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
King Abdullah University of Science and Technology

Version: June 18, 2026

DRAFT

Please do not quote or distribute without permission.

Abstract

When people talk and write about nature, they provide a window into how their culture represents the natural world, whether they see it as important, worthy of concern and protection. Such collective beliefs may be a key driver of global environmental health or degradation, yet no study to date has been equipped to map cultural representations of nature at scale. This research aims to quantify global variation in linguistic representations of nature and explore implications for sustainable development indicators. The authors use Natural Language Processing (multilingual word embeddings) built from Wikipedia text for 120 languages spoken in 189 countries, and from European parliamentary speeches for 28 languages in 27 countries. In English, the concept of *Nature* is moderately but robustly associated with *Importance* (alongside other dimensions, e.g., *Concern, Protection*), and such associations increased over 200 years. Yet, examining other languages revealed that Niger-Congo and Austronesian languages (mostly spoken in Global South nations) express the strongest *Nature-Importance* associations whereas English and Indo-European languages (generally spoken in the Global North) express the weakest. Robustness tests within 27 European countries reinforced that such country-level variation of *Nature* representations are consistent across text sources, time (2015-2022), and speaker demographics. Most critical, the robust global variation is significantly correlated with Sustainable Development Goals, including protection of water and land biodiversity areas, even explaining variance over economic covariates. In addition to introducing new methods and empirical records for mapping nature representations worldwide, the results underscore the importance of non-English perspectives in guiding policy and research on environmental protection.

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]. Work from across the social and environmental sciences suggests a critical role for *culture-level* values and representations in shaping anthropogenic impacts (such as biospheric values[5–7], relational values and nature connectedness[8–11], and prioritizing nature as *important*). Beliefs of individual people clearly matter for individual behaviors like environmental donations[12,13], recycling or water usage[14]. But the global magnitude of environmental degradation is more likely to be the emergent property of culture-level *collective representations*—shared social beliefs and values communicated and constructed through language and maintained through social pressure[15,16]. Yet, until now, systematic global data on how nature and the environment are collectively represented has remained elusive.

Although there are many international surveys of environmental attitudes[8,17–23], most remain focused on Global North countries, underrepresenting the true global variation in ecosystems and human-environment relations. For instance, a recent audit of environmental values research found that 96% of papers were written in English[24], which suggest a focus on English-speaking perspectives and its correlated focus on Western, Educated, Industrialized, Rich, Democratic (WEIRD) populations[25]. This has led to urgent calls for increasing multilingual coverage of environmental research to better expand to populations and emic values from other cultures[26–29]. Even the few surveys that go beyond the Global North[23] remain limited either because they rely on proprietary data (with prohibitive costs for most researchers) and/or on self-reported survey measures (which can be vulnerable to self-presentation[30]). This project aims to overcome these

limitations using open-source multilingual data from 189 countries to naturalistically examine the cultural representations of nature.

We do so by relying on the rapidly-growing tools of natural language processing (NLP) which have opened up new avenues to study cultural representations at-scale. Already, NLP tools—especially word embeddings[31,32] which represent word meaning as quantitative vectors based on word-to-word co-occurrence—have been used to study representations of race, gender, and social class in books, Internet text, newspapers, and more[33–35]. Such language-based representations (e.g., associating *women-home/men-office*, or *white people-good/black people-bad*) correlate robustly with measures of implicit attitudes[36–38], including across countries[39] and across history[40,41]. Yet, to date, no study has extended such NLP applications to identify nature-related representations, nor to evaluate variation in representations across the globe.

Four guiding questions of nature representations in language and culture.

Using NLP, we are able to address four long-standing questions spanning the social and environmental sciences. First, *how* is nature represented in everyday English language? What are the dimensions (e.g., positivity/negativity, protection, concern, and importance) that arise as most strongly and consistently associated with nature in everyday text? In answering these questions, we can test the relative strength of many dimensions suggested (but rarely directly compared) in prior research.

Second, how have the associations with such dimensions *changed*? Numerous theories, such as post-materialist values theory[42] (see critiques e.g.,[43]) suggest that nature will gain importance when people feel security in their immediate needs (for food, shelter) and can therefore embrace more abstract values such as environmentalism. So, as populations in the English-speaking world

have become wealthier and more secure over the past 200 years, we might expect that language associations of nature with dimensions of value (i.e., importance) increased since the 1800s.

Third, how do such associations *vary across languages and cultures*? Many perspectives[5,6,11] agree that cultures with different economic structures and ecologies will also have different values and relations to nature. For instance, some perspectives suggest that richer Global North countries will have stronger associations of nature with importance (whether because of post-material values[42] or other reasons[44]). Other perspectives suggest that greater subsistence relationships with nature, as seen in the Global South and poorer nations, may lead to a prioritization and protection of nature[45,46]. And yet other theories combine both ideas to argue for a curvilinear relationship between wealth and environmental values[47,48]. Evidently, examining a global scale of cross-cultural variation will help us understand *where* nature is prioritized today.

