e.g., *Wild habitats: habitat, fauna, etc.*) and the words representing each top-down semantic dimension (e.g., *Importance: important, importance, significant, meaningful*; Figure S2). Across all semantic dimensions, *Importance* consistently emerged as being moderately or strongly correlated with all bottom-up factors (overall $r = .37$), and was especially

correlated with the factor on human culture interactions, $r = .51$, with the weakest association to importance emerging for the factor of cultivated habitats, $r = .24$. Notably, the sample of pairwise correlations of all *Nature* and *Importance* words (e.g., *nature-important, nature-meaningful, climate-important, climate-meaningful,* and so on) was significantly stronger than the aforementioned empirical null of all random word-word correlations (which had a mean $r = .11$), $t(19) = 11.96$, $p < .001$. By contrast, *Negativity* was consistently the least associated with the discovered factors (overall $r = .23$), especially for wild habitats ($r = .13$), although even correlations of *Nature-Negativity* were still significantly stronger than any random word-word correlations from the empirical null, $t(34) = 11.29$, $p < .001$. Nevertheless, this implies that nature is generally *not* perceived as negative in contemporary English but, instead, is most often perceived as important.

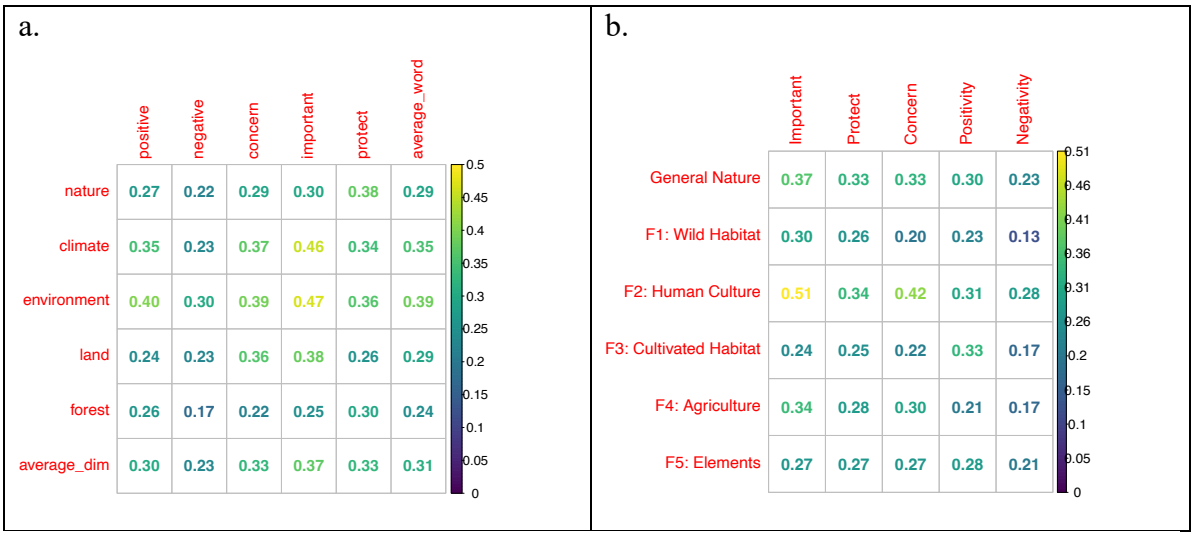

Fig S2. Inter-correlations between (a) all individual words of the overall nature concept and average semantic dimensions, and (b) average nature concept, and all subdimensions of nature, with semantic dimensions. Note that Fig S2b. is reproduced in the main text.

3. Changes over time: How have Nature representations changed in historical English books?

Next, we considered how the *Nature* concept has changed in English, spanning 200 years of English book text, using pre-trained word2vec embeddings[2]. Notably, because words like *biodiversity* and *reforestation* are relatively more recent in historical usage, for these historical analyses we focused on the shorter 4-word list of *Nature* terms (*nature, climate, environment, land*) especially since the pilot study results showed robust results across both word lists. First, for the bottom-up discovery of the neighborhood of the *Nature* concept, we show that there are similar, robust, and face-valid results across history (Table S4).

Table S4.
Bottom-up discovery of the top-50 words associated with *Nature* in historical English book text for selected decades

1800	**1900**	**1990**
climate, soil, nature, land, fertility, foil, torrid, zone, diversity, pasture, situation, fertile, country, proximity, spontaneous, culture, badness, moist, vary, extent, tract, fruitful, temperature, variable, inhabit, hemisphere, vegetation, polar, peculiar, scenery, moisture, atmosphere, quality, sea, produce, drought, season, thrive, decomposition, inhabitant, surplus, crop, district, change, chemical, geographical, constitution, weather, native, phenomenon	climate, environment, nature, land, soil, surroundings, configuration, adaptation, fertility, adapt, organism, physical, peculiar, indigenous, character, vegetation, fertilize, temperament, social, inherent, scenery, natural, emotional, diversity, fertile, humidity, relation, interaction, habitat, situation, plant, economic, variety, environmental, intellectual, depend, location, change, seasonal, dependent, material, food, phenomenon, individual, sustenance, tropics, geography, country, fauna, sensuous	environment, climate, nature, land, environmental, vegetation, soil, landscape, ecological, atmosphere, fauna, change, ecosystem, terrain, pollution, conducive, habitat, global, geography, impact, conservation, wildlife, fertility, tropical, tropics, natural, surroundings, humid, indigenous, degradation, diversity, development, resource, aquatic, coastal, situation, abundance, awareness, quality, depend, relationship, agriculture, deterioration, interaction, nurture, human, preservation, countryside, infrastructure, erosion

Next, we addressed a possible concern that the representations of *Nature* we are using today (e.g., with factors including wilderness or farming) could be anachronistic, and that other meanings or dimensions may have been more relevant in the past. We therefore performed the same exploratory factor analysis (EFA) as we did in the earlier pilot study on contemporary English but now repeated on all 20 decades of historical English text. Results confirm that a similar set of latent factors appear to explain nature representations across all 200 years: for instance, in every decade, there are usually three sets of factors with words related to (1) wilderness/ecosystems, (2) farming/agriculture, and (3) culture/human-nature interactions (Table S5).

Still, there are some interesting nuances in the specific words associated with different factors over time. For instance, the human-environment interaction factor in 1990s included words such as *degradation* and *pollution;* however the similar factors in 1900 focused more on *adaptation,* as well as more *social* and *emotional* relations. This underscores that the negative aspects of pollution and degradation have more recently become so central in human-environment discussions, especially following the turn of the century. Overall, though, we emphasize that the robustness of the latent representation of *Nature* confirms that what is changing is the association of *Nature-Importance* rather than only changes in the definition or meaning of nature itself.

Table S5.
Exploratory factor analysis across selected decades (1900, 1950, 1990) of historical English books.

1900	**"adaptation"** environment, adapt, adaptation, environmental, organism	**"agriculture"** fertilize, sustenance, soil, plant, food, indigenous, land, fertile, vegetation	**"habitat"** climate, scenery, fauna, vegetation, tropics, humidity, seasonal, surroundings, fertility,	**"culture"** intellectual, emotional, social, physical, sensuous, economic, nature, natural, individual,	**"general"** depend, location, dependent, relation, change, individual,

			diversity, variety, peculiar, configuration, change, situation, geography, indigenous, environmental	temperament, character	configuration, situation
1950	**"general"** particular, nature, peculiar, relation, inherent, unique, configuration, character, knowledge, diversity, terrain	**"habitat"** tropical, tropics, indigenous, fauna, vegetation, fertile, land, climate, soil, habitat, fertility, thrive, ecological, geographical	**"cultural"** cultural, social, economic, intellectual, culture, geography, physical	**"adaptation"** environment, surroundings, organism, atmosphere, environmental, adapt, adaptation, interaction, thrive, conducive, climate	**"agriculture"** landscape, scenery, texture, character, vegetation
1990	**"habitats"** wildlife, habitat, fauna, conservation, aquatic, ecosystem, diversity, natural, ecological, tropical, preservation, resource, indigenous, vegetation, tropics, agriculture, abundance, human	**"interaction"** relationship, nature, interaction, surroundings, awareness, situation, change, depend, nurture, environment, human	**"climate"** atmosphere, climate, humid, tropics, tropical, soil	**"human effect"** environmental, pollution, erosion, deterioration, infrastructure, impact, development, global, degradation, quality	**"countryside"** countryside, terrain, land, landscape

Note. The factors are colored to indicate the general themes reflected in the top-loading words for each factor, as determined through discussion among the authors. Green reflects themes of habitats/ecosystems (similar to the wilderness factor in contemporary English), purple reflects themes of agriculture or food production (similar to the agriculture factor), and blue reflects themes of human-environment interactions (similar to the pollution and preservation factors).

Third, how have the Nature representations changed in their associations to all other semantic dimensions (e.g., *Concern, Protect*) across 200 years of English text? As reported in Table S6 (and summarized in the main text), we saw that the association of *Nature* and *Importance* increased more strongly than any other dimension, although both *Concern* and *Protection* also increased at slower rates. Still, by the end of the century, *Importance* was now the most-associated dimension. By contrast, both *Negative* and *Positive* dimensions decreased in associations (albeit not significantly), and were essentially at neutral associations by the end of the century.

Table S6.

Relationships of *Nature*-semantic dimensions across historical English text; starting (1800) and ending (1990) values and Spearman's correlation with time.

