communications psychology

Article

https://doi.org/10.1038/s44271-024-00183-5

Pathogen stress heightens sensorimotor dimensions in the human collective semantic space

Check for updates

Ze Fu[1,8], Huimin Chen[2,8], Zhan Liu[3], Maosong Sun[3], Zhiyuan Liu[3] ✉ & Yanchao Bi[1,4,5,6,7] ✉

Infectious diseases have been major causes of death throughout human history and are assumed to broadly affect human psychology. However, whether and how conceptual processing, an internal world model central to various cognitive processes, adapts to such salient stress variables remains largely unknown. To address this, we conducted three studies examining the relationship between pathogen severity and semantic space, probed through the main neurocognitive semantic dimensions revealed by large-scale text analyses: one cross-cultural study (across 43 countries) and two historical studies (over the past 100 years). Across all three studies, we observed that increasing pathogen severity was associated with an enhancement of the sensory-motor dimension in the collective semantic space. These patterns remained robust after controlling for the effects of sociocultural variables, including economic wealth and societal norms of tightness. These results highlight the universal dynamic mechanisms of collective semantics, such that pathogen stress potentially drives sensorially oriented semantic processing.

Infectious diseases have been major causes of death in human history[1] and are considered critical stress variables in human genetic and cultural evolution[2,3]. Pathogen prevalence has been linked to various sociocultural patterns, including societal pressures for conformity (i.e., cultural tightness[4]), the endorsement of religious values[5], political conservatism[6], and the moralization of language[7,8]. Under the assumption that ecological events are ultimately processed by and acted through the cognitive processes of the human brain to produce psychological and sociobehavioral changes, we aim to systematically explore the effects of pathogens on human cognition across cultures and time. Addressing this question not only introduces perspectives of cultural evolution to studies of cognition (see the historical psychology framework[9–12]), but also has the potential to enhance understanding of cognitive mechanisms that underlie sociocultural patterns, guiding coping strategies to address the potential long-term cognitive effects of the recent COVID-19 pandemic.

We focus on the effects of pathogens on a central component underlying diverse human behaviors—the conceptual/semantic space. Semantic memory stores knowledge and beliefs about the world, which serve as the basis for broad human cognitive functions, including language, perception, reasoning, and decision-making[13–15]. Decades of cognitive neuroscientific research on semantics have yielded a consensus that it is derived from sensory and language experiences, with a compositional representation that respects the neural architecture of the human brain, entailing sensorimotor and affective dimensions across salient information domains, including time, space, number, cognition, and social (neurocognitive dimensions[16–18]). For instance, the meaning of "rose" is (at least partly) supported by brain regions processing the color, shape, and olfactory properties of roses[19–21]. This neurocognitive dimensional structure has consistently been observed in populations with varying cultures and languages (e.g., the United States, the United Kingdom, Italy, China, and Israel) and in individuals with drastic experiential differences (e.g., congenital blindness or deafness)[22]. While it offers a potentially universal neurobiological mechanism through which the human brain processes and represents semantic information, a dynamic component of how this information may be modulated by broad stress variables during cultural evolution remains to be understood[23].

Given that human semantic representation is derived from sensory and language experiences, semantic variations reflecting changes in physical and sociocultural environments are naturally accommodated—words like

[1]State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China. [2]School of Journalism and Communication, Tsinghua University, Beijing, 100084, China. [3]Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China. [4]School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, 100871, China. [5]IDG/McGovern Institute for Brain Research, Peking University, Beijing, 100871, China. [6]Institute for Artificial Intelligence, Peking University, Beijing, 100871, China. [7]Chinese Institute for Brain Research, Beijing, 102206, China. [8]These authors contributed equally: Ze Fu, Huimin Chen.
✉e-mail: liuzy@tsinghua.edu.cn; ybi@pku.edu.cn

"illness" or "curfew" become more salient during the peak of a pandemic[24,25]. Broader effects beyond these obvious semantic content changes are not well-predicted. Previous research has reported the short-term effects of disaster events on human emotion and social domains (e.g., terrorist attacks[26], hurricanes[27], and the COVID-19 outbreak[28]). The contemporary cultural evolutionary framework has also identified the long-term impacts of pathogens within human social and moral domains in history[29,30]. Thus, one possibility is that socioemotional dimensions in the semantic space are associated with pathogens in some manner[7,8]. Another non-mutually-exclusive possibility is that pathogens, being biological threats to bodily well-being, affect how humans conceptually interpret sensory signals, thereby affecting the sensorimotor dimensions, i.e., those semantic dimensions derived from sensory-motor experiences.

Motivated by the rationales of the cultural evolution and historical psychology perspectives, we systematically examine the dynamic mechanisms of the collective semantic space associated with pathogens across space (cross-cultural data) and time (historical data). Cross-cultural variations represent the results of cumulative adaptive responses to varying pathogen loads in different regions. Temporal associations in history allow for the exploration of the dynamic process, establishing potential directionality and causal mechanisms. The backbone dimensions were operationalized to be those identified in neurocognitive semantic research (see review in ref. 17) : *sensory-motor* (color, shape, sound, touch, smell, taste, and bodily actions), *time*, *space*, *number*, *mental-cognition*, *emotion*, and *social*. We consider the large-scale language data of a particular group as a proxy for the collective global semantic space and examine its structural properties in relation to pathogen severity. Language, as a product of human minds, reflects (at least partly) human semantic representations[22,31]. To measure the semantic structural properties of this collective space, we adopt an approach from the cognitive neural research discipline and analyze the activation strength of the backbone dimensions of semantic representation[32]. In the cognitive framework of word production, a word is selected for production if its semantic and lexical activation exceeds others to a threshold[33,34]. We extend this rationale to the group level, using the production intensity of words (i.e., frequency) related to a semantic dimension during a time period as a measure of the degree to which that particular dimension is active in the global semantic space in that particular region during that time. The recent digitization of extensive linguistic records (e.g., "Google Books" contains 500 billion words spanning more than 200 years) and the availability of various large-scale text methods allow for such analyses[25,35].

Combining large-scale text data and historical pathogen data from multiple sources, we conducted three studies: one cross-cultural study (Study 1) and two historical studies (Studies 2 and 3). Study 1 investigated cross-cultural associations between historical pathogens and the contemporary intensity of neurocognitive semantic dimensions in 43 countries by analyzing frequency of words denoting these dimensions (i.e., univariate analysis). Study 2 examined historical associations between pathogens and semantics across four countries (the US, the UK, Italy, and China) using Google Ngram word frequency data and pathogen records spanning the past 100 years. Study 3 extended the univariate analysis findings of Study 2 to examine whether the detected semantic dimension changes (measured by word frequency) represent a general principle for broader concepts' representation (measured by word-dimension relation patterns). Using decade-wise word embedding data from the US, we analyzed how different concepts relate to neurocognitive semantic dimensions through multivariate pattern analysis. Pathogen data and corpora were selected to optimize the analysis approaches in each study (see Fig. 1 for the overall analysis scheme).