Fourth and finally, we ask: does such cultural variations *matter* for environmental protections and health? Theoretically, attitudes are expected to translate to corresponding behaviors[12,13,49]; likewise, at a collective level, representations of nature as important should translate to collective behaviors and policies that protect that nature. We therefore test the link between language-based representations and nation-level indicators of environmental health from remotely sensed (e.g., tree cover from satellite data) and self-reported sources (UN Sustainable Development Goal reports). Moreover, if collective representations play a practically meaningful role, we expect that any relationships will persist over other economic covariates and self-reported surveys.

Text sources to examine naturalistic representations of nature across languages and history

To test these questions we draw on four complementary data sources of naturalistic text. For initial explorations, we examine English representations of nature in 840 billions of words of naturalistic contemporary Internet text[50]. For change over time, we examine 200 years of English

book texts[51]. For questions of cross-cultural variation and correlations, we used multilingual word embeddings built from Wikipedia text written in 120 languages spoken across 189 countries (covering 98% of UN member countries). Wikipedia text is not a representative sample of the population: audits of the millions of volunteer Wikipedia contributors[52] suggest that most contributors are men (~80% of contributors in 2019-2024), and more educated than the average population. Still, these biases in the sample actually align with an interpretation of Wikipedia embeddings as reflecting generally elite, dominant narratives which, in turn, aligns with our goal of measuring *collective representations* that are also created as dominant narratives.

Wikipedia has issues, of course: norms of contribution may vary across cultures (e.g., people in individualist cultures are more likely to clarify and add information[53]) and representation may also vary according to Internet access (which we control for in our models). Furthermore, while most non-English Wikipedia pages (e.g., German) are written and edited by people with nationalities corresponding to those languages (e.g., Germans in Germany), there is also likely cross-cultural blending: German-speakers from Austria, Switzerland, or even the US may write and edit German Wikipedia pages. Thus, Wikipedia text—despite having the critically important contribution of a global scope and naturalistic contribution format—is still limited in not providing a clear one-to-one mapping between language and country.

Although we attempt to address this latter concern using country-level weighting (see *Methods*), we also addressed it more directly by including a complementary fourth dataset of parliamentary speeches[54] (2015-2022) available in 28 languages spoken in 27 European countries. Despite the more limited geographic coverage, these data reinforce any interpretations because they are directly mapped to each country (rather than inferred, as in Wikipedia). Additionally, the parliamentary data provides demographic data on speaker gender, age, and

political ideology, allowing us to examine whether country-level variation is consistent across demographic groupings.

Results and discussion.

How is nature represented in everyday Internet text?

We first discover the representation of nature in contemporary English word embeddings (from 840 billion words of Internet text; pre-trained GloVe word embeddings[50]), and embeddings from the 5 most widely spoken non-English languages of French, Spanish, Arabic, Bengali, and Chinese (from Wikipedia text; pre-trained *fastText* embeddings[31]). We identified the top 50 nearest neighbors (the 50 words that had the highest average cosine similarities) to a list of seed words representing the concept of *Nature* (*nature, climate, environment, land;* see *SI Appendix* for leave-one-out-cross-validation to show robustness across word choices).

The results showed clear face validity. Nearest neighbors were consistently centered on nature-related concepts, in both English Internet text (Figure 1A) and in the 5 non-English languages (*SI*). For instance, exploratory factor analysis on the word embeddings of the top 50 nearest neighbors showed that the *Nature* concept referred to face-valid latent factors (Figure 1B), including: wilderness/ecosystems (e.g., *woodlands, jungle, savanna*); agriculture/farming (e.g., *agriculture, husbandry, cultivation*); and culture/human-nature interactions (e.g., *society, culture, policy*).

Next, we identified the strength of association between these diverse nature concepts and five semantic dimensions central to past nature attitude research, specifically: *(1) positivity* and *(2) negativity* (foundational dimensions of evaluative attitudes[55]) *(3) concern* and *(4) protection,* (two widely-used dimensions in sustainability attitudes[56–58])*;* 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[59,60]. These five dimensions are just a starting point; the flexibility of the current methods can enable future work to test other values (e.g., *biospheric values*[5,61]) and constructs (e.g., *nature connectedness*[8,10]).

As expected, the general *Nature* concept (and its five latent factors identified from the EFA) was positively associated with all five semantic dimensions (Figure 1B). However, the dimension of *Importance* was descriptively stronger and more robustly associated than other dimensions (comparisons in *SI Appendix*). Additional benchmarking comparisons showed that the *Nature-Importance* associations were meaningfully stronger than other evaluative associations (e.g., *Flowers-Good, Insects-Bad;* Caliskan et al., 2017) or social associations (e.g., *Women-Arts, Men-Science*; Charlesworth et al., 2022). We therefore focus on this *Importance* dimension to streamline all subsequent reporting (*SI* has all other dimensions). Notably, in classic attitude research, *importance* is a key variable that indexes the prioritization, strength, centrality, and consequentiality of an attitude object[59,60]. At a societal level, we interpret the collective representation of *Nature-Importance* as a broad evaluative signal about nature prioritization, determining whether environmental considerations are even addressed in public discourse or policy.

How have Nature-Importance associations changed over historical English?