Importance	**Concern**	**Protect**	**Positive**	**Negative**
Start = 0.04, *End* = 0.12	*Start* = 0.04, *End* = 0.09	*Start* = 0.05, *End* = 0.08	*Start* = 0.11, *End* = 0.04	*Start* = 0.07, *End* = 0.01
rho = .83, $p < .001$	*rho* = .76, $p < .001$	*rho* = .77, $p < .001$	*rho* = -.17, $p = .48$	*rho* = -.36, $p = .12$

Note. Results are reported for the associations between word lists (i.e., the mean average cosine similarity between the synonyms for *Nature* and *Importance*).

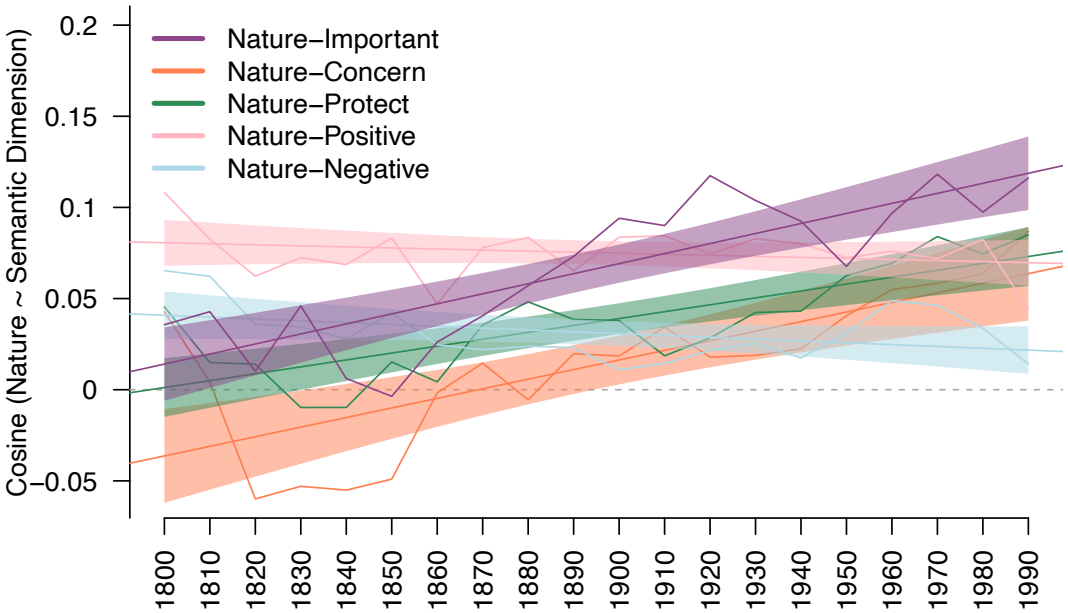

Fig S4. Changes in cosine similarities between nature and semantic dimensions. Error bars represent 95% confidence intervals around simple bivariate linear regression predicting the cosine similarity timeseries from a time vector (of decade 1 to 20).

Finally, we look at changes in the relationships between *Importance* and the 5 latent discovered factors (e.g., *Importance-Culture, Importance-Wilderness*). The largest and most notable increase is observed in *Culture-Important* associations (Figure S5), which moved from a neutral association in 1830 ($M = 0.07$) to a small but significant association in 1990 ($M = 0.16$; *rho* = .90, $p < .001$). By contrast, the association of *Important* with most other subfactors have

always been relatively small in magnitude and with weak to no slopes over time, all *rhos* < .46, *p* > .06, with the exception of *Important-Elements* (e.g., ocean, water, earth) which also increased at *rho* = .67, *p* = .003, but not nearly as much as *Important-Culture.* Thus, English text persistently conveyed some importance for a consideration of human-nature interactions, but it is only in the past century (indeed, mostly since the since the industrial revolution), that concerns of human impacts have gained the most importance in language. This result also reinforces how cultural representations, discovered bottom-up from language, are intertwined with real-world changes in policy and social movements, not only for social stigma[3,4], but also newly for environmental attitudes.

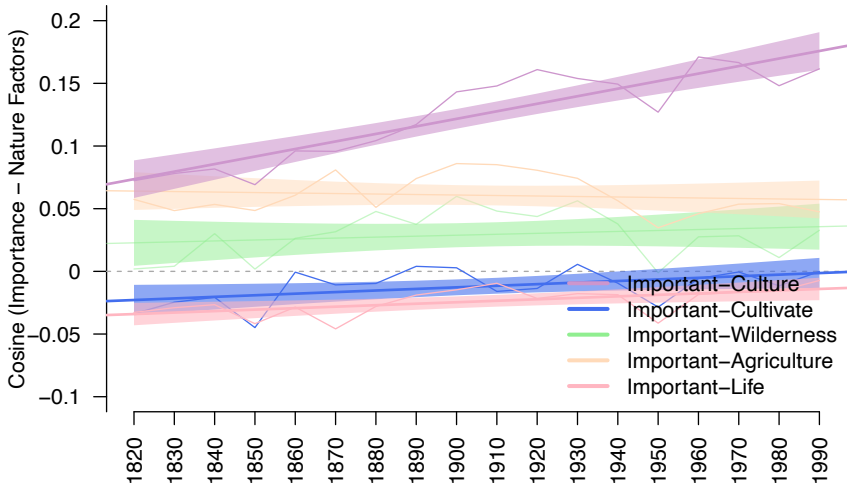

Figure S5. **Changes in *Importance*-factor associations across historical English book text (1800-1990)**. Time (decades) is on the x-axis and cosine similarities (essentially correlations) on the y-axis. The strongest and most increasing association was of importance and culture (human-nature interactions).

4. Variation across languages: Additional covariates of speaker and corpus size

We collected data from Wikipedia estimates of the number of speakers of each language (compiled from our own searches on Wikipedia), as well as the number of Wikipedia pages of each language (scraped from https://www.tensorflow.org/datasets/catalog/wikipedia). The cross-linguistic data varied substantially in (1) the number of speakers, as well as (2) the corpus size, ranging from small languages like *Romansch* (spoken in Switzerland, ~40,000 speakers, and ~3,939 Wikipedia pages) and *Maori* (spoken in New Zealand, estimated at ~100,000 speakers, and ~7,855 Wikipeda pages) up to *English* (spoken worldwide, estimated at ~1.5 billion speakers, and ~ 6,672,479 Wikipedia pages). We planned to test the raw correlations between the number of speakers, corpus size, and the strength of *Nature-Importance* associations. However, because of the extreme dominance of some languages like English and Chinese, we rank-transformed these variables so that English was the final rank (i.e., the largest corpus and the largest number of speakers) and then computed the Spearman correlations on these rank-transformed values.

Results showed no significant correlation between the number of speakers of a language and the magnitude of *Nature-Importance* associations, $rho = -.04$, $p = .69$ (Figure S6A). Additionally, there was also no significant correlation between the number of Wikipedia pages (i.e., the corpus size) and the magnitude of *Nature-Importance* associations, $rho = -.17$, $p = .09$ (Figure S6B). In other words, it was not the case that only small (rarely-spoken or small data source) languages were the ones that had more extreme *Nature-Importance* associations.

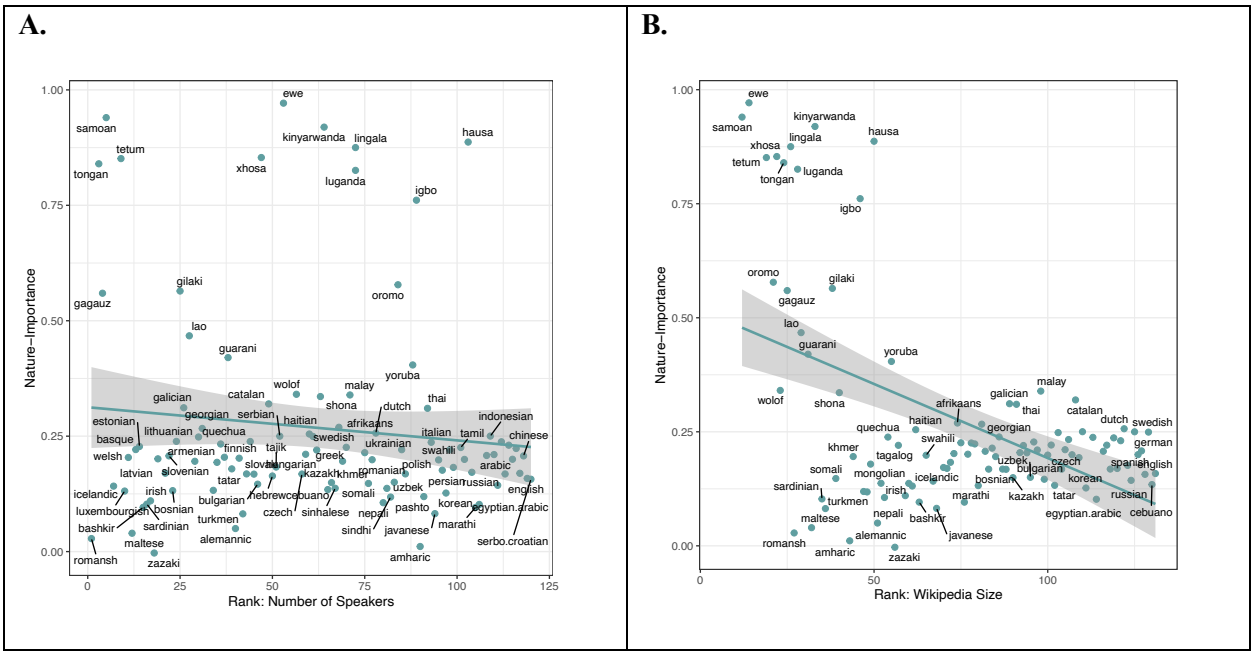

Figure S6. **Magnitude of *Nature-Importance* associations across 120 languages as a function of the number of speakers (A) or corpus size (B)**. Number of speakers and corpus size were rank-transformed, such that larger languages had larger ranks.