Methods

Study 1

In this study, we examined cross-cultural associations between the historical pathogen index and variations in the semantic backbone dimensional intensity in contemporary corpora at the country level. Specifically, we tested whether the current intensity of a given dimension, measured by relative word frequency, was associated with the historical pathogen severity across 43 countries where both text data and historical pathogen data were available (see Table S1 for information on the selected country samples). The contemporary word frequency data were obtained from a cross-linguistic word frequency database[36], which compiled millions of texts from public websites (archived between 2011 and 2013). Previous cross-cultural studies have emphasized potential covariation patterns among our variables of interest (i.e., pathogens and semantics) and sociocultural properties[4,37,38]. We thus tested whether the effects of pathogen-semantic associations were robust after controlling for specific sociocultural properties (i.e., cultural tightness and economic wealth), and further explored the mediating relationships between pathogens, semantics, and cultural variables (e.g., cultural tightness). The results from 43 countries are reported in the main analysis and further validated by analyses using larger sample sizes from 53 countries and an independent word dictionary from the linguistic inquiry and word count (LIWC) category (see Supplementary Note 1, Fig. S1 and S2). This study, along with Studies 2 and 3, was not preregistered. These studies analyzed publicly available text data and historical pathogen records. No human subjects were directly involved in the research, and no personal or identifiable information was collected or processed. The use of aggregated public data at the language/country level did not require formal ethics committee approval.

Historical pathogen severity. We utilized the historical pathogen severity index compiled by Murray and Schaller[39]. For each of the 230 geopolitical regions, they computed a severity index by averaging the prevalence rates of nine diseases: leishmaniasis, schistosomiasis, trypanosomiasis, leprosy, malaria, typhus, filariasis, dengue, and tuberculosis, which were obtained from epidemiological maps during the early 20th century.

Semantic dimension identification. Our study employed a neurocognitive semantic dimension framework to represent the collective semantic space. For the selection of specific dimensions, we relied on a consensus derived from semantic research in cognitive neuropsychology, brain imaging studies, and developmental psychology[16–18,22,40,41].

We first identified dimensions that had explicit neural semantic correlates, including the senses (color, shape, sound, touch, smell, taste), motor-action, space, time, causality, cognition, social, and emotion (see review in ref. 17). We took into account core knowledge domains that underpin early conceptual spaces, including objecthood, agency, number, space, and social (see review in ref. 18). These two lines of dimensions were collapsed and grouped into the "neurocognitive backbone semantic dimensions", which were analyzed throughout the study: *sensory-motor* (including the basic senses and bodily motor functions), *space*, *time*, *number*, *mental-cognition* (including the causal understanding and other cognitive processes), *emotion*, and *social* dimensions. Agency/animacy was not considered as a separate dimension since its information content was largely encompassed by other dimensions such as *sensorimotor*, *social*, and *mental-cognition*. The dimensions of *attention* and *drive* from ref. 17 were not included due to the lack of clear empirical evidence regarding their neural semantic correlates and semantic dimensional word correspondences. The resulting seven semantic dimensions show low to medium correlations with each other, excluding the possibility of high collinearity among the dimensions.

Contemporary semantic dimensional intensity variations. The intensity of specific semantic dimensions within a country's collective semantic space is determined by the frequency of words associated with those dimensions in contemporary linguistic corpora.

Corpora. Word frequency data were obtained from the WorldLex database, an open-source collection that provides frequency information for multiple languages[36]. The database contains texts archived between 2011 and 2013 from various web sources. We analyzed word frequency data from blogs and news articles. To ensure the quality and comparability of the corpora across languages, we excluded languages with fewer than 10 million total words or incomplete linguistic resources from the main analysis. This resulted in a sample of 43 countries (see Table S1). Replicating the analysis using the original samples of 53 countries yielded similar results (see Fig. S1).

Dimensional dictionary selection. We utilized an anchor word list for the semantic backbone dimensions, obtained from the NorthEuralex (NEL) lexicostatistical database[42]. NEL compiles data from dictionaries and other linguistic records, covering a substantial core vocabulary shared across languages. The concepts directly related to each dimension were chosen from the NEL concept list. Three native Chinese speakers manually selected words that define or correspond to each dimension and reached a consensus on a final word list. The complete list of concepts is summarized in Table S2. NEL provided translated word forms for each language, and in cases where word forms were not available in NEL, we used the Google Translation API to translate them from English. The results were validated using another translation API (DeepL; www.deepl.com) and using concepts from an independent LIWC dictionary[43] with larger word samples (see Table S3 and Figure S2).

Frequency normalization. The intensity of each semantic backbone dimension in each language was computed by summing the frequencies of words associated with each dimension. To address potential differences in word frequencies across countries, we used a scalar method to calculate the relative frequency differences for each dimension (one vs. all: e.g., *sensory-motor* vs. (*sensory-motor* + *time* + *space* + *number* + *mental-cognition* + *emotion* + *social*)).

Sociocultural differences. We considered two variables that could confound the pathogen-semantic associations: cultural tightness and economic wealth. We examined whether the pathogen-semantic associations could be attributed to these sociocultural variables and, conversely, whether the pathogen-semantic associations could explain sociocultural variations. Cultural tightness was assessed using a composite index compiled from the 2000 wave of questionnaires from the