Across 200 years of English books (using pre-trained *word2vec* embeddings from Google Books across 1800-1990[51]), *Nature-Importance* showed the largest increases over time of any dimension, tripling in magnitude from essentially no association in 1800 $M = 0.04$ to a small 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 $< .001$). There were no significant changes for *Nature-Positive* or *-Negative (rho*

= -.17, -.36, respectively, *p*s > .12). This aligns with perspectives that environmentalism may rise alongside increases in material security over the long history of 200 years[42] (despite possible stability since ~1990[44]).

[Figure 1]

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).

How do Nature-Importance associations vary across languages and cultures?

Cross-linguistic variation: Wikipedia text. We next turned to variation across 120 languages using data from pre-trained multilingual *fastText* embeddings on Wikipedia text[31]. 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 in Wikipedia. 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 languages (all those that had available professional translators from the 120 languages); and (2) back-translated all automated and professional translations manually in Google translate to identify errors. On average, the accuracy for our key word lists of *Nature* showed 72-85% of languages had identical translations word-for-word across the automated and native speaker translations (similarly, "important" had identical translations for 77% of languages). Translation "misses" were most often due to near-synonyms or to alternative spellings across languages (see

SI). Thus, we infer that the automated translations provide reasonable accuracy for understanding cross-lingual representations of nature concepts.

Averaging across all languages, we found that *Nature* was moderately associated with *Importance* ($M = 0.27$, $SD = 0.22$) though the strength of *Nature-Importance* associations varied across languages (Figure 2). Compared to Indo-European languages ($M = 0.20$), *Nature-Importance* associations were significantly stronger in Niger-Congo languages ($M = 0.65$), $\beta = 0.60$, $p < .001$, and Austronesian languages ($M = 0.40$), $\beta = 0.28$, $p = .001$ (full model in *SI*). English was particularly notable in having relatively weak *Nature-Importance* associations ($M = 0.16$), lower than 71% of all other languages across the dataset. We ruled out concerns that stronger associations may emerge only because of small sample sizes, whether smaller speaker populations or sample sizes of text (i.e., smaller datasets in Wikipedia). The strength of *Nature-Importance* associations was not correlated with either (1) the number of speakers of a language, *rho* = -.04, $p = .69$; or (2) sample size of Wikipedia text, *rho* = -.17, $p = .09$ (see *SI Appendix*).

[Figure 2]

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, Māori*) are smaller circles, while larger speaker populations (e.g., *English, Chinese, Serbo-Croatian)* are plotted with larger circles.

Robustness test with an alternative NLP approach: Topic classifications of Wikipedia titles. Because capturing cross-linguistic and, by extension cross-cultural, variation is a central contribution of this paper, we wanted to ensure robustness of our conclusions using a complementary NLP tool: topic classifications. Specifically, we used a dataset[62] that had classified all Wikipedia titles (as of 1 January 2021) from all languages into 64 topics

(determined by Wikiprojects[63]), one of which was *Earth and Environment*. For each language, we therefore compared the probability of the *Earth and Environment* topic compared to the probability of the other 63 topics (e.g., *Geography: Western Europe; Culture: Biography;* see *SI Appendix*). Across languages, the odds of the *Earth and Environment* topic was significantly correlated with the strength of *Nature-Importance* associations from word embeddings, $r = .27$ [.08, .44], $t(101) = 2.82$, $p = .006$. Thus, even across very different approaches, *Nature-Importance* associations from word embeddings appear to capture a robust signal of how each language discusses and prioritizes the environment.

From language to cultural variation: Wikipedia text. Drawing on this robust linguistic variation, we estimated country-level *Nature-Importance* associations for 189 countries (countries that spoke one of the 120 languages). Specifically, we generated weighted national averages of *Nature-Importance* associations by averaging the *Nature-Importance* scores for the first four primary languages of each country, weighted by the population speaking each of these four languages (*Methods* and *SI Appendix*). For validation, we further ensured that these nation-level language-based results were indeed significantly and positively correlated with international survey data from nearly 60,000 respondents across 63 countries[64] (*SI Appendix*). The positive, significant correlations (with an average $r = 0.22$ across 38 belief outcomes; *SI*) established that language analyses can help expand the global map of nature representations.

Nation-level averages (Figure 3) 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 no North American countries were in the top quartile of *Nature-Importance* associations.

[Figure 3]

Figure 3. **Variation in Nature-Importance representations across global countries in Wikipedia text.** Weighted estimates of *Nature-Importance* associations across 189 countries inferred from the top four languages spoken in each country based on Wikipedia text associations for 120 languages. Darker green indicates that the country is in the top quartile (strongest *Nature-Importance* associations), darker red 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 *Importance* dimension (i.e., the importance synonyms were not available in the languages of those countries; N = 3 countries).

Cultural variation: Tests in European parliamentary speeches. Wikipedia text is advantageous because it provides a large sample of languages (mapped to countries) and reflects relatively naturalistic contributions that could be written by any speaker. But the text is limited because we had to infer country-level estimates (based on languages spoken in each country) and we do not know the demographics of contributors to each Wikipedia language. Thus, for robustness, we examined a smaller set of data from the ParlaMint corpus, covering 27 European countries speaking 28 languages across 8 years (2015-2022), as well as four demographic contrasts (gender: men/women, ideology: left/center/right, age generation: baby boomers/genX/millennials, and government status: party-in-power/opposition). Because these parliamentary texts were smaller than Wikipedia (especially when divided by speaker demographics), we adopted a different word embedding approach—à la carte (ALC) embeddings[65]—recommended for multilingual inference from small corpora (see *Methods*).