5. Variation across languages: Additional details on other dimensions

For simplicity in the main text, we focus on the variation across languages in *Nature-Importance* associations. However, here, we summarize the results for other key semantic dimensions and some of the discovered exploratory factors from the pilot studies in English. First, it is noteworthy that all of the dimensions and factors are generally correlated across languages (Figure S7). This means that languages that have high associations of *Nature-Importance* also have high associations of *Nature* with other dimensions, including *Concern* (r = .91), and *Protection* (r = .92). Thus, although the main text focuses on *Nature-Importance,* these high correlations by languages across dimensions suggest that similar results would be found across other semantic dimensions as well. We see similar conclusions when visualizing the mean cosine similarities across all languages and all dimensions (Figure S8).

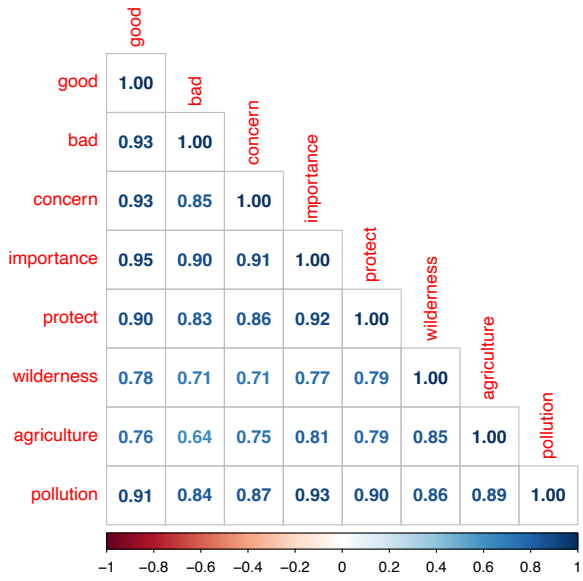

Fig S7. **Correlations across 120 languages in associations between *Nature*-semantic dimensions**. In addition to the five core semantic dimensions of positivity (good), negativity (bad), concern, importance, and protection, we also explored three other factor dimensions that continued to emerge in historical English and across languages, focusing on wilderness (i.e., wild habitats), agriculture, and pollution (i.e., the human-culture relationships). All numbers are

reporting Pearson's correlation values, with darker colors indicating stronger correlations. The correlations indicate that how the languages vary in their association of *Nature-Good,* for example, is strongly correlated with how the languages vary in their association of *Nature-Bad,* and so on.

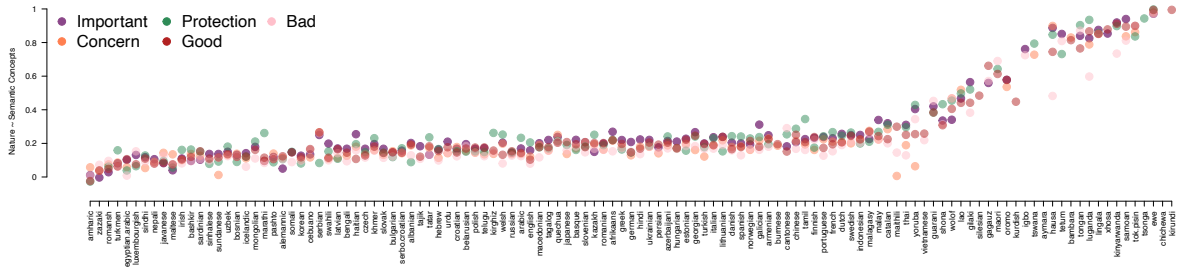

Fig S8. **Variation in *Nature*-semantic dimension associations across languages.** *Nature-Importance* (purple), *Nature-Concern* (orange) and *Nature-Protection* (green) cosine similarities (essentially correlation scores) for each language. Languages are ordered along the x-axis by their average cosine similarity across all dimensions, with languages showing the smallest associations on the left side of the plot and the largest associations on the right side of the plot. The similar trends across all dimensions underscore that languages high on one dimension (e.g., *Important* are also high on other dimensions).

Next, we used a simple one-way ANOVA to compare the magnitudes of *Nature-Importance* associations across five major language families: Indo-European ($N = 55$ languages), Niger-Congo ($N = 15$), Austronesian ($N = 11$), Afro-Asiatic ($N = 8$), and Other ($N = 30$). There was a significant overall effect of language, $F\ (4, 99) = 15.69$, $p < .001$. Compared to a baseline of Indo-European associations ($M = 0.20$), the strength of *Nature-Importance* was significantly stronger in both Austronesian ($M = 0.40$), $b = 0.21$, SE = 0.06, $t = 3.37$, $p = .001$, $\beta = 0.28$, and in Niger-Congo ($M = 0.65$) languages, $b = 0.45$, SE = 0.06, $t = 7.39$, $p < .001$, $\beta = 0.60$. Indo-European languages were not significantly different from either Afro-Asiatic or Other languages, $b < 0.07$, $p > .32$.

Of note, although *Importance* was the most central and strongest associated dimension in English text, when we look across all other languages, we find that the mean association to *Importance* (M = 0.27. SD = 0.22) is about on-par with *Positive* (M = 0.26, SD = 0.23) and *Protection* (M = 0.27, SD = 0.23). Moreover, we also see that, across all non-English languages, *Nature* is most strongly associated with *Wilderness* (M = 0.44, SD = 0.21) and *Agriculture* (M = 0.35, SD = 0.20). These results have face validity and lend confidence in the findings since *Wilderness* and *Agriculture* are indeed more deeply tied to the very meaning of *Nature* – after all, they were discovered bottom-up from the words associated with *Nature* rather than as external semantic dimensions.

6. Variation across languages: Bottom-up discovery of factors for commonly-spoken non-English languages

One potential concern with the current approach is that non-English languages could have fundamentally different definitions of *Nature* (i.e., different latent factors) that we might have missed by starting with the English stimuli and translating them into different languages. To address this concern, we explore, bottom-up, the top-50 words and factor structure emerging as associates with *Nature* within the 5 most commonly-spoken non-English languages (French, Spanish, Arabic, Bengali, Chinese).

Results confirm that these bottom-up, latent representations of *Nature* across non-English languages contain the same signatures, including a factor on wilderness/ecosystems, and agriculture/farming, as well a culture/human-nature interaction factor. Such consistency implies that non-English definitions of nature share the same conceptual space as the English, and thus we can accurately compare how this similar concept of *Nature* is differentially associated with other semantic dimensions (such as *Importance*) across languages.

Table S7.
Bottom-up representations in widely-spoken languages, including exploratory factor analysis with labelled factors

Language	Top 50 words (translated to English)	Top 5 words associated with a 5-factor solution				
Arabic	environ, environment, climate, thicket, forest, ground, precinct, district, area, zone, region, soil, woodlands, garden, floor, inside, jungle, airs, desert, midst, amid, amidst, central, middle, waist, plateau, knoll, perimeter, species, surroundings, ocean, circumference, city, ranch, farm, plantation, islet, island, isle, acreage, cultivation, loch, lake, strand, coast, afield, farmer, tree, village, hamlet	**"controlled space":** city, islet, knoll, strand, zone	**"controlled space":** inside, afield, species, waist, zone	**"climate":** environ, climate, soil, airs, floor	**"wilderness":** woodlands, jungle, forest, tree, lake	**"agriculture"**: ranch, farmer, cultivation, garden, tree

Bengali	environs, environment, environ, woodlands, nature, climate, climatology, scenic, booking, conservation, saving, save, preserving, preservation, reservation, riches, assets, resources, restitution, retrieve, regain, retrieval, reconnaissance, restoring, recovery, reclamation, restored, recoup, vegetation, fauna, swamp, swampy, marsh, management, region, hydrology, territory, area, locality, landscaping, agriculture, farming, agricultural, biosphere, contamination, pollution, lowlands, culture, geography, weather	**"agri management":** reservation, management, riches, environment, restitution	**"space":** region, territory, lowlands, swampy, woodlands	**"climate":** climatology, climate, weather, hydrology, geography	**"habitats":** vegetation, woodlands, swampy, pollution, lowlands	**"culture":** biosphere, nature, fauna, culture, scenic
Simplified Chinese	environs, surroundings, environ, circumstance, environment, climate, forest, nature, naturally, vegetation, economy, economical, scape, landscape, farming, agriculture, agricultural, science, calamity, lands, land, fend, conservation, belay, protecting, protect, protection, resources, source, society, soc, human, mankind, timberland, woodlands, soil, park, fishery, nation, country, area, development, evolve, developing, diversity, multiplicity, husbandry, scenery, habitat, lush	**"agriculture":** husbandry, agriculture, fishery, economical, developing	**"wilderness"**: woodlands, forest, lands, vegetation, soil	**"cultivated":** park, scenery, scape, nation, lush	**"human society":** nature, science, soc, human, calamity	**"climate":** vegetation, habitat, climate, multiplicity, scape
French	environment, fauna, agriculture, farming, husbandry, durable, lasting, climate, flora, landscape, savage, wild, untamed, upland, habitat, prairie, agricultural, littoral, preserving, conservation, civilization, protection, vie, savanna, underwood, heritage, patrimony, jungle, hydrology, middle, mid, midst, plain, territory, planting, plantation, politics, policy, politic, vert, moist, wet, humid, damp, dank, harmony, peace, rural, botanical, botany	**"space":** littoral, landscape, territory, environment, patrimony	**"habitats":** prairie, plain, savanna, jungle, upland	**"protect":** peace, vie, harmony, protection, policy	**"preserve":** flora, fauna, botanical, preserving, hydrology	**"agriculture":** farming, agricultural, rural, hydrology, policy
Spanish	woodlands, nature, forest, husbandry, farming, agriculture, fauna, environment, ambient, landscape, biosphere, jungle, flora, earth, land, tenable, weather, culture, landscaping, planet, horticulture, lifelike, prairie, grove, restoration, littoral, seaboard, savage, wild, untamed, life, mankind, humanity, diversity, territory, weedy, scrub, mid, means, medium, midst, middle, medial, park, health, greenhouse, ancestral, zone, orchard, spirituality	**"wilderness":** forest, prairie, grove, jungle, woodlands	**"cultural":** culture, diversity, nature, fauna, spirituality	**"climate":** ambient, weather, landscape, medial, landscaping	**"agriculture":** horticulture, husbandry, orchard, greenhouse, landscaping	**"general intx":** planet, mankind, land, biosphere, life

Note. The factors are colored to indicate the general themes reflected in the top-loading words for each factor, as determined through discussion among the authors. Green reflects themes of habitats/ecosystems (similar to the wilderness factor in contemporary English), red reflects themes of agriculture or food production (similar to the agriculture factor), and blue reflects themes of human-environment interactions (similar to the pollution and preservation factors).