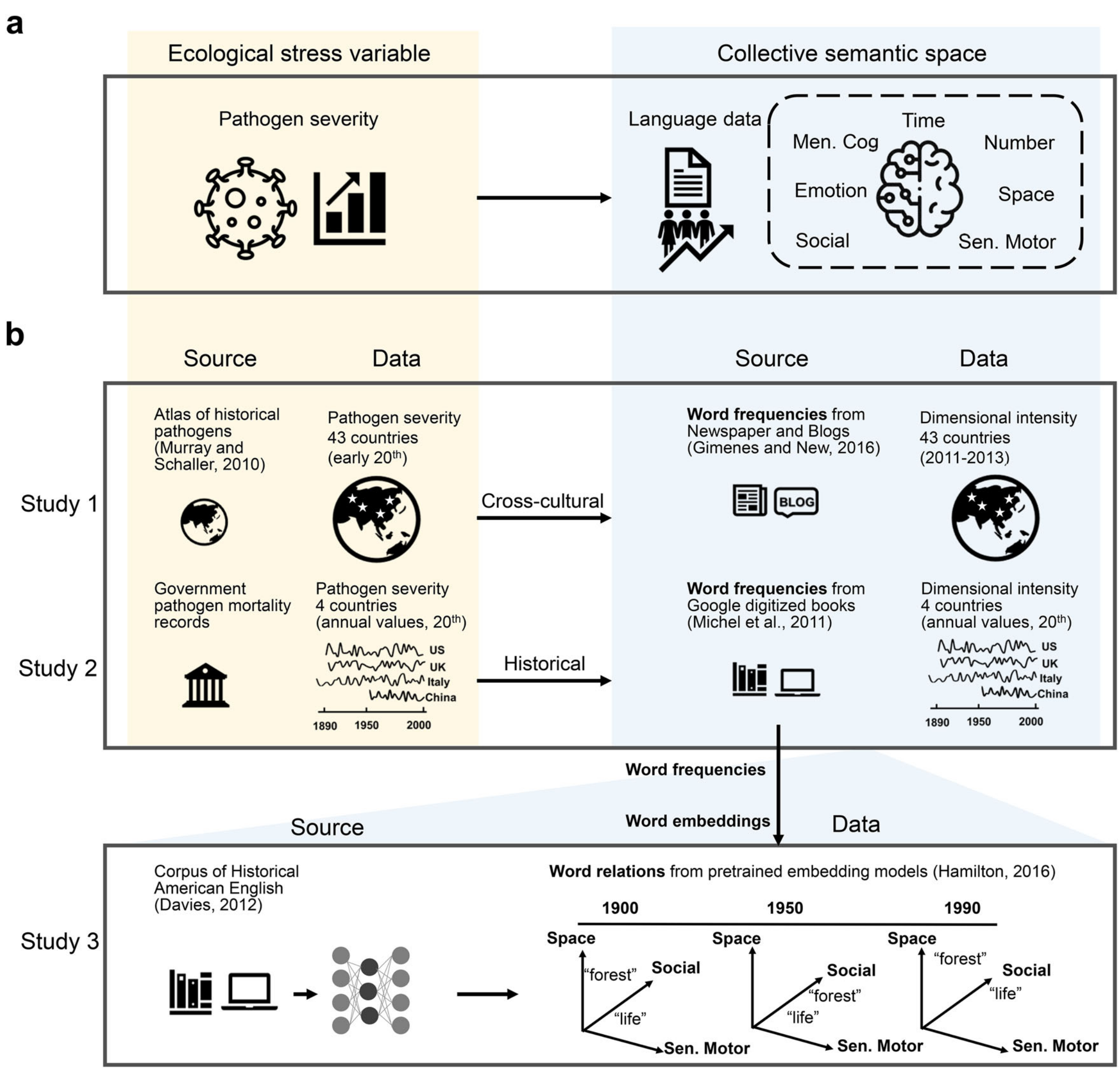

Fig. 1 | Overall schematic diagram. a The research framework. Our aim was to investigate the dynamic mechanisms of the collective semantic space associated with language in response to pathogens. To represent the collective semantic space, we adopted a neurocognitive semantic dimensional framework (e.g., Binder et al.[17]) that encompasses *sensory-motor* (sen. motor), *time*, *space*, *number*, *mental-cognition* (men. cog), *emotion*, and *social* dimensions. (**b**) The pipeline of data collection and analysis. We conducted two studies: a cross-cultural study (Study 1) and a historical study (Study 2) to examine the associations between pathogens and semantics using word frequency data (i.e., univariate analyses). Study 3 (historical study) was further conducted in the US to investigate the representation patterns of broader concepts across the neurocognitive semantic dimensions using word embedding data (i.e., multivariate word pattern analyses). Data sources for pathogens and language corpora were selected to optimize the analysis approaches for each study.

European Values Study and World Values Survey[44]. Economic wealth was determined by obtaining the average GDP per capita values from the Our World in Data database[45] for the 2000–2009 period.

Analytical procedures. We primarily investigated the direct associations between pathogens and the variations of semantic intensity for each dimension using Kendall correlation analyses. We used Kendall's tau (r_τ) to measure the associations between variables because it does not assume normal distributions of the data, making it suitable for our cross-cultural dataset and the historical data used in Studies 2 and 3. We explored whether sociocultural differences, specifically cultural tightness and GDP per capita, could explain the relationships between pathogens and semantics using regression models. We also incorporated corpus size as an additional regressor to address global disparities in word frequencies across languages. To address the geographic autocorrelations within our country-level data[7,46], we employed multi-level regression models. Specifically, we treated geographic regions (i.e., continents: Europe, Asia, and Africa) as a random factor in our analysis. Given that only one estimate was available for America (i.e., the US), it was excluded from the multi-level regression analyses. To control for inherent geographic differences, we further validated the results by including absolute latitude as a control variable[47]. Finally, we conducted mediation analyses to examine the potential mechanisms underlying the relationships among pathogens, semantic variations, and sociocultural psychological differences (cultural tightness).

Study 2

This study aimed to investigate the temporal dynamic associations between pathogens and semantics. Specifically, we examined whether changes in infectious disease levels led to alterations in the semantic dimension intensity throughout the last century. This analysis allows us to validate the relationships observed in Study 1 from a different perspective and to further assess the historical relationships between pathogens and changes in the intensities of semantic dimensions. We empirically tested this dynamic hypothesis in four countries, varying in cultural and linguistic properties, and with relatively comprehensive historical records (i.e., the Google Ngram[25]) from the past century: the US, the UK, Italy, and China. Similar to the cross-cultural data, we measured dimensional intensity by summing the word frequencies for specific dimensions. We examined the temporal associations and causality between pathogens and the intensity of semantic dimensions, while controlling for potential confounding factors, including changes in cultural tightness and economic wealth.

Historical pathogen severity. Yearly mortality rates from major types of pathogens were collected from multiple sources, including official government records, statistics tabulations from the Centers for Disease Control and Prevention (CDC), and the World Health Organization (WHO) mortality database (https://platform.who.int/mortality), for four countries: the US (1900–2000); the UK (1901–2000); Italy (1887–2000); China (1955–2000). These measures included the most prevalent infectious diseases specific to each country throughout the 20th century. Detailed collection procedures are shown in Supplementary Note 2.

Semantic dimensional intensity changes. The intensity of a particular semantic dimension in the collective semantic space at a specific point in time (a year) is estimated based on the frequency of words representing that dimension in the linguistic record.

Dimensional dictionary selection. Validated LIWC dictionaries[43] were used for querying historical semantic dimensional changes. The LIWC dictionaries were used in the main analysis of this study because they offered a greater number of word samples per dimension compared to NEL, enabling a more reliable assessment of stable signals across the different years. Note that we also used NEL word lists here for validation, and obtained comparable patterns. The word dictionaries used for Italy and China were translated versions of LIWC-Simplified Chinese (LIWC-2015) and Italian (LIWC-2007), respectively.

Corpora. We extracted word frequency data from Google Books n-gram archives[25] for years with available pathogen severity data. Data after 2000 were excluded from the main analyses due to potential changes in Google's data acquisition patterns when it started receiving books from individual publishers, in addition to library holdings. We calculated historical changes in semantic dimensional intensity by summing the yearly frequency counts for words categorized based on major semantic dimensions.

To enhance the reliability of signal estimation in historical texts, we adhered to the guidelines of historical linguistic analysis[48] and implemented the following preprocessing steps.

Word exclusion and filtering. To address the issue of non-words resulting from limitations in Google Books' Optical Character Recognition system, words not found in comprehensive lexical databases (e.g., WordNet for English, Multilingual WordNet—ITA for Italian, and OpenHowNet for Simplified Chinese) were excluded when matching the LIWC dictionaries. Additionally, words assigned to multiple semantic dimensions were also excluded.

Frequency normalization. To account for non-semantic changes, such as the approximately 20% decrease in the overall frequency of the word "the" over the past century[48], we examined the relative frequency changes of words. We divided the frequency of those words by the frequency of a set of generic semantically neutral functional words. Building on Scheffer et al.[49], who used the article 'a' as a representation of total volume, we expanded our list of functional words for each language: American English (a, an, the); British English (a, an, the); Italian (un, uno, una, il, lo, i, gli, la, le). For Simplified Chinese, which does not use articles in the same way as English or Italian, we selected the three particle words: 的, 地, and 得. These particles are common and relatively semantically neutral, and they serve various grammatical functions, making them suitable for our frequency normalization analysis.

Data smoothing and standardization. To improve the signal detection capability of association tests on time-series word frequency data, we standardized the original time-series data and employed a small rolling window (width of 3) to smooth the data.

Social cultural changes. Similar to Study 1, we considered two sociocultural factors as control variables: cultural tightness and economic wealth. Annual estimates of the cultural tightness index in these four countries were calculated by analyzing the relative word frequency ratio between tight and loose words. We adopted the word list used in Jackson et al.[50] for the US and the UK, which was subsequently translated and reviewed by native speakers to create the Chinese and Italian versions. The annual estimates of GDP per capita values for each country were directly obtained from the Our World in Data database for the period 1850–2000[45].

Analytical procedures. We first examined the associations between pathogens and intensity changes for each semantic dimension. For dimensions that showed consistent pathogen-associated signals, we conducted Granger causality tests to determine the direction of association patterns. Finally, a single-case design of Difference in Differences (DID) was employed to evaluate the potential causal relations between pathogens and semantic changes.