On average, we found a significant positive relationship between *Nature* and *Importance* in European parliamentary speeches, $b = 0.02$, SE = 0.003, $t\ (27.13) = 5.97$, $p < .001$ (note that the means are not directly comparable to the Wikipedia text because of the different embedding approach; but the relative differences of country-level variation remain informative). Critically, this country-level variation in *Nature-Importance* associations was significantly and positively

correlated between Wikipedia and parliamentary texts, r = .37 [.002, .66], t (26) = 2.06, p = .05. While this correlation is only moderate in magnitude, it is a strong and conservative test given the number of methodological differences, and the small sample of (only European) countries.

Next, we explored whether demographic groups across European countries had different *Nature-Importance* associations. Nearly all demographic groups had similar and positive means of *Nature-Importance*, M range = [0.011, 0.023], and paired samples t-tests showed no significant differences by political ideology, age generation, or government status, all $|t|$s < 1.99, all ps > .06. Only gender differences emerged: men politicians (M = 0.016) showed significantly weaker *Nature-Importance* associations than women politicians (M = 0.023), t(28) = -4.03, p <.001, d = 0.75. More critical, however, we found that the country-level variation was consistent regardless of demographic groups (Figure 4A). Countries that had strong Nature-Importance associations among men's speeches also had strong Nature-Importance associations among women, r = .81 [.63, .91], t(27) = 7.22, p < .001, with similar correlations for ideology, generation, and government (Figure 4A), rs range [.65, .74], all ps < .001. What is reflected in the "culture" (e.g., through Wikipedia) is also revealed consistently within different demographic groups of speakers. Such cross-demographic consistency aligns with our definition of nature representations as *collective* representations[15] that are shared across many people in a culture.[62]

[Figure 4]

Figure 4. **Variation in Nature-Importance representations across European countries by speaker demographics (A) and by time (B) in Parliamentary speeches.** Panel A indicates four demographic contrast by gender, ideology, age generation, or government) with countries ranked on the x-axis according to their average Nature-Importance association across all contrasts. Countries are listed first, followed by the language spoken in that country (note that Spain has four entries for Spanish, Basque, Catalan, and Galician). Countries listed on the left (e.g., Bosnia, Spain) have the weakest average associations; countries on the right (e.g., Sweden, Finland) have the strongest associations. Panel B indicates the average *Nature-Importance* associations across time by country. Darker green indicates stronger associations, lighter green indicates weaker associations; hashed shading indicates the relationship is negative. Countries are ordered from top-to-bottom from weak (and negative) associations in Bulgaria to the strongest positive average associations in Sweden and Norway. White in either plot indicates no data was available.

Finally, in parliamentary speeches, we examined the effect of time overall (Figure 4B) and in each demographic groups (*SI Appendix*). Using a mixed effects linear regression with random intercepts for countries, we found that the *Nature-Importance* associations remained stable the 8 years of data (2015-2022), β = -0.0018 [-0.09, 0.09], t (200.3) = -0.038, p = .97. Furthermore, in models including interactions of time and demographic groups, we found no significant interaction between time and group, $F(9, 2000) = 1.67$, p = .09, meaning that all ten groups have similarly remained largely stable. Taken together, the results underscore that country-level variation in *Nature-Importance* associations from Parliamentary discourse are generally strong, demographically-robust, and temporally-persistent.

Do Nature-Importance associations matter? Correlations with environmental protections and health. We used two sources of data on environmental health. First, the 2020 United Nations Sustainable Development Goals (SDGs), specifically focusing on indicators for climate-relevant goals of SDG 14 (Life Below Water) and 15 (Life on Land), as these represent the gold-standard for policy discussions around the environment. For 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 (top-loading indicators on the first factor in an EFA; see *SI Appendix*). All other SDG14 and 15 indicators are reported in the *SI.* Second, to complement self-reported indicators, we used a large dataset of 109 indicators from remote-sensing satellite data, including tree cover[66] and forest biomass[67] (see *SI Appendix*). For parsimony, we first identified 13 indicators that showed significant zero-order correlations with *Nature-Importance* across countries*,* then examined whether the relationships persisted beyond covariates.

Next, we used hierarchical regressions to examine macro-level relationships with environmental health (following similar approaches[68]) and test whether *Nature-Importance* explained variance beyond six nation-level covariates hypothesized to relate to environmental health: (1) Gross Domestic Product (GDP), with research suggesting possible curvilinear relationships with low GDP countries having environmental health as pre-extraction economies, but high GDP countries having environmental health because they are post-industrial and divest from extraction[47,48,69]; (2) Gini inequality[70], for similar reasons as GDP; (3) Gender Inequality Index[71], a proxy indicator for a country's development in social issues, which may trade-off with environmental issues. We also included three covariates to rule out methodological confounds: (4) Internet use[72]; (5) Wikipedia data size[73,74]; and (6) speaker population.