7. Variation across countries: Additional dimensions

In the main text, we visualize the variation across countries (from weighted language estimates) on the *Nature-Importance* associations. Here, we expand to consider all other key semantic dimensions and discovered factors. As expected based on the similarity of all dimension associations across languages, we find that the cross-country results are also moderately-to-strongly correlated across all dimensions, *rs* > .37 (Figure S9). Notably, these correlations are weaker than the cross-language correlations, indicating that there is more country-level variation (versus language-level variation) in how these dimensions are used to describe *Nature.* Indeed, in Figure S10 we can see that some countries (e.g., Canada, United States) are in the lowest quartile for *Nature-Positive* associations but more middle quartiles for *Nature-Wilderness* associations. This suggests that the combination of languages spoken in Canada (e.g., English, French, Mandarin) distinguish nature as relatively wild, but also not necessarily positive.

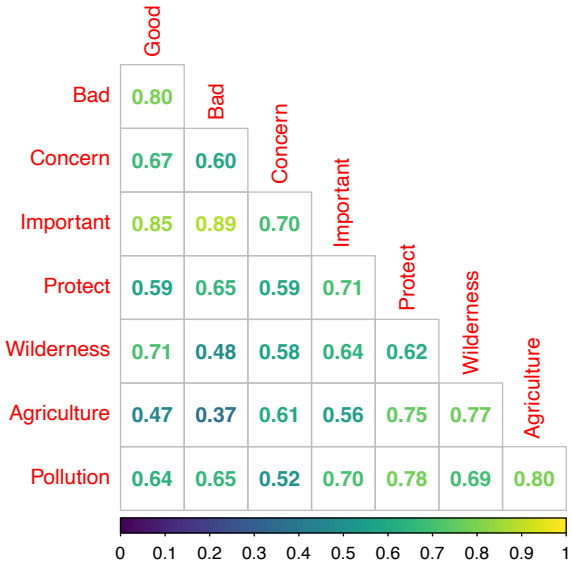

Fig S9. **Correlations across 189 countries in associations between *Nature*-semantic dimensions and other factors**. All numbers are reporting Pearson's correlation values, with more yellow (brighter colors) indicating stronger correlations. The correlations indicate that how

the countries vary in their association of *Nature-Good,* for example, is strongly correlated with how the languages vary in their association of *Nature-Bad,* and so on.

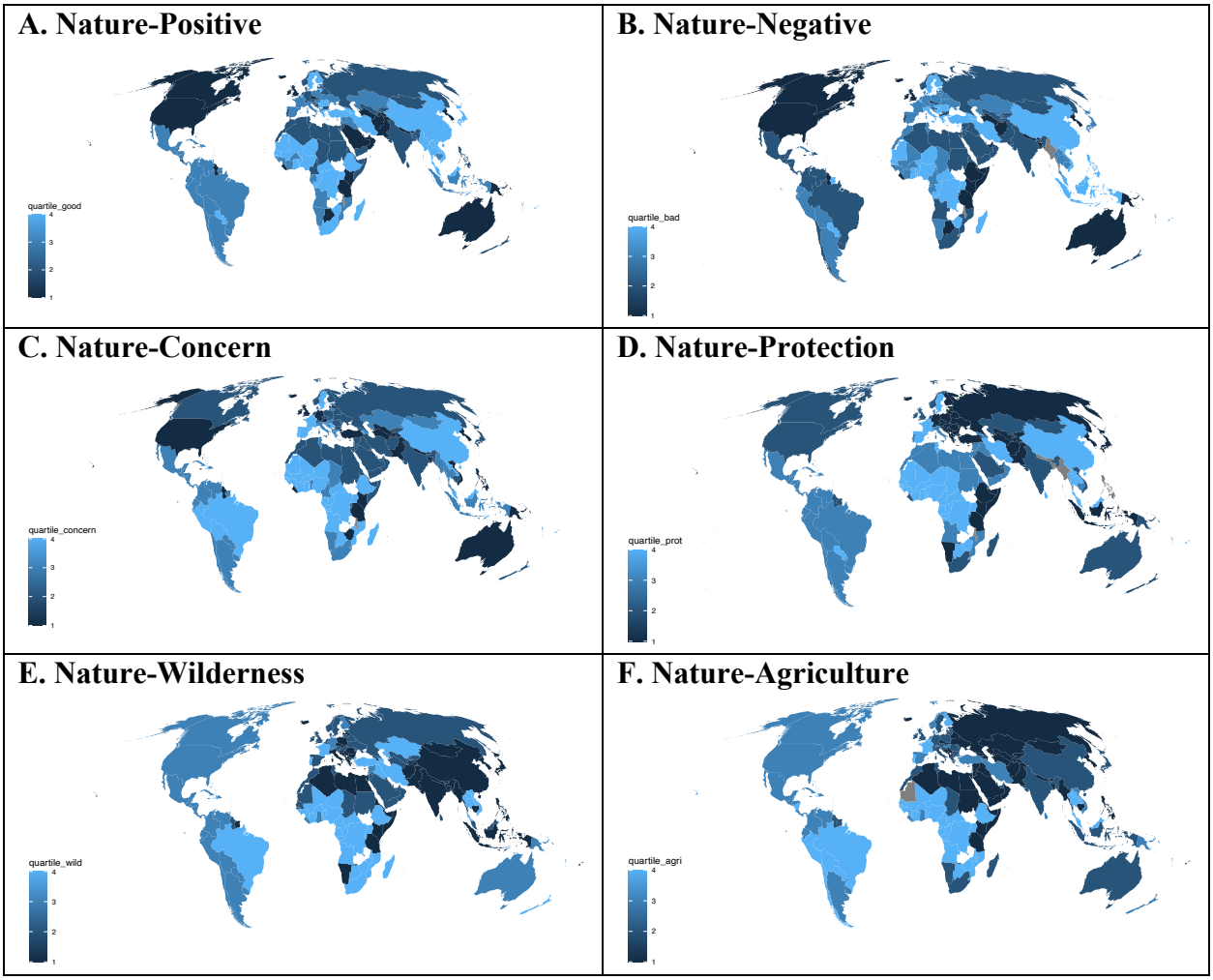

Fig S10. **Variation in *Nature*-semantic dimension representations across countries.** Each panel shows weighted estimates of *Nature-dimension* associations across 189 countries speaking at least one of 120 languages. Lighter blue indicates that the country is in the top quartile (strongest *Nature-dimension* associations), darker blue is the bottom quartile (weakest *Nature-dimension* associations). White indicates no language data was available for that country, gray indicates that language data was available but not for the specific semantic dimension .

8. Variation across countries: Associations with covariates of GDP, Gini, Gender inequality, Internet use

Can economic, social, and linguistic variables help us understand the geographic patterning of global Nature-Importance attitudes? We first examined the inter-correlations among all economic, social, and linguistic covariates (Figure S11A). Results suggested no major correlations or substantial problems of multicollinearity, except for the relationships among the two GDP indicators (GDP from 2014, when the Wikipedia data was collected; and GDP from 2022, closer to the other outcome variables). Because there were few substantial changes in GDP over this period, we chose to only include the more recent GDP measure from 2022. Additionally, there were strong positive associations of Wikipedia page views and GDP measures, likely because richer places both have more Internet access and more literacy that then becomes amplified in Wikipedia use and access. Given our more central focus on economic factors, we chose to leave out the Wikipedia page views data in future analyses.