Time series association tests. We used Kendall correlations for the association tests, as these non-parametric tests are suitable for time series analysis. Before conducting the association tests, we applied regression models to account for the shared covariance between the time series, including the variables of cultural tightness and economic wealth, as well as the autocorrelation structure of the time series themselves. These models were applied to both the time series of pathogen severity and semantic changes to eliminate the influence of cultural tightness and GDP per capita. Furthermore, we utilized autoregressive integrated moving average (ARIMA) models to remove the autoregressive components from each time series[7,51]. We assessed the stationarity of all final time series using Dickey-Fuller tests. Subsequently, Kendall tau correlation tests were conducted to evaluate the associations between changes in semantic dimensional intensity and pathogen severity.

Granger causality tests. To investigate the direction of association patterns, we conducted Granger causality tests with one-year and five-year lags[7,50]. These tests use regression models to determine whether present semantic variations could be better predicted by past pathogen severity values than by past semantic variations themselves.

Single-case "Difference in Differences" tests. To further test the potential causal relations between pathogens and semantic changes, we adopted the DID schema[52] and searched for time periods during which one country experienced a pathogen outbreak (i.e., the treatment group) while the other three did not (i.e., the control group). Due to the limited number of time points, we utilized a Bayesian Standardized Difference Test[53], typically used in single-case studies, for statistical inference.

Study 3

Study 2 demonstrated a robust association between an increase in pathogens and an increased usage of words denoting *sensory-motor* properties, reflecting a (univariate) strengthening of *sensory-motor* dimensions in the global semantic space. In Study 3, we moved beyond word frequency and further examined whether the semantic representation of other words also leaned towards *sensory-motor* experiences in general. That is, we investigated whether when pathogen levels increased, people associate the meaning of words such as "forest" and "life" in a manner closer to *sensory-motor* experiences. As it is impossible to go back in time and rate the *sensory-motor* attribute strength of a concept, we utilized distributional word embedding models to acquire these semantic associations, which have been shown to correlate with humans' subjective ratings of the corresponding dimensional strengths[54,55]. By projecting the word space onto specific dimensional anchor words using historical word embedding data, we can acquire dimensional associations for any concept in the linguistic history[56,57]. We focused on analyzing historical changes within the US and employed a pre-trained embedding model trained using an independent corpus of historical American English (COHA[58,59]). We tested word embedding association patterns between broad concept lists (NEL basic concept words) and the seven semantic dimensional anchor words, examining whether pathogens were associated with a shift in semantic representation towards *sensory-motor* attributes in word meanings in general (i.e., a multivariate pattern analysis).

Changes of semantic dimensional representations (embedding relations)

Embedding model. We used a pre-trained embedding model from HistWords[59], which was trained on COHA, a genre-balanced corpus providing an independent source for detecting semantic shifts in the US over time[58]. The diachronic embedding model provided decade-wise word embedding representations for the top 100,000 most frequently occurring word tokens from the 1810s to the 2000s.

Target concept list. We examined a comprehensive list of 1016 conceptual words that encompass diverse semantic domains. These words were sourced from NEL concept lists[42]. The identification of semantic domains was based on the Concepticon frameworks[60], encompassing a range of topics such as 'Agricultural and Vegetation', 'Animals', and 'Social and Political Relations'. Querying the 1016 conceptual words in COHA yielded representations of a list of 947 unique English words.

Dimensional dictionary selection. We utilized the same dimensional anchor word lists used in Study 1. We obtained word representations for both target concepts and anchor dimensional words, each consisting of 300-dimensional vectors. Cases where words appeared in both target and anchor words were excluded.

Computation of dimensional strength. For each decade between 1900 and 2000, we calculated the mean cosine similarity between each target concept and its related dimensional anchor words to obtain dimensional semantic associative patterns across seven dimensions. For example, for the word "life", we calculated the *sensory-motor* specificity index by determining the ratio of its *sensory-motor* strength to its total strength across all seven dimensions (e.g., *sensory-motor* vs. (*sensory-motor* + *time* + *space* + *number* + *mental-cognition* + *emotion* + *social*)).

Analytical procedures. We investigated the association patterns using Kendall correlation analyses between the trends of changes in dimensional embedding associations, dimensional word frequencies, and historical pathogen severity. We further identified the target concept domains that exhibited the most consistent patterns with actual pathogen changes.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Results

Study 1

Cross-cultural association patterns between pathogens and changes in semantic dimensional intensity. We first examined the association patterns among the major semantic dimensions. The analysis revealed low-to-medium correlations ($r_\tau = -0.44$ to 0.45), confirming the relative independence of these dimensions as distinct constructs. The scatter plots in Fig. 2 illustrate the Kendall correlational patterns between pathogen severity and variations in major semantic dimensions across 43 countries. Out of the seven major semantic dimensions, pathogen severity was significantly and positively associated with contemporary *sensory-motor* word frequency intensity ($r_\tau = 0.38$, 95% CI [0.18, 0.58], Bonferroni-corrected $p = 0.002$, number of correlations: 7; see Table S4), while the other dimensions exhibited weak or no statistically significant associations ($r_\tau = -0.11$ to 0.16; 95% CI for the strongest correlation [−0.07, 0.39], $p = 0.129$). That is, countries with a greater severity of infectious diseases tend to use more *sensory-motor* words. We plotted maps depicting the geographic distributions of both historical pathogen severity and the intensity of *sensory-motor* words in Fig. 3a. The results remained robust despite variations in sample sizes across countries (Fig. S1) and were validated using an independent LIWC dictionary (Fig. S2). When examining individual sensory (color, shape, sound, taste, touch, and smell) and motor (i.e., bodily action) dimensions, we observed significant positive correlations in most cases, except for shape and smell (Table S5). Due to the overall similar pattern and limited anchor words for each individual dimension, we present the combined results of the *sensory-motor* dimension in the main text.

We tested whether sociocultural differences could explain the positive association between pathogens and *sensory-motor* semantic intensity. Kendall correlation analysis revealed moderate associations between *sensory-motor* and cultural tightness ($r_\tau = 0.26$, 95% CI [0.05, 0.46], $p = 0.031$), and no statistically significant associations between *sensory-motor* and GDP ($r_\tau = -0.18$, 95% CI [−0.45, 0.08], $p = 0.126$). Multilevel regression models were employed to account for the random effects of macro-geographic areas (i.e., continents: Europe, Asia, and Africa). As shown in Table 1, pathogen severity remained a significant predictor of country-level *sensory-motor* semantic intensity ($b = 0.47$, 95% CI [0.13,