Even after controlling for these six economic and linguistic covariates, we found persistent small-to-moderate relationships between *Nature-Importance* representations in Wikipedia text and SDG15.1.2, for the protection of freshwater biodiversity areas (Figure 5A-B), $\beta = 0.24$, $t(137) = 2.91$, $p = .004$, and protection of terrestrial biodiversity areas, $\beta = 0.27$, $t(111) = 3.12$, $p = .002$. Similarly, remotely-sensed indicators revealed persistent relationships with *Nature-Importance,* even beyond covariates (Figure 5C-D). The strongest correlation was observed with the estimated probability of an area experiencing a forest fire (essentially a proxy for more forested areas, and more human-forest interaction[75]), with greater *Nature-Importance* scores associated with higher probability of potential fires, $\beta = 0.32$, $t(143) = 4.35$, $p < .001$. The second-strongest correlation was with biomass carbon storage (essentially a generalized indicator of ecosystem health[67]), which increased with *Nature-Importance* scores $\beta = 0.28$, $t(144) = 3.78$, $p < .001$. A further 11 remote-sensed indicators also showed persistent and significant relationships after covariates (reported in *SI Appendix* and open data), with higher *Nature-Importance*

characterizing places with high forest cover, vegetation, lower silt soil content, and more stable temperatures. Notably, all of these were ecological features that largely characterize sub-Saharan Africa nations, and that also reflect generalized environmental health (e.g., biomass and forest-human interaction are taken as broad indicators of the overall health of an ecosystem).

Critically, the contribution of *Nature-Importance* suggested a meaningful as an effect size. Model comparisons against covariate-only models showed that *Nature-Importance* explained an additional 4-9% of country-level variance (full models in *SI*). The effect magnitude is on-par with the contribution of Gini inequality. Thus, *Nature-Importance* representations in language may be at least as critical as economic inequality for understanding global variability in which countries engage in biodiversity protections and have environmental health.

[Figure 5]

Figure 5. **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.

Using a language-based approach we advanced understanding of how humans represent nature, not only in the Western world today (as most previous studies have done[21,76]), but across nearly the entire globe (189 countries) and even back 200 years in history. In English, the concept of *Nature* is robustly associated with *Importance*, alongside other dimensions more commonly studied in attitude and sustainability research (e.g., positivity/negativity, concern[57,58]), and these associations have increased (tracking the rise of environmental concerns and post-materialist values broadly). Yet, comparisons across 120 languages showed that English today

still remains among the weakest *Nature-Importance* associations, especially lower than Niger-Congo and Austronesian languages spoken in the Global South.

Finally, such global variation in *Nature-Importance* representations matters. We found that language-based representations are correlated significantly and positively with indicators of national environmental health, especially protection of biodiverse land and water ecosystems, 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[48,69], we show that collective representations of *Nature-Importance* add meaningful explanatory value beyond those past variables, at least on-par with economic inequality. Thus, even though the raw correlations and effect sizes may be small-to-moderate, even small relationships aggregated to a global scale across billions of people can help answer the "wicked problem" of environmental degradation.

Despite its contributions, the present research faces empirical limitations to be addressed in future work. First, although we validated non-English translations with native speakers 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. Pairing the current high-scale approach with collaborative work with local communities and Non-Governmental Organizations would support culture-specific understandings of nature representations[26,77].

Second, we focused on *Nature-Importance* because it provides a broad evaluative signal about the centrality and prioritization of nature in a culture. But understanding *why* nature is seen as important is a critical next step and can help integrate the current findings with long-standing values-based perspectives[5]. For instance, perhaps nature is associated with importance via

biospheric values (nature is important for its own intrinsic value[6]), via nature connectedness (nature is important for its relevance to oneself), or even for cultural reasons (nature is important for preserving cultural diversity[27,28]).

Finally, we relied on data that, to our knowledge, are the largest raw cross-cultural texts available (from Wikipedia and parliamentary speeches from ParlaMint), but these still have limited temporal scope. Given that we saw little evidence of change in linguistic representations over the 8-year span for parliamentary text, and even less change over this period in the outcome data (see *SI Appendix*), future efforts will need much longer timespans to examine cross-lagged or causal relationships. Such relationships are bidirectional: associating nature with importance may spur a country to introduce policies that protect nature, while the protection of nature may, itself, increase associations of nature-importance. In addition to the longer timespans of data, future research can also employ diagnostics to establish the strength of causal inference from observational data[78], including looking at the roles of omitted variables or counterfactual results.

In conclusion, using an unprecedented scale of language data, the current work shows that nature is associated with importance across global languages, with gradual increases over the last two centuries. However, the strength of *Nature-Importance* associations remains lowest in Western languages, correlating with higher levels of environmental degradation, even after controlling for socio-economic factors. Given the prominence of English within international environmental policy frameworks[1], the relatively low *Nature-Importance* associations in English and languages from other WEIRD nations is perhaps alarming. Addressing global environmental challenges will benefit from integrating perspectives and collective representations from those regions where the importance of nature is already culturally embedded.

Materials and Methods.

Data sources.

Contemporary English text. We first explored how nature is represented in everyday English language, using GloVe word embeddings[50] trained on Common Crawl text, which comprises 840-billion-word tokens, with a vocabulary of approximately 2.2 million unique words. Note that, across all analyses, we show robust conclusions not with only the *GloVe* embedding algorithm but with *word2vec, fastText,* and *ALC* algorithms.