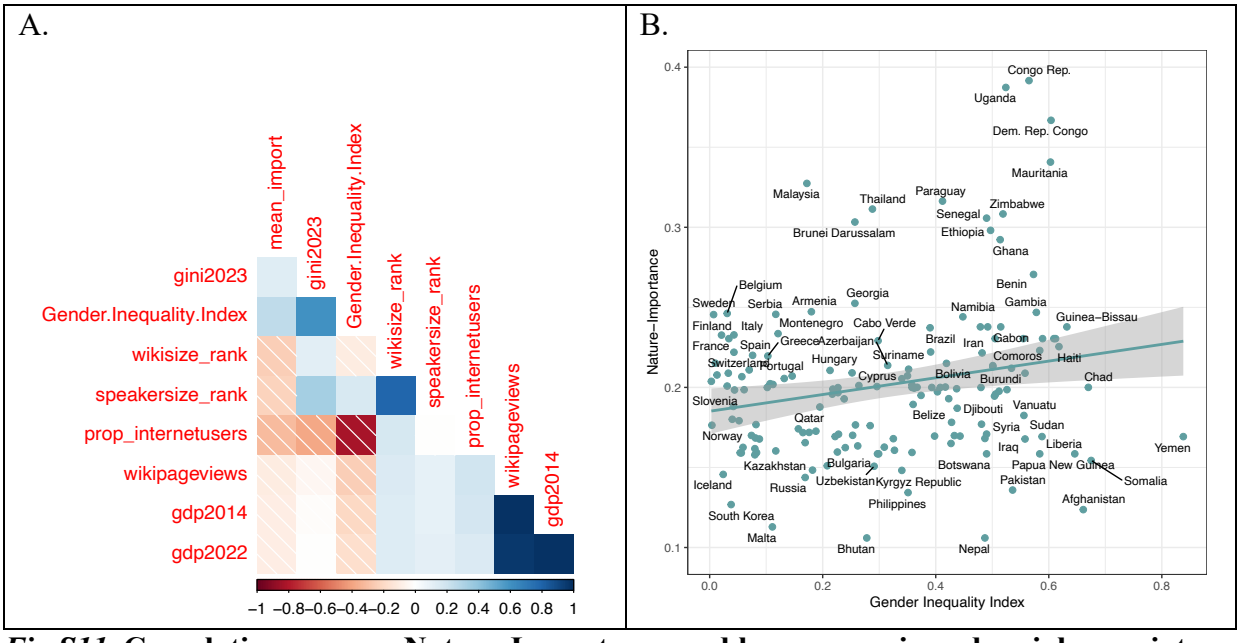

Fig S11. **Correlations among Nature-Importance and key economic and social covariates.** Panel A visualizes all bivariate relationships between economic, social, and linguistic variables.

Positive correlations are shaded in blue, negative correlations are shaded in red. Panel B visualizes the correlation between gender inequality (from the UN SDG Gender Inequality Index; x-axis) and the Nature-Importance associations (y-axis) across countries. Countries with larger gender inequality also have larger Nature-Importance associations.

We also see that, in terms of bivariate relationships with *Nature-Importance* associations across places, the only positive correlates were measures of: (1) income inequality (Gini) which was not significant, $r = .12$ [-.03, .26], $t\ (175) = 1.57$, $p = .12$; and (2) gender inequality (Gender Inequality Index), which was significant but small, $r = .21$ [.06, .36], $t\ (155) = 2.69$, $p = .008$ (Figure S11B). This indicates that more unequal places, and especially more gender-unequal places, may nevertheless have stronger *Nature-Importance* associations; this aligns with the findings so far that the most *Nature-Importance* places are largely in Global South, which have sometimes struggled with achieving gender equality. It is unlikely that these two variables are causally related (i.e., gender inequality does not beget stronger nature importance) but, rather, we consider them to both be the result of cultures in early industrialization stages. These places (a) have more direct connections to nature, perhaps due to subsistence nature relationships, yet (b) are early in the Kuznet's curve in industrialization and thus (c) have not yet gained the additional national capital that supports women development. In a similar vein, we find a significant negative bivariate relationship with Internet access and *Nature-Importance* associations, $r = -.28$ [-.41, -.14], $t\ (171) = -3.79$, $p < .001$, which can be understood as yet another indicator of economic development and democratized access to knowledge and industrialization.

There were also small but significant negative correlations with the country-weighted corpus size and speaker size (i.e., the size of the corpora and speaker populations of the four languages that went into the country's final weighted mean), Specifically, *Nature-Importance* associations were weaker in places where there were larger corpora, $r = -.18$ [-.32, -.03], $t(175) =$

-2.44, p = .02, and where there were languages spoken by larger populations, r = -.20 [-.33, -.04], t(175) = -2.63, p = .01. These effects are likely because of the weak *Nature-Importance* associations in English and other colonial languages of the Global North. But they too align with the idea that weaker *Nature-Importance* associations occur in places that have more industrialization, and economic dominance through dominant languages; in contrast, stronger *Nature-Importance* associations occur in places that have less industrial and technological development, and smaller, rarer languages.

9. Correlation of Nature-Importance with United Nations Sustainable Development Goals: Additional details and regression models

For our key outcome variable we relied on the gold-standard for environmental policy discussions: the United Nations Sustainable Development Goals for SDG14 (life below water) and SDG15 (life on land). First, we explored the inter-relationships between the 13 indicators across SDG14 and 15 (Figure S12), finding both positive and negative correlations among the indicators, mean r = .02, range: [-.47, .76].

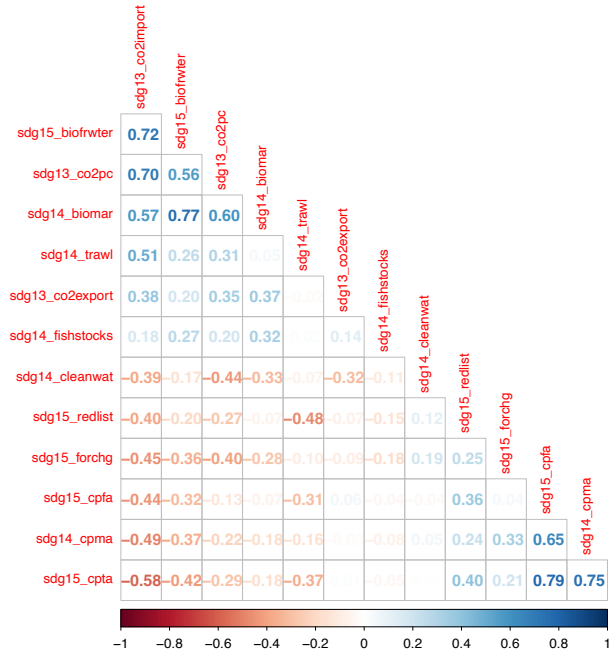

Fig S12. **Inter-correlations among all UN SDG14 and 15 indicators.** Blue colors indicate positive correlations; red colors indicate negative correlations. The full codebook for variable names is provided in the open data on the Open Science Framework.

Visual inspection suggested two factors and, indeed, an Exploratory Factor Analysis with oblimin rotation and maximum likelihood estimation showed that two factors explained 44% of variance across correlations, with the first factor explaining 26% and the second explaining 18%.

The first factor was centered on biodiversity protection, with highest-loading indicators on the first factor were: sdg15_cpta (SDG15.1.2; protected terrestrial biodiversity areas), sdg15_cpfa (SDG15.1.2 as well; protected freshwater biodiversity areas), and sdg14_cpma (SDG14.5.1; protected marine biodiversity areas). For simplicity and streamlining in the main text, we therefore chose to focus on the first two indicators because they are both the most centrally correlated (i.e., the highest loading indicators on the primary factor) are clearly conceptually overlapping. We visualize the geographic variability and availability of 15.1.2 indicators in Figure S13.

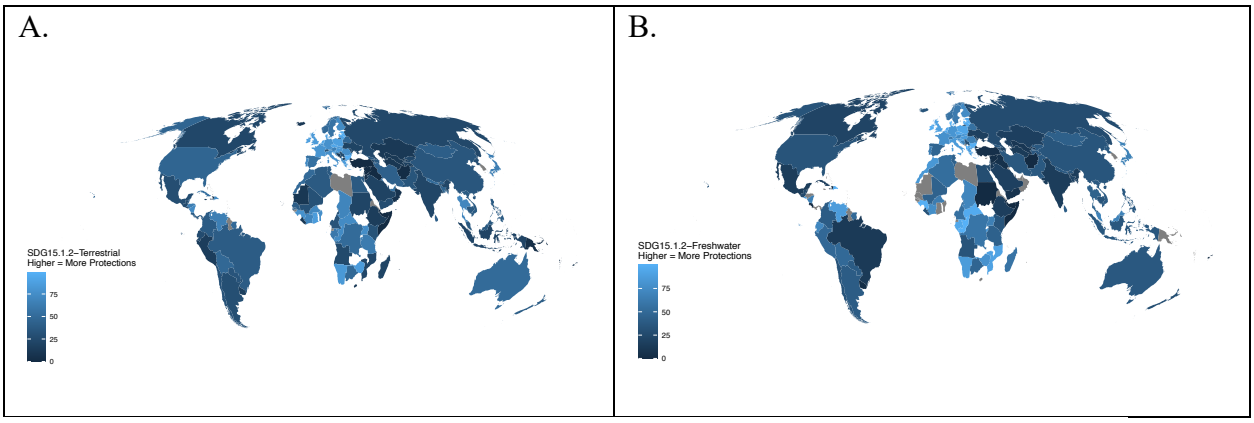

Fig S13. **Geographic variability in SDG15.1.2, protection of key terrestrial (A) and freshwater (B) biodiversity areas.** Lighter colors indicate more protection, darker colors indicate less protections. White indicates no data (on any outcome indicator) for that country, gray indicates the data are available for other indicators but just not for the current indicator). Already, we can see similar geographic patterning as the *Nature-Importance* associations, whereby nations in sub-Saharan Africa (and in Western Europe) have generally larger protections than most of Asia and North America.

The second factor in the SDG14 and 15 indicators was more centered on ocean and freshwater threats from trade and economy, with the highest-loading indicators on this second factor were: sdg14_biomar (threats to marine species due to imported goods and services), sdg15_biofrwter (threats to both terrestrial and freshwater species due to imported goods and services), and sdg14_fishstocks (percentage of a country's fishstocks that are over-exploited).