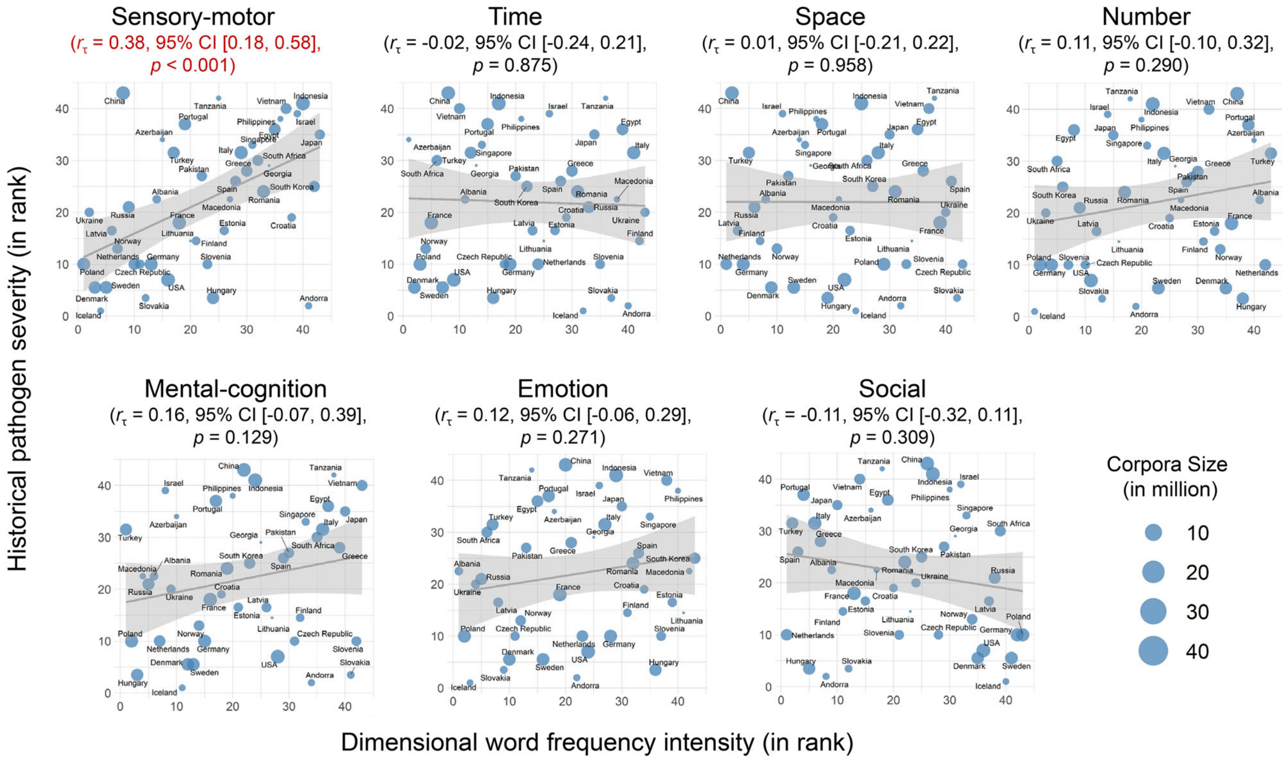

Fig. 2 | Variations of semantic dimensional intensities associated with pathogens in 43 countries. Scatterplots display the relationship between pathogens and word frequency intensity for each dimension (N = 43). Kendall tau correlation coefficients are shown above each plot. Countries that had higher historical pathogen severity also had a higher intensity of sensory-motor related word frequency in their languages (r_τ = 0.38, 95% CI [0.18, 0.58], p < 0.001). However, such pathogen-semantic associations were not statistically significant in other semantic dimensions. The size of the node corresponds to the size of the corpora (in millions of words). The shading of grey regions represents the 95% confidence intervals.

0.82], p = 0.013), even when controlling for covariates, including cultural tightness, GDP per capita, and corpora sizes. Cultural tightness and economic wealth did not exhibit significant independent predictive effects on *sensory-motor* semantic intensity (tightness: b = 0.17, 95% CI [−0.17, 0.51], p = 0.340; GDP per capita: b = 0.20, 95% CI [−0.14, 0.54], p = 0.266). To further address potential geographic variations within the macro-geographic areas, a more stringent test was conducted by incorporating absolute latitude as a regressor in the regression model. Pathogen severity remained a significant predictor of *sensory-motor* semantic intensity (b = 0.58, 95% CI [0.13, 1.02], p = 0.017).

We further conducted mediation analyses to examine the relationships among pathogens, *sensory-motor* semantic intensity, and cultural tightness, while controlling for economic wealth and corpora sizes as covariates (Fig. 3b). Moderate correlations were found between these three variables (r_τ = 0.25 to 0.50). Two separate mediation regression models were conducted, with *sensory-motor* semantic intensity and cultural tightness as mediating factors, respectively. The results revealed a statistical complete mediation effect of *sensory-motor* semantic dimensions on the relationship between pathogen severity and cultural tightness (b = 0.28, 95% CI [0.002, 0.61], p = 0.048, number of bootstraps: 5000). Note that this mediation effect does not survive after correction for multiple comparisons (two models were computed, resulting in a corrected p = 0.096). Even at the uncorrected threshold, statistical mediation effect of cultural tightness was not observed on the relationship between pathogen severity and *sensory-motor* semantic dimensions (b = 0.06, 95% CI [−0.04, 0.21], p = 0.250, number of bootstraps: 5000). That is, the observation that pathogen severity leads to increased cultural tightness might be explained by the semantic effects: pathogens led to increased *sensory-motor* semantic intensity, which further led to a tight culture, and not vice versa. This suggestive directional effect from pathogens to semantic *sensory-motor* dimensions would be further tested in the following studies.

Study 2

Historical association patterns between pathogens and semantic dimensional intensity. Figure 4a presents annual mortality data associated with major pathogens over the 20th century. In 1900, major infectious diseases accounted for approximately one-third of the overall mortality rates in the US, the UK, and Italy. These rates exhibited a downward trend until 1980, aligning with documented epidemiological transitions in Western countries, where acute infectious diseases shifted to chronic degenerative diseases[1,61]. China also exhibited similar trends, albeit over a shorter time period (1955–2000). Fluctuations in pathogen severity were evident throughout the last century, exemplified by notable events including the influenza pandemic in 1917–1918 in the US, the UK, and Italy, as well as the measles epidemic in 1959 and the meningitis outbreak in 1967 in China. Figure 4b displays changes in intensity (i.e., standardized frequency values) for each of the seven semantic dimensions in the four countries. The time series patterns align with earlier research[49]. For instance, in the three Western countries, words related to "cognitive processing" increased before the 1970s and decreased afterward, in line with a recent study reporting the rise and fall of rationality[49].

To examine whether the semantic space changes with pathogens in the four countries, Kendall tau correlation tests were performed on the time series of pathogen severity and each semantic dimension (Fig. 5a). All four countries had changes in words of certain semantic dimensions significantly associated with pathogens, with high convergence on *sensory-motor* dimensions (the US: r_τ = 0.59, 95% CI [0.46, 0.71], Bonferroni-corrected p < 0.001; the UK: r_τ = 0.35, 95% CI [0.21, 0.48], Bonferroni-corrected p < 0.001; Italy: r_τ = 0.41, 95% CI [0.29, 0.53], Bonferroni-corrected p < 0.001; China: r_τ = 0.29, 95% CI [0.11, 0.47], Bonferroni-corrected p = 0.02; number of corrections: 4). No other dimensions showed patterns consistent across the four countries (see