We relied on static embeddings (i.e., one embedding per word) because these are (a) computationally efficient and pre-trained (thus reducing environmental and energy costs [79]), (b) flexible to low data settings (e.g., ParlaMint text), and (c) widely-used and validated in previous work on cultural representations[37,41]. However, we recognize that these approaches are limited because they do not capture nuanced contexts and polysemy of words. Although there are initial efforts at generalizing contextualized embeddings (*BERT*[80]) and/or generative and large language models (LLMs, e.g., *GPT*) to multilingual settings, there remain concerns about their validity and representativeness of other languages[81]. Continued development and validation of non-English LLMs is necessary.

Historical English text. Pre-trained historical word embeddings were obtained from *Histwords*[51], which provides *word2vec* embeddings trained on English Google Books data from 1800-1990. These historical data have robust internal validity (e.g., they capture semantic shifts in words like *gay* or *broadcast*) and external validity, aligning with real-world events (e.g., the women's movement altered gender stereotypes in the language[40,41]).

Non-English Wikipedia embeddings. *FastText* word embeddings were pre-trained on over 200 different languages with available text on Wikipedia[31,32]. 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 clear geographic match (see below on the approach of language-country matching).

Non-English parliamentary embeddings. *À La Carte* (ALC) embeddings were created on all machine-translated text from European parliamentary speeches obtained from the ParlaMint dataset[54]. In brief, ALC embeddings use two sources of data: the context window (+/- 5 words) surrounding a target word (e.g., *nature*) to identify how a target word is discussed in a small corpus (e.g., in the Parliamentary speeches of Austria); and a base pre-trained embedding model (e.g., from English *fastText*). For each set of context window words, a word's ALC embedding is the average of the pre-trained embedding for all context words. For example, if the word *nature* appears in the context "we are very concerned that *nature* is going to be eradicated" we would average across the pre-trained embeddings for [*we, are, very, concerned, that, is, going, to, be, eradicated*]. We do this for every context in which *nature* appears and take the average to get the raw ALC embedding. Finally, we multiply the raw ALC embedding of each country by a transformation matrix to remove generic directions and highlight the unique semantic meaning of each word, yielding the final ALC embeddings that capture the country-specific representations.

Environmental health indicators. The UN Sustainable Development Goals (SDGs) are used for international 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. We therefore supplemented the SDGs with a dataset of >100 remotely-sensed geospatial indicators including tree cover, forest biomass, and terrain ruggedness (*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 codebook provided on OSF. Notably, sustainable development encompasses not only the environment, but also political, social, and economic outcomes as well. In exploratory analyses (*SI Appendix*) we also consider all other available UN SDG indicators, encompassing 83 indicators.

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[37]). We also validated these automated translations with professional translations from native speakers on *BLEND.com*. Inconsistencies were flagged and then checked using back-translations in Google translate and consulting native speakers (see *SI* for further details on accuracy statistics). Results showed that, on average, the identical word was used by both automated translations and native speakers for 59% of languages. But these scores were much higher for our central concepts of *Nature* (words were

identical for 72-85% of languages), with the lower averages coming from near-synonym translations for valenced concepts (e.g., words like "*gross*" and "*horrendous*" had identical translations for only 15-17% of languages). Still, given the high accuracy for our central concepts we 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 ensures comparability across all languages (including those without native-speaker translations).

Next, using these lists, we computed the mean average cosine (MAC) similarity[40] 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.

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, and do not provide information on the number of speakers. Such data is essential because we are interested in the collective representation[16] 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 and are constructed by elites.

As such, to match languages with countries, we introduce a 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. We underscore that this weighted Wikipedia-text estimation approach was nevertheless moderately and positively correlated with data from the European parliamentary speeches that had direct geo-tagged data (i.e., we knew exactly which countries the speakers were coming from). Ultimately, 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, TL, and HS collected and prepared data for analysis; TESC performed analyses and drafted the initial manuscript; all authors revised and approved the final manuscript.

Acknowledgements
We are grateful to Klaus Weber and William J. Brady for their comments on earlier drafts of this manuscript. Funding for this project was provided for TESC from the Roberta Buffett Institute at Northwestern University; 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.