Analyses on all of these indicators are provided in the open data and code, as well as briefly summarized below in the results across all UN SDG indicators.

Regression outputs.

Next, we ask: how are the key environmental SDGs (focusing on those two indicators that emphasize biodiversity protections) related to the *Nature-Importance* associations? And, moreover, does *Nature-Importance* add significant explanatory value above and beyond the economic, linguistic and social covariates already explored?

First, covariate only models showed that covariates alone explained significant and meaningful variance for: sdg15_cpta (R^2 = 0.16*, Adjusted* R^2 = 0.12), sdg15_cpfa (R^2 = 0.20*, Adjusted* R^2 = 0.16). As reported in the open data, the significant covariates in the covariate-only models included: gender inequality (for sdg15_cpta), and internet users and population size of language speakers (for sdg15_cpfa).

Second, and most critical, adding *Nature-Importance* associations significantly improved model fit and the amount of variance explained, based on model comparisons using ANOVA for sdg15_cpta, $F(1, 137) = 8.49$, $p = .004$; and for sdg15_cpfa, $F(1, 111) = 9.74$, $p = .002$. We report the full model outputs for these two variables below. For both outcome variables, the *Nature-Importance* associations added another 4% (sdg15_cpta), and 7% of explained variance (sdg15_cpfa), and consistently with small-moderate effects: for sdg15_cpta, $\beta = .24$, $p = .004$, for sdg15_cpfa, $\beta = .27$, $p = .002$. The only other covariate that related to both indicators was Gini inequality which, as we discussed above, appears negatively related with biodiversity protections and environmental health because more industrialization and development brings both less economic inequality and more environmental degradation.

Table S8.1.
Multiple regression: Predicting SDG15.1.2 – Protection of Key Terrestrial Biodiversity Areas

	β	*b*	SE	t	p
Intercept		15.33	9.59	1.60	.11
Nature-Importance	**0.24**	**134.38**	**46.12**	**2.91**	**.004**
GDP 2022	0.02	0.46	2.08	0.22	.82
Gini 2023	*-0.22*	*-5.85*	*2.66*	*-2.20*	*.03*
Internet users	-0.13	-3.54	3.92	-0.91	.37
Gender Inequality	*-0.33*	*-9.18*	*4.54*	*-2.02*	*.04*
Wikipedia corpus size	0.03	0.84	3.64	0.23	.82
Speaker population size	-0.07	-2.00	3.86	-0.52	.60

$R^2 = 0.20$, *Adjusted* $R^2 = 0.16$, $F(7,137) = 5.03$, $p < .001$

Table S8.2.
Multiple regression: Predicting SDG15.1.2 – Protection of Key Freshwater Biodiversity Areas

	β	*b*	SE	t	p
Intercept		15.19	10.88	1.40	.17
Nature-Importance	**0.27**	**163.20**	**52.28**	**3.12**	**.002**
GDP 2022	-0.03	-0.71	2.20	-0.32	.75
Gini 2023	*-0.25*	*-7.03*	*3.11*	*-2.26*	*.03*
Internet users	-0.20	-5.97	4.36	-1.37	.17
Gender Inequality	-0.27	-7.94	5.32	-1.49	.14
Wikipedia corpus size	0.24	7.75	4.14	1.87	.06
Speaker population size	-0.26	-8.63	4.39	-1.96	.05

$R^2 = 0.27$, *Adjusted* $R^2 = 0.22$, $F(7,111) = 5.81$, $p < .001$

10. Correlation of Nature-Importance with Remote-sensed indicators: Additional details and regression models

For the secondary set of outcome variables – the satellite-derived and remote-sensed indicators – we also sought to streamline analyses (rather than discussing all 119 potential indicators). To do so, we took an entirely bottom-up approach. First, we examined the full set of bivariate (Pearson's correlation) correlations between *Nature-Importance* and all 119 potential indicators. These correlations ranged from slightly negative (r = -.23, with CHELSA_BIO_Temperature_Seasonality, indicating that higher temperature seasonality occurred in places with lower *Nature-Importance*) to moderately positive (r = .31, with EsaCci_BurntAreasProbability, indicating that higher probability of burned land occurred in places with higher *Nature-Importance*). However, it was notable that the majority of indicators (all reported in section 12 below) were not significantly correlated with *Nature-Importance.* Thus, we set a threshold of looking only at indicators whose absolute value correlation was greater than a standard "small" effect, i.e., $|r| > .20$. This threshold resulted in 13 indicators which, for simplicity, we have reproduced from the codebook here with their datasource and meaning (Table S9).

Table S9.
Remote-sensed indicators with more than small bivariate correlations to Nature-Importance, $|r| > .20$

Name	Meaning	Data source
CHELSA_BIO_ Isothermality	Essentially the amount of temperature stability across seasons, comparing day-to-night temperature variation against annual temperature variation. Higher values indicate more stability across seasons because the within day-to-night temperatures fluctuate more than seasonal temperatures. Higher values are more characteristic of places close to the equator (which therefore have low seasonal temperature changes) and more coastal places (again with water tempering extreme climate fluctuations). Measured in degrees Celsius.	http://chelsa-climate.org/bioclim/ and https://doi.org/10.1038/sdata.2017.122
CHELSA_BIO_ Mean_Temperature_of_Coldest_ Quarter	The lowest temperature of any monthly daily *mean* temperature, within the country's coldest quarter of the year. Higher scores indicate that, even in the coldest months (e.g., December, January in the Northern hemisphere), the place is relatively warm on average. Measured in degrees Celsius.	http://chelsa-climate.org/bioclim/ and https://doi.org/10.1038/sdata.2017.122

CHELSA_BIO_Min_Temperature_of_Coldest_Month	The lowest temperature of any monthly daily *minimum* temperature, within the country's coldest quarter of the year. Higher scores indicate that, even in the coldest months (e.g., December, January in the Northern hemisphere), the place is relatively warm, even at its minimum. Measured in degrees Celsius.	http://chelsa-climate.org/bioclim/ and https://doi.org/10.1038/sdata.2017.122
CHELSA_BIO_Temperature_Seasonality	Essentially the amount of temperature variability across seasons, measured as the standard deviation of the monthly mean temperatures. Higher values are more characteristic of places that are typically inland continental climates, like the Midwestern United States, Interior of Canada, or Northeast China. Measured in degrees celsius.	http://chelsa-climate.org/bioclim/ and https://doi.org/10.1038/sdata.2017.122
ConsensusLandCoverClass_Deciduous_Broadleaf_Trees	Percentage of the pixel area covered by deciduous broadleaf trees (e.g., Elm, Oak, Maple). Higher scores indicate that the country has more land covered by forests with deciduous broadleaf trees. Measured as average percentage across the country.	https://www.earthenv.org/landcover and https://doi.org/10.1111/geb.12182
ConsensusLandCoverClass_Mixed_Other_Trees	Percentage of the pixel area covered by mixed trees of various types. Higher scores indicate that the country has more land covered by mixed tree forests. Measured as average percentage across the country.	https://www.earthenv.org/landcover and https://doi.org/10.1111/geb.12182
EarthEnvTexture_CoOfVar_EVI	Enhanced Vegetation Index (EVI) quantifies the spatial heterogeneity or unevenness of a landscape's vegetation. It essentially measures the variation in "greenness" of the landscape. Higher scores would indicate that the country varies more in its vegetation (i.e., some areas are very green, other areas very barren).	http://www.earthenv.org/texture and https://doi.org/10.1111/geb.12365
EsaCci_BurntAreasProbability	Probability of an area experiencing a fire (largely a forest fire), detailed on a monthly frequency, based on observations over the 2001-2019 period. Notably, the probability of an area experiencing a fire is higher in densely forested areas, but also that those forests are more subject to catch fire because of causes such as human interaction (agricultural and industrial expansion), soil degradation, lighting and storms.	https://maps.elie.ucl.ac.be/CCI/viewer/download.php and ESA Land Cover CCI project team; Defourny, P. (2016): Centre for Environmental Data Analysis, https://catalogue.ceda.ac.uk/uuid/7c114fc6e2884c1f9ca107e7a502fdbf

IPCC_Global_Biomass	Global Biomass (from the International Panel on Climate Change Tier 1 Calculated Values) for the year 2000. This includes carbon stored in land both aboveground (i.e., in leaves, branches) and belowground (i.e., in soil). Higher biomass values indicate that the land stores more energy (carbon), an indication of a healthier ecosystem. Measured as tonnes of biomass carbon per hectare.	http://cdiac.ess-dive.lbl.gov/epubs/ndp/global_carbon/carbon_documentation.html and https://doi.org/10.15485/1463800
MODIS_NDVI	The Normalized Difference Vegetation Index (NDVI; i.e., the "greenness" of a place) across a 16-day average, averaged across all 16-day periods from 2015-2019. The NDVI is calculated as the difference between the near-infrared and red light reflectance from satellite images, since plants absorb those light waves differently. Higher scores indicate that the place is, on average, "greener". Values can range from -1 to +1, as the relative ratio of light reflectance.	https://explorer.earthengine.google.com/#detail/MODIS%2F006%2FMYD13Q1 and https://doi.org/10.5067/MODIS/MYD13Q1.006
SG_Silt_Content_015cm	Indicates the amount of silt in the soil at a depth of 0.15m. More silt in the soil means that the land can hold water and nutrients but is more prone to erosion. It is helpful for agricultural development, with places in the Middle East Fertile Crescent having particularly high silt content. Measured as the percentage of mass of the full soil sample.	https://www.isric.org/explore/wosis/accessing-wosis-derived-datasets and https://doi.org/10.1371/journal.pone.0169748
SG_Soil_pH_H2O_015cm	Indicates the acidity (pH < 7) versus alkalinity (pH >7) in the water pulled from the soil at a depth of 0.15m. Places with higher pH in soil are generally found in dry climates (e.g., Australia has alkaline soil in arid and semiarid regions). Places with lower pH in soil are generally found in humid places (e.g., Brazil, sub-Saharan Africa rainforest regions).	https://www.isric.org/explore/wosis/accessing-wosis-derived-datasets and https://doi.org/10.1371/journal.pone.0169748
SpawnEtAl_HarmonizedBGBiomass	Global belowground biomass carbon density for the year 2010. As with the IPCC values, biomass carbon density is taken as an indicator of a healthy ecosystem with high energy storage. Here, the indicator is only for belowground biomass, stored in soil (not in leaves, plants). Measured as the Mg of carbon per hectare	https://doi.org/10.3334/ORNLDAAC/1763 and https://doi.org/10.1038/s41597-020-0444-4

Regression outputs.