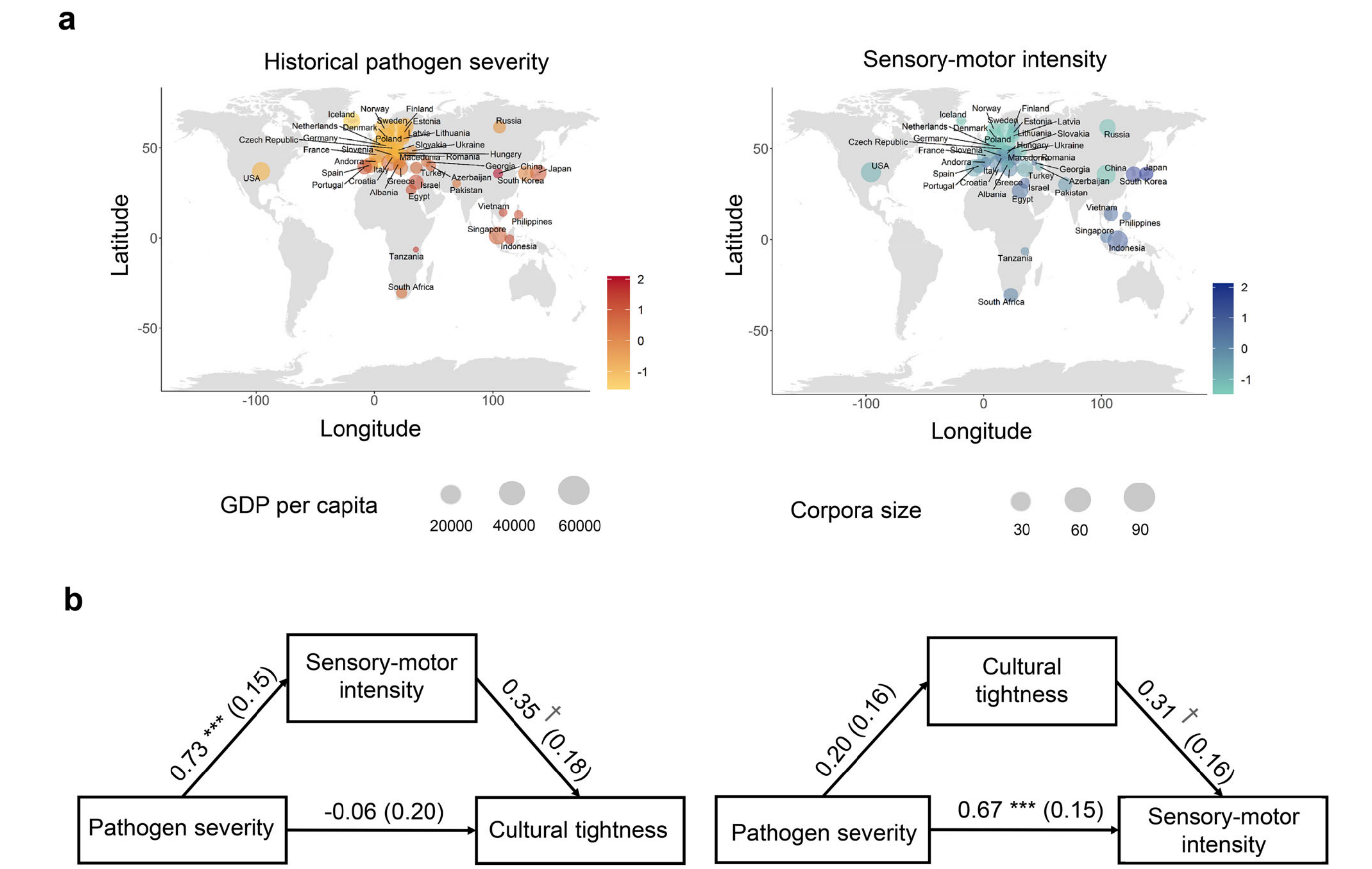

Fig. 3 | The relationships among historical pathogens, sensory-motor semantics, and cultural variables. a The distributions of historical pathogen severity (in red, left graph) and the intensity of sensory-motor words (in blue, right graph) across countries ($N = 43$). The size of the dots represents two control variables: GDP per capita for the historical pathogen severity map and corpora size for the sensory-motor intensity map. Note that the association pattern persisted after the latitude variable was controlled for (see text). **b** The mediation results ($N = 35$) for three variables—historical pathogen severity, sensory-motor intensity, and cultural tightness—are presented after controlling for GDP per capita and corpora sizes (***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; †, $p < 0.10$).

Table 1 | Multilevel regression models predicting contemporary sensory-motor semantic variations

	Model 1	Model 2	Model 3
(Intercept)	0.27 [−0.56, 1.09]	0.19 [−0.49, 0.88]	0.31 [−0.59, 1.21]
Historical pathogen severity	0.38* [0.06, 0.70]	0.47* [0.13, 0.82]	0.58* [0.13, 1.02]
Corpora size	−0.03 [−0.26, 0.20]	−0.07 [−0.32, 0.18]	−0.06 [−0.31, 0.18]
Cultural tightness	–	0.17 [−0.17, 0.51]	0.14 [−0.20, 0.49]
GDP per capita	–	0.20 [−0.14, 0.54]	0.15 [−0.19, 0.50]
Absolute latitude	–	–	0.27 [−0.24, 0.78]
Num.Obs.	35	35	35
R2 Marg.	0.14	0.24	0.14
R2 Cond.	0.55	0.50	0.57
AIC	90.50	97.02	99.03
BIC	98.27	107.91	111.48
ICC	0.47	0.34	0.51
RMSE	0.64	0.64	0.62

Note. Values show standardized β coefficients with 95% confidence intervals in brackets. ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; †, $p < 0.10$.

Table S6). For example, in the *social* dimension, positive associations were found in China, negative associations were found in the US and Italy, while no statistically significant associations were observed in the UK.

The raw data revealed significant non-stationarity. Following previous works[7,51], we preprocessed the time series data and rendered them stationary. We removed covariance between other pathogen-associated variables, such as cultural tightness and GDP per capita, using regression models, and autoregressive components using ARIMA models. In accordance with previous literature studying US culture[50], we found significant positive associations between pathogen changes and cultural tightness for the US, and extended these findings to three other countries (Fig. S3), including a culturally quite distant country (China). Importantly, we found that changes in words of the *sensory-motor* dimension were significantly associated with pathogens in all four countries after regressing out the effects shared by cultural tightness and economic wealth growth ($r_\tau = 0.22$ to 0.31, 95% CI for the lowest correlation [0.07, 0.36], Bonferroni-corrected $p = 0.005$; number of corrections: 4; Fig. 5b). Other semantic dimensions showed generally weaker and less consistent pathogen association effects (Fig. 5c). For *time* and *space*, no correlations reached statistical significance in any of the four countries. For *mental-cognition*, negative association patterns were observed only in the US ($r_\tau = -0.27$, 95% CI [−0.40, −0.13]). For *emotion*, pathogen association effects were observed in three countries, but with different directions (the US: $r_\tau = 0.29$, 95% CI [0.16, 0.42]; the UK: $r_\tau = 0.17$, 95% CI [0.02, 0.33]; Italy: $r_\tau = -0.18$, 95% CI [−0.31, −0.05]). For *social*, pathogens were positively associated only in the UK ($r_\tau = 0.20$, 95% CI [0.04, 0.35]).