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* **36**, 193–209 (2015).
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. Steg, L. Psychology of Climate Change. *Annu. Rev. Psychol.* **74**, 391–421 (2023).
6. de Groot, J. I. M. & Steg, L. Value Orientations to Explain Beliefs Related to Environmental Significant Behavior. *Environ. Behav.* **40**, 330–354 (2008).
7. Dietz, Thomas. *Decisions for Sustainability : Facts and Values*. (Cambridge University Press, Cambridge, 2023).
8. Richardson, M. *et al.* Macro-level determinants of nature connectedness: An exploratory analysis of 61 countries. *Ambio* **55**, 80–100 (2025).
9. Lou, X., Mei, D. & Li, L. M. W. Development and validation of a scale measuring intrinsic, relational, and instrumental environmental values. *J. Environ. Psychol.* **106**, 102744 (2025).
10. Mayer, F. S. & Frantz, C. M. P. The Connectedness to Nature Scale: A measure of individuals' feeling in community with nature. *J. Environ. Psychol.* **24**, 503–515 (2004).
11. Klain, S. C., Olmsted, P., Chan, K. M. A. & Satterfield, T. Relational values resonate broadly and differently than intrinsic or instrumental values, or the New Ecological Paradigm. *PLoS One* **12**, e0183962 (2017).
12. Klöckner, C. A. A comprehensive model of the psychology of environmental behaviour—A meta-analysis. *Global Environmental Change* **23**, 1028–1038 (2013).
13. 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).
14. Kormos, C. & Gifford, R. The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *J. Environ. Psychol.* **40**, 359–371 (2014).
15. Durkheim, E. *Sociology and Philosophy*. (The Free Press, New York, 1974).
16. Moscovici, S. Notes towards a description of Social Representations. *Eur. J. Soc. Psychol.* **18**, 211–250 (1988).
17. Schultz, P. W. Environmental Attitudes and Behaviors Across Cultures. *Online Readings in Psychology and Culture* **8**, 4 (2002).
18. Ihemezie, E. J., Nawrath, M., Strauß, L., Stringer, L. C. & Dallimer, M. The influence of human values on attitudes and behaviours towards forest conservation. *J. Environ. Manage.* **292**, 112857 (2021).
19. Freymeyer, R. H. & Johnson, B. E. A cross-cultural investigation of factors influencing environmental actions. *Sociological Spectrum* **30**, 184–195 (2010).
20. Mede, N. G. *et al.* Perceptions of Science, Science Communication, and Climate Change Attitudes in 68 Countries: The TISP Dataset. *Sci. Data* **12**, (2025).
21. 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).
22. Kleespies, M. W., Hahn-Klimroth, M. & Dierkes, P. W. Relational values of nature—a global empirical study of environmental students in 37 countries. *Ecology and Society* **30**, 3 (2025).

23. Suchyta, M., Dietz, T. & Frank, K. A. Cross-National Social and Environmental Influences on Life Satisfaction. *Social Forces* **103**, 66–86 (2024).
24. Berdejo-Espinola, V. & Amano, T. Assessing diverse values of nature requires multilingual evidence. *Nature Reviews Biodiversity 2025 1:1* **1**, 5–6 (2025).
25. Henrich, J., Heine, S. J. & Norenzayan, A. The weirdest people in the world? *Behavioral and Brain Sciences* **33**, 61–83 (2010).
26. Droz, L., Brugnach, M. & Pascual, U. Multilingualism for pluralising knowledge and decision making about people and nature relationships. *People and Nature* **5**, 874–884 (2023).
27. Maffi, L. Linguistic, cultural and biological diversity. *Annu. Rev. Anthropol.* **34**, 599–617 (2005).
28. David Harrison, K. Environmental Linguistics. *Annu. Rev. Linguist.* **9**, 113–134 (2023).
29. Inglis, D. & Pascual, U. On the links between nature's values and language. *People and Nature* **5**, 326–342 (2023).
30. Gong, Y., Wang, S., Li, Y. & Sun, Y. Discrepant implicit and explicit attitudes toward climate change: implications for climate change communications. *Sustainability Science 2023 18:3* **18**, 1367–1377 (2023).
31. Grave, E., Bojanowski, P., Gupta, P., Joulin, A. & Mikolov, T. Learning Word Vectors for 157 Languages. in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (2018).
32. Bojanowski, P., Grave, E., Joulin, A. & Mikolov, T. Enriching Word Vectors with Subword Information. *Trans. Assoc. Comput. Linguist.* **5**, 135–146 (2017).
33. 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.) 494–508 (Guilford Publications Inc., New York, 2022).
34. Jackson, J. C. *et al.* From Text to Thought: How Analyzing Language Can Advance Psychological Science. *Perspectives on Psychological Science* **17**, 805–826 (2022).
35. Kozlowski, A. C., Taddy, M. & Evans, J. A. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. *Am. Sociol. Rev.* **84**, 905–949 (2019).
36. Caliskan, A., Bryson, J. J. & Narayanan, A. Semantics derived automatically from language corpora contain human-like biases. *Science (1979).* **356**, 183–186 (2016).
37. 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).
38. Bhatia, S. & Walasek, L. Predicting implicit attitudes with natural language data. *Proceedings of the National Academy of Sciences* **120**, (2023).
39. Lewis, M. & Lupyan, G. Gender stereotypes are reflected in the distributional structure of 25 languages. *Nat. Hum. Behav.* **4**, 1021–1028 (2020).
40. 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).
41. 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).
42. Inglehart, R. The Silent Revolution in Europe: Intergenerational Change in Post-Industrial Societies. *American Political Science Review* **65**, 991–1017 (1971).