For each of these 13 indicators, we next replicated the regression approach that we did for the UN SDGs. Specifically, we first fit covariate-only models and then examined the additional contribution of *Nature-Importance* associations above and beyond those covariates. For 12 out of the 13 indicators (all except land cover by broadleaf trees), the *Nature-Importance* association was significant and meaningfully related, even after controlling for all covariates. The three indicators with the strongest incremental relationships beyond covariates were: (1) the probability of burned areas, with *Nature-Importance* explaining an additional 9% of variance (beyond the R^2 = .27 of the covariate-only model); (2) isothermality (i.e., temperature stability across seasons; *Nature-Importance* explained 6% additional variance, beyond the R^2 = .50 of the covariate-only model); and (3) the global aboveground and belowground biomass (*Nature-Importance* explained 7% variance, beyond the R^2 = .23 of the covariate-only model).

Table S10.1.
Multiple regression: Predicting probability of fires and burned areas (EsaCci_BurntAreas)

	β	*b*	SE	t	p
Intercept		-0.46	0.17	-2.69	.008
Nature-Importance	**0.32**	**3.56**	**0.82**	**4.35**	**<.001**
GDP 2022	0.002	0.001	0.03	0.02	.74
Gini 2023	-0.014	-0.007	0.05	-0.16	.72
Internet users	*-0.32*	*-0.18*	*0.07*	*-2.61*	*.01*
Gender Inequality	0.15	0.08	0.08	1.05	.43
Wikipedia corpus size	0.07	0.04	0.06	0.59	.78
Speaker population size	-0.02	-0.009	0.07	-0.13	.94
R^2 = 0.36, *Adjusted* R^2 = 0.33, $F(7,143) = 11.45$, $p < .001$					

Table S10.2.
Multiple regression: Predicting isothermality (CHELSA_BIO_Isothermality)

	β	*b*	SE	t	p
Intercept		274.92	45.19	6.08	<.001
Nature-Importance	**0.25**	**923.32**	**218.54**	**4.23**	**<.001**
GDP 2022	-0.09	-15.71	9.85	-1.60	.11

Gini 2023	*0.24*	*41.94*	*12.52*	*3.35*	*.001*
Internet users	-0.10	-18.35	18.45	-1.00	.32
Gender Inequality	*0.37*	*67.48*	*21.33*	*3.16*	*.002*
Wikipedia corpus size	0.20	36.42	17.13	2.23	.04
Speaker population size	0.03	5.27	18.15	0.29	.77
R^2 = 0.56*, Adjusted* R^2 = 0.54, $F(7,144) = 25.83$, $p < .001$					

Table S10.3.
Multiple regression: Predicting biomass (above and belowground; IPCC_Global)

	β	*b*	SE	t	p
Intercept		-955.22	1486.43	-0.64	.52
Nature-Importance	**0.28**	**27140.07**	**7188.15**	**3.78**	**<.001**
GDP 2022	-0.02	-73.45	323.82	-0.23	.82
Gini 2023	0.13	600.77	411.79	1.46	.15
Internet users	0.14	645.51	606.90	-1.06	.29
Gender Inequality	0.42	*2026.19*	*701.62*	*2.89*	*.004*
Wikipedia corpus size	0.37	*1740.65*	*563.38*	*3.09*	*.002*
Speaker population size	-0.17	-813.51	596.93	-1.36	.18
R^2 = 0.30*, Adjusted* R^2 = 0274, $F(7,144) = 8.97$, $p < .001$					

11. Correlation of Nature-Importance with United Nations Sustainable Development Goals: All SDG indicators

In the main text, we focus on the UN SDG correlations only for SDG14 and SDG15 which are the key environment-related indicators. However, researchers may also be interested in the broader sample space of how countries might trade-off between various UN SDGs (as hinted at by the fact that African nations have both high environmental achievement but also low social welfare achievement on poverty indicators). To that end, we repeated our primary analyses correlating UN SDGs with *Nature-Importance* associations but across all 83 UN SDG outcomes with available data by country (Figure S14).

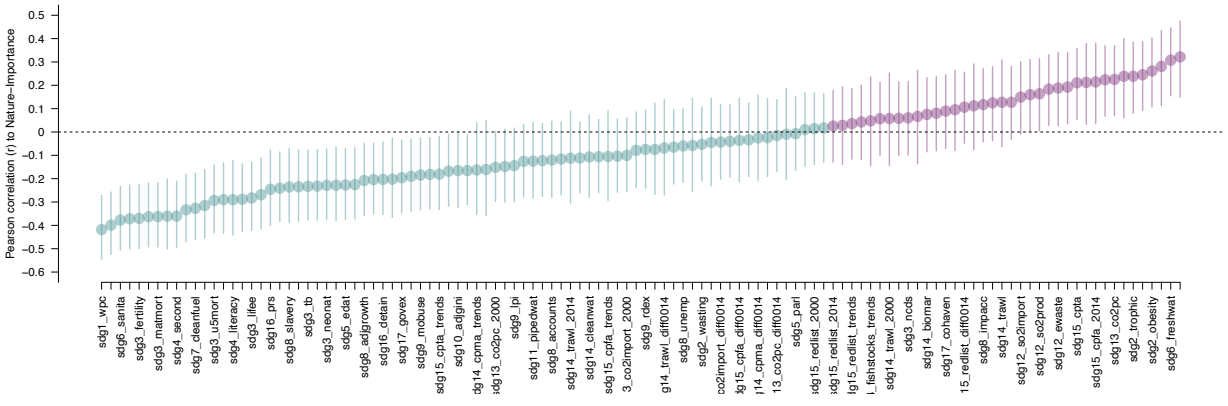

Fig S14. Bivariate correlations of *Nature-Importance* across all 83 UN SDG outcomes. Purple dots indicate positive correlations, blue dots indicate negative correlations; error bars represent 95% confidence intervals.

Results across this broader sample space provide two important conclusions. First, they reinforce that environmental indicators are consistently among the most positive correlates (almost all purple correlations in Fig S14 have to do with environmental outcomes such as freshwater resources or even the consumption of meat, which would imply high agricultural demands). Second, the additional results also highlight a few particularly informative negative correlations for discriminant validity in showing what is *not* related to *Nature-Importance*.

Specifically, as hinted in the above results focusing on GDP and poverty across continents, we found negative bivariate correlations for SDG1 (on *sdg1_wpc*, an indicator of poverty), r = -.42 [-.55, -.28], $t(140)$ = -5.45, p <.001. Again, overall, countries with more poverty (especially in sub-Saharan Africa) had stronger *Nature-Importance* associations. Similar negative correlations were found for other poverty-related indicators including SDG7 (electricity access, r = -.40) and SDG6 (sanitation, r = -.38, and clean water access, r = -.37).

12. Correlation of Nature-Importance with Remote-sensed Environmental Variables: All remote-sensed indicators

In the main text, we focus on only a small subset of positive correlations with remotely-sensed environmental health. Here, we show the range of correlations across all indicators that are not redundant with GDP or population size (N = 106 indicators; Figure S15).

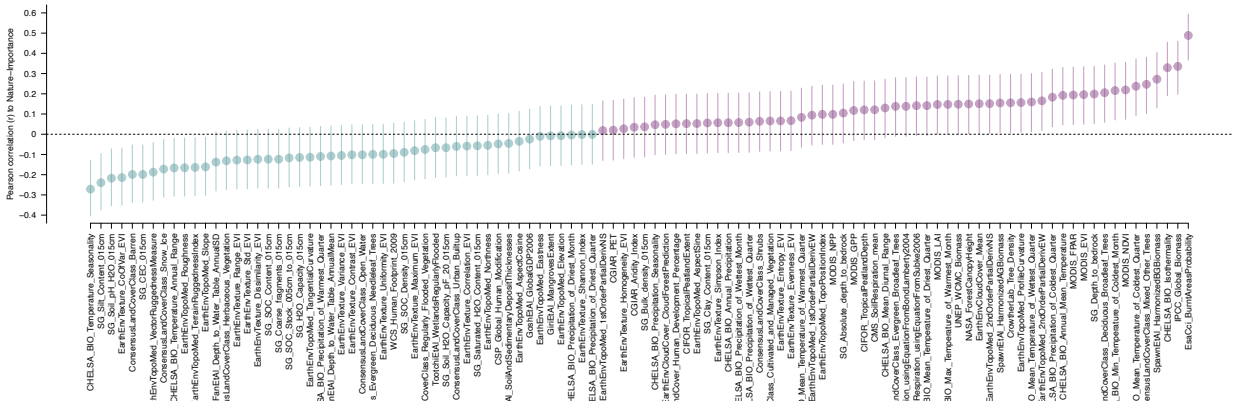

Fig S15. Correlations of *Nature-Importance* across all remotely-sensed environmental outcomes. Purple dots indicate positive correlations, blue dots indicate negative correlations; error bars represent 95% confidence intervals. Codebook for all variable names can be found in the open code and data on the OSF.