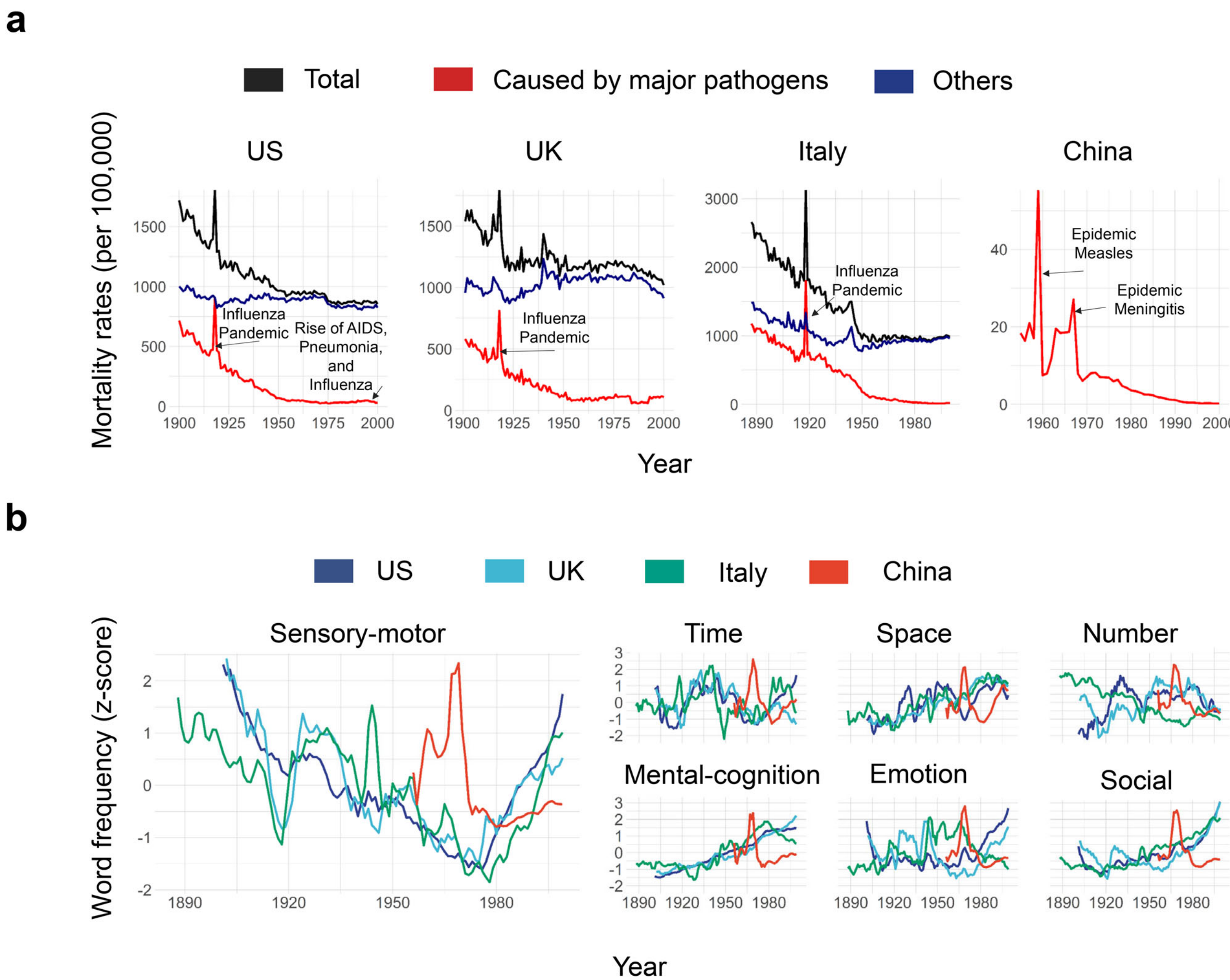

Fig. 4 | Quantification of pathogen severity and semantic variations across four countries in the past hundred years. **a** Mortality rates from major pathogens (per 100,000) in the US (1900–2000; $N = 101$), the UK (1901–2000; $N = 100$), Italy (1887–2000; $N = 114$), and China (1955–2000; $N = 46$). The mortality rates of major infectious diseases were obtained through the complete death registration systems in the three Western countries, whereas mortality rates in China were estimated using a sample registration method based on disease surveillance. **b** Changes in the intensity of semantic dimensions over time revealed in the four countries using large-scale linguistic corpora (Google Ngram) and semantic dictionaries (LIWC), with corresponding time periods and sample sizes as in (**a**). Time series of word frequency data were smoothed using a three-year rolling window and then standardized to *z*-scores.

Directional relationships between pathogen severity and the semantic sensory-motor intensity

Granger causality results. Having established a robust and consistent association pattern between pathogens and the intensity of *sensory-motor* semantic dimension over time, we explored the lead-lag patterns and conducted Granger causality tests with 1-year and 5-year lags. The Granger causality tests examined whether the current semantic variations could be better predicted using previous pathogen severity values than previous semantic variations themselves. As reported in Table 2, we identified significant pathogen-leading association patterns (at both 1-year lag and 5-year lags) in Italy and China ($Fs > 2.70$), indicating that pathogen severity predicts an increase in *sensory-motor* word frequencies in these two countries. For the US and the UK, the models in either direction or time lag were not statistically significant.

Single-case "Difference in Differences" results. Another approach to test statistically causal associations is the DID scheme[52]. This method involves forming a treatment group and a control group using existing observational data. In the available time series data (Fig. 4a), China had regional outbreaks of measles and meningitis in 1959 and 1967, while the other three countries (the US, the UK, Italy) did not have such an abrupt change. We can thus estimate the treatment effects by comparing the changes in *sensory-motor* word frequency between China and the other three countries during the two time periods (three years before vs. after the outbreaks in China) using the unprocessed time series data. For the single case study, we used the Bayesian Standardized Difference Test for dissociations[53]. The results show that the *sensory-motor* dimension (relative to other dimensions) exhibited significant changes before and after 1959 ($t(2) = 90.32$, $p < 0.001$, effect size zcc = 104.29, 95% CI [0.43, 0.48]) and 1967 ($t(2) = 14.86$, $p = 0.004$, zcc = 17.16, 95% CI [0.08, 0.14]). For additional control analyses, we performed the same tests on the year 1976, without abrupt pathogen outbreaks in any of the four countries and found no statistically significant dissociations between China and the other three countries ($t(2) = -2.54$, $p = 0.126$, 95% CI [−0.12, 0.03]). That is, the single-case DID analyses of specific pathogen outbreak events (1959, 1967) in China supported a significant directional effect of pathogens on the increase in the *sensory-motor* semantic dimension.

Study 3

Historical association patterns between pathogens and semantic representation for broader concepts. Here we tested whether the increased semantic dimension observed in the above univariate analyses