43. Dunlap, R. E. & York, R. The Globalization of Environmental Concern and The Limits of The Postmaterialist Values Explanation: Evidence from Four Multinational Surveys. *Sociol. Q.* **49**, 529–563 (2008).
44. Franzen, A. & Meyer, R. Environmental Attitudes in Cross-National Perspective: A Multilevel Analysis of the ISSP 1993 and 2000. *Eur. Sociol. Rev.* **26**, 219–234 (2010).
45. Martinez-Alier, J. The environmentalism of the poor: Its origins and spread. in *A Companion to Global Environmental History: Second Edition* (eds. McNeill, J. R. & Mauldin, E. S.) 479–492 (John Wiley & Sons, Hoboken, New Jersey, 2025).
46. Fairbrother, M. Rich People, Poor People, and Environmental Concern: Evidence across Nations and Time. *Eur. Sociol. Rev.* **29**, 910–922 (2013).
47. Panayotou, T. Demystifying the environmental Kuznets curve: turning a black box into a policy tool. *Environ. Dev. Econ.* **2**, 465–484 (1997).
48. 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).
49. Busch, J. & Ferretti-Gallon, K. What drives deforestation and what stops it? A meta-analysis. *Rev. Environ. Econ. Policy* **11**, 3–23 (2017).
50. 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.
51. 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).
52. Oeberst, A. & Ridderbecks, T. How article category in Wikipedia determines the heterogeneity of its editors. *Scientific Reports 2024 14:1* **14**, 740- (2024).
53. Pfeil, U., Zaphiris, P. & Ang, C. S. Cultural Differences in Collaborative Authoring of Wikipedia. *Journal of Computer-Mediated Communication* **12**, 88–113 (2006).
54. Erjavec, T. *et al.* The ParlaMint corpora of parliamentary proceedings. *Lang. Resour. Eval.* **57**, 415–448 (2023).
55. Eagly, A. H. & Chaiken, S. Attitude Structure and Function. in *The handbook of social psychology* (eds. Gilbert, D., Fiske, S. T. & Lindzey, G.) 269–322 (McGraw-Hill, New York, 1998). doi:10.2307/2072868.
56. Dunlap, R. E. & Jones, R. E. Environmental Concern: Conceptual and Measurement Issues. in *Handbook of Environmental Sociology* (eds. Dunlap, R. & Michelson) 482–524 (Greenwood Press, 2002).
57. Cruz, S. M. & Manata, B. Measurement of Environmental Concern: A Review and Analysis. *Front. Psychol.* **11**, (2020).
58. 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).
59. 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).

60. Krosnick, J. A. Attitude Importance and Attitude Accessibility. *Pers. Soc. Psychol. Bull.* **15**, 297–308 (1989).
61. Bouman, T., Steg, L. & Kiers, H. A. L. Measuring values in environmental research: A test of an environmental Portrait Value Questionnaire. *Front. Psychol.* **9**, (2018).
62. Johnson, I. Wikipedia Article Topics for All Languages (based on article outlinks) . *Dataset figshare* Preprint at https://doi.org/10.6084/m9.figshare.12619766.v3 (2020).
63. Wikiprojects. Wikipedia:WikiProject Council/Directory - Wikipedia. https://en.wikipedia.org/wiki/Wikipedia:WikiProject_Council/Directory (2026).
64. 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).
65. Khodak, M. *et al.* A La Carte Embedding: Cheap but Effective Induction of Semantic Feature Vectors. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* 12–22 (2018).
66. Hansen, M. C. *et al.* High-resolution global maps of 21st-century forest cover change. *Science (1979).* **342**, 850–853 (2013).
67. 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.
68. Grant, D., Jorgenson, A., Longhofer, W. & Vasi, I. B. ("Bodi"). The Long Shadow of Partisan Hostility: How Affective Polarization Hinders Democracies' Ability to Mitigate Climate Change. *Am. Sociol. Rev.* **91**, (2025).
69. Inglehart, R. Public Support for Environmental Protection: Objective Problems and Subjective Values in 43 Societies. *PS Polit. Sci. Polit.* **28**, 57–72 (1995).
70. World Bank Poverty and Inequality Platform. Income inequality: Gini coefficient, 2024. https://ourworldindata.org/grapher/economic-inequality-gini-index?time=latest (2025).
71. UNDP Human Development Report. Gender Inequality Index, 2023. https://ourworldindata.org/grapher/gender-inequality-index-from-the-human-development-report (2025).
72. International Telecommunication Union & World Bank. Share of the population using the Internet, 2023. https://ourworldindata.org/grapher/share-of-individuals-using-the-internet (2025).
73. Wikipedia. List of Wikipedias. https://meta.wikimedia.org/wiki/List_of_Wikipedias (2025).
74. 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).
75. Dataset Record: ESA Land Cover Climate Change Initiative (Land_Cover_cci): Land Surface Seasonality Products. https://catalogue.ceda.ac.uk/uuid/7c114fc6e2884c1f9ca107e7a502fdbf/.
76. 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).
77. Droz, L. *et al.* Exploring the diversity of conceptualizations of nature in East and South-East Asia. *Humanities and Social Sciences Communications 2022 9:1* **9**, 186- (2022).
78. Frank, K. A., Lin, Q., Xu, R., Maroulis, S. & Mueller, A. Quantifying the robustness of causal inferences: Sensitivity analysis for pragmatic social science. *Soc. Sci. Res.* **110**, 102815 (2023).

79. Bender, E. M., Gebru, T., Mcmillan-Major, A. & Shmitchell, S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (2021). doi:10.1145/3442188.3445922.
80. 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 (2019).
81. Karthikeyan, K., Wang, Z., Mayhew, S. & Roth, D. Cross-lingual ability of multilingual BERT: An empirical study. in *Proceedings of the International Conference on Learning Representations (ICLR 2020)* (2020).