Looking across this broader sample space of indicators, it is noteworthy that the strongest positive correlates were typically found with seemingly more malleable (or at-risk) indicators (such as biomass and burn risk), rather than more permanent features (such as topography). However, another theme in positive correlates was that they related to more accessible and reliable nature-based resources, e.g., more stable yearly temperatures and high amounts of vegetation (e.g. tree cover, biomass). In contrast, negative correlates relate to more harsh environments with a higher variability in yearly temperature and water availability, as well as greater ruggedness or barrenness of the environment.

In fact, this interpretation of the results may suggest that part of the reason that *Nature* is discussed as *Important* in places like Uganda, Congo, and the Central African Republic is because nature's resources can be accessed; by contrast, *Nature* may not be associated with *Important* in harsher places like Nepal, Bhutan, or Russia because nature's resources may be seen as less accessible and reliable. This result dovetails with social science research showing that accessible nature (e.g., green spaces, forests) sometimes activates more positive attitudes and pro-environmental behaviors towards nature[5]. Future work may use more granular geographic variation (e.g., local newspapers) to investigate how within-country variation in ruggedness and accessibility to nature may help explain the global variation in *Nature-Importance* associations.

13. Correlation of Nature-Importance with Environmental Attitudes from 63 Countries

A primary aim of the current manuscript is to expand the global measurement of human attitudes and collective representations. In the analyses above we have shown that this newly-mapped global variation in *Nature-Importance* associations indeed correlates with consequential environmental outcomes, including UN SDGs and remotely-sensed indicators of ecological health (particularly forest health). Such results emphasize the strong external and real-world validity of the language measures.

However, there may still be a question of whether the language measures are capturing meaningful human psychology and attitudes or, instead, whether they are just capturing descriptions of the surrounding environment (e.g., descriptions of the prevalence of forests). To address this question and provide more validation we therefore test the correlation between the language measures and new data from the International Climate Psychology Collaboration (ICPC)[6], providing data from 59,503 participants in 63 countries (collected July 2022 - July 2023).

We focus on country-level averages of key outcomes of climate change beliefs (e.g., "*Taking action to fight climate change is necessary to avoid a global catastrophe.*"), support for climate change policies (e.g., *"I support protecting forested and land areas"*) and identities around climate change (e.g., *"How interested are you in reducing your carbon emissions?"*). If the current language estimates are capturing meaningful psychological representations then we should expect positive correlations between greater concern about climate change, support for policies addressing climate change, and stronger interest and identity in fighting change.

Indeed, across 38 belief outcomes, we found that the average country-level correlation was positive and small-to-moderate in size, $r = 0.22$ [-0.03, 0.39]. The strongest correlation (Figure S16A) emerged for the average environmental identity (e.g., seeing oneself as someone who cares about the environment), $r = .33$ [.08, .53], $t(60) = 2.68$, $p = .009$, with similar small-to-moderate correlations for indicators including support for emission reductions and beliefs in the crisis of climate change (Figure S16B-D). The few neutral (weakly negative correlations) were for perceptions of the average country-level support for reducing carbon footprints (e.g., *"How many Americans do you think make an effort towards reducing their carbon footprint?"*, $r = -.02$ [-.27, .23], $t(59) = -0.16$, $p = .87$).

Overall, however, results are clear: language representations across countries are meaningfully and robustly correlated with the average climate attitudes measured on surveys across countries. Although, until now, the ICPC data had the greatest global coverage to-date (even including a handful of countries in Africa and South America) the current linguistic approach has a clear advantage of drastically expanding the scale of assessing collective representations and psychological constructs in languages spoken around the world.

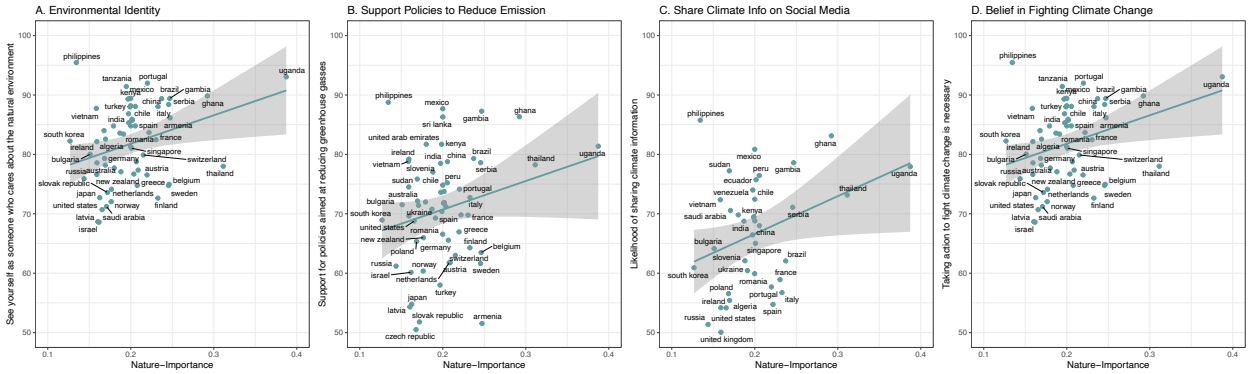

Fig S16. Correlations of contemporary language representation (*Nature-Importance*) with environmental attitudes. All plots have the same x-axis, indicating the *Nature-Importance* associations from weighted estimates across countries. Y-axes show the average score on each environmental attitude item (e.g., environmental identity) within each of the 63 surveyed countries. Blue line indicates the correlation, shaded area represents the 95% confidence interval estimate around the correlation.

14. Comparing Nature-Importance Attitudes Over Surveyed Environmental Attitudes

Just as we did above with the contribution of *Nature-Importance* language-based attitudes, we can ask: do the surveys of environmental attitudes from the International Climate Psychology Collaboration add explanatory value above and beyond the covariates? And even beyond *Nature-Importance* attitudes from language? Or are the Nature-Importance attitudes unique in being able to add further explanatory value beyond covariates? As above, we compared covariate-only models (including *Nature-Importance* attitudes, GDP, Gini, Gender Inequality Index, Internet users, Wikipedia size, and speaker population) to the full model that also included the country-average of environmental attitudes, merged from all survey variables. Here we focused only on the two SDG 15.1.2 indicators of protection for freshwater and terrestrial biodiversity areas.

Results showed that surveyed environmental attitudes did not add significant explanatory value beyond the *Nature-Importance* and covariate combinations: for terrestrial protection $F(1, 50) = 0.24$, $p = .63$; and for freshwater protection, $F(1, 46) = 0.05$, $p = .83$. That is, the combination of covariates and *Nature-Importance* were sufficient to explain approximately 32% of variance in terrestrial protection and 43% of variance in freshwater protections across the subset of included countries; the addition of surveyed attitudes did not increase the R^2 value for either of the models. Additionally, whereas the standardized beta effect size for *Nature-Importance* in these models was $\beta = 0.19$ for terrestrial protections and $\beta = 0.10$ for freshwater protections, the parallel effect sizes for surveyed attitudes were substantively smaller, $\beta = 0.08$ and $\beta = 0.03$, respectively, neither of which were significant, $p > .63$. In summary, the advantage of including language based attitudes (e.g., *Nature-Importance* representations) appears to not

only be its expanded global coverage, but even within the same countries, they may be better able to capture and explain the variation in environmental protection and health.

15. References for Appendix

1. Pennington, J., Socher, R. & Manning, C. D. GloVe: Global vectors for word representation. in *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference* 1532–1543 (2014). doi:10.3115/v1/d14-1162.
2. Hamilton, W. L., Leskovec, J. & Jurafsky, D. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 1489–1501 (2016).
3. Charlesworth, T. E. S., Caliskan, A. & Banaji, M. R. Historical representations of social groups across 200 years of word embeddings from Google Books. *Proc Natl Acad Sci U S A* 119, e2121798119 (2022).
4. Garg, N., Schiebinger, L., Jurafsky, D. & Zou, J. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proc Natl Acad Sci U S A* 115, E3635–E3644 (2018).
5. Martin, L. *et al.* Nature contact, nature connectedness and associations with health, wellbeing and pro-environmental behaviours. *J Environ Psychol* 68, 101389 (2020).
6. Doell, K. C. *et al.* The International Climate Psychology Collaboration: Climate change-related data collected from 63 countries. *Scientific Data 2024 11:1* 11, 1–17 (2024).

47. Rad, M. S., Martingano, A. J. & Ginges, J. Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proc Natl Acad Sci U S A* **115**, 11401–11405 (2018).
48. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. in *Proceedings of North American chapter of the Association for Computational Linguistics-Human Language Technologies 2019* 4171–4186 (2018).
49. Bender, E. M., Gebru, T., Mcmillan-Major, A., Shmitchell, S. & Shmitchell, S.-G. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* https://doi.org/10.1145/3442188.3445922 (2021) doi:10.1145/3442188.3445922.
50. Hamilton, W. L., Leskovec, J. & Jurafsky, D. Diachronic word embeddings reveal statistical laws of semantic change. in *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers* vol. 3 1489–1501 (2016).
51. Bojanowski, P., Grave, E., Joulin, A. & Mikolov, T. Enriching Word Vectors with Subword Information. *Trans Assoc Comput Linguist* **5**, 135–146 (2017).