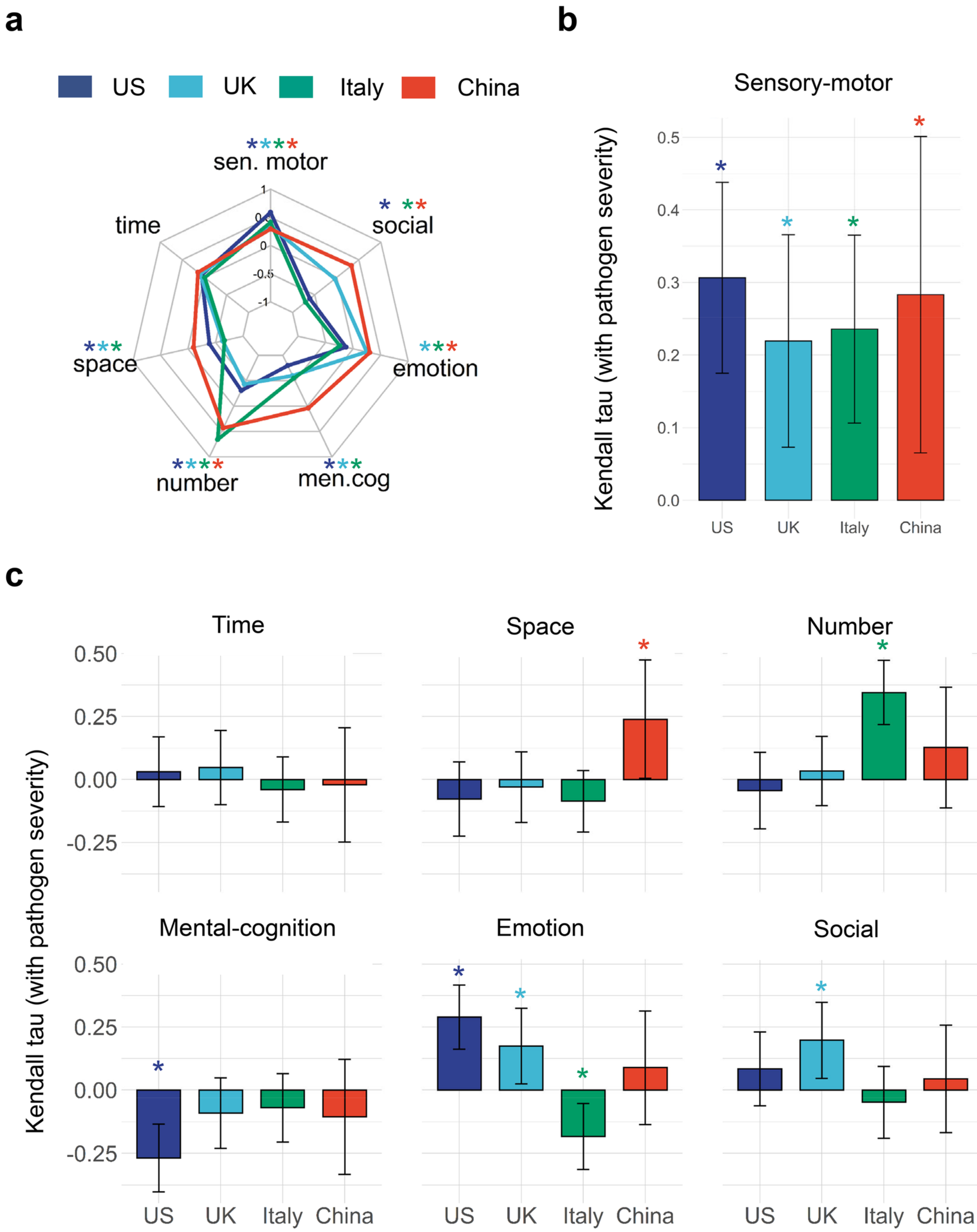

Fig. 5 | Historical changes in semantic dimension intensity associated with pathogen severity in four countries. The radar plot in **a** displays Kendall correlation coefficients between the time series of pathogen severity and changes in collective semantic dimensions, where values further from the center indicate stronger positive associations. The time series were smoothed using a 3-year rolling window. Data span the US (1900–2000; N = 101), the UK (1901–2000; N = 100), Italy (1887–2000; N = 114), and China (1955–2000; N = 46). Kendall correlations between pathogens and the sensory-motor dimension, as shown in (**b**), remained significant across the four countries even after controlling for cultural tightness, economic wealth, and autoregressive components. However, correlations were generally not significant in other dimensions, as indicated in (**c**) (error bars represent 95% confidence intervals; *, $p < 0.05$).

(increased related word use) represents a general principle for broader concepts in terms of their semantic dimensional representations. For NEL concepts, 947 had decade-wise word embedding data from the US, allowing us to test their representation along the seven neurocognitive semantic dimensions based on embedding similarity (i.e., a multivariate pattern analysis).

We first plotted the mean (decade-wise) semantic dimensional strength across these 947 concepts (Fig. 6a). The most notable change observed was the decreasing trend in words' *sensory-motor* dimensional strength (i.e., embedding similarity with *sensory-motor* attributes), which is further plotted in Fig. 6b. This pattern aligns with the historical changes in *sensory-motor* word frequencies ($r_\tau = 0.56$, 95% CI [0.12, 1.00]), as well as changes in pathogen mortality rates ($r_\tau = 0.78$, 95% CI [0.51, 1.00]; Fig. 6c). That is, higher pathogen mortality rates are associated with a broader shift across various semantic domains of words towards a stronger association with *sensory-motor* attribute words. As shown in Fig. 6d, pathogen changes

Table 2 | Granger causality test results for temporal association between pathogens and sensory-motor semantic dimensions

Granger causality model	Country	1-year-lag		5-year-lag	
		***F*-value**	η^2	***F*-value**	η^2
Pathogens → Sensory-motor					
	US	F (1, 95) = 2.98†	0.03 [0.00, 0.13]	F (5, 83) = 0.37	0.02 [0.01, 0.18]
	UK	F (1, 94) = 0.87	0.01 [0.00, 0.08]	F (5, 82) = 1.43	0.08 [0.03, 0.26]
	Italy	F (1, 108) = 7.76**	0.07 [0.01, 0.18]	F (5, 96) = 2.84*	0.13 [0.06, 0.30]
	China	F (1, 40) = 13.22***	0.25 [0.06, 0.48]	F (5, 28) = 2.70*	0.32 [0.16, 0.62]
Sensory-motor → Pathogens					
	US	F (1, 95) = 0.00	0.00 [0.00, 0.05]	F (5, 83) = 0.93	0.05 [0.02, 0.22]
	UK	F (1, 94) = 0.07	0.00 [0.00, 0.06]	F (5, 82) = 1.72	0.09 [0.04, 0.28]
	Italy	F (1, 108) = 1.82	0.02 [0.00, 0.09]	F (5, 96) = 0.26	0.01 [0.01, 0.15]
	China	F (1, 40) = 1.99	0.05 [0.00, 0.24]	F (5, 28) = 2.35†	0.30 [0.14, 0.60]

Note. Values in brackets represent 95% confidence intervals. ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; †, $p < 0.10$.

predominantly associate with changes in *sensory-motor* specificity of concepts related to the external world and its interactions with humans (e.g., animals, action, body, house, and agriculture).

Discussion

Semantic representation is assumed to be derived from sensorimotor interaction with the external world, yet the dynamic modulation by salient external variables is poorly understood. Here, we found highly convergent association patterns between pathogen severity and *sensory-motor* semantic processing in the global human conceptual space. Among the 43 countries in the cross-cultural analysis, historical pathogen severity positively associated with the *sensory-motor* semantic intensity, measured by word frequencies related to basic senses (color, shape, sound, touch, smell, taste) and bodily motor. In the historical analysis spanning the past century, increases in pathogen severity were consistently associated with increases in *sensory-motor*-related word frequencies in the four analyzed countries (the US, the UK, Italy, and China), with pathogen-leading patterns observed in Italy and China through Granger causality tests and single-case DID analyses. When pathogens increase, not only are more sensory-motor words produced, but also other words' representations become more sensory-motor-leaning—they become closer to *sensory-motor* words in the embedding space (US data). While associations with pathogens were also observed in other dimensions (*time*, *space*, *number*, *mental-cognition*, *emotion*, and *social*) in some analyses (e.g., *emotion* and *social* using the LIWC dictionary in Study 1; *emotion* in the US, the UK, and Italy in Study 2), the direction of effects, causal direction, and robustness were inconsistent across analyses. These findings suggest complex patterns that warrant further examination.

Can the observed relationship between pathogens and *sensory-motor* dimensions be explained by confounding variables? We mainly considered two sociocultural variables that have been found to be associated with pathogens: social tightness[4,38] and economic wealth[62]. However, neither of them accounted for relationships between pathogens and semantic variations. Ecological and societal change patterns are highly complex and interactive (e.g., many are related to resource shortages and/or broader ecological variables such as climate change[63,64]). It may not be possible to exhaust all potentially related variables. Particularly illuminating here is the consistency of results across different sets of data and approaches. Worth highlighting is that, in Study 2, the four countries varied in language, culture, and modernization stages (e.g., they landed distantly in the contemporary cultural maps in terms of traditional/secular values and survival/self-expression values[65]), naturally excluding attributions to any simple variable along these lines.

These findings highlight the dynamic mechanisms of semantics. As outlined in the Introduction, existing neurocognitive semantic theories naturally accommodate variations that mirror the outside world by having more words describing illnesses, symptoms, and corresponding actions. Our findings go beyond this by demonstrating that heightened pathogen severity is associated with broad and lasting cognitive changes in the semantic space. One possibility for such an effect is that heightened *sensory-motor* semantic dimensions are a general response mechanism to any kind of threat[66,67]. Alternatively, the conceptual system may respond differently to different types of threats, modulated by the type of human interactions. Pathogens affect body senses due to their direct effect on physical health (e.g., COVID-19 is associated with alteration of olfactory systems[68]), and/or the interpretation of sensory cues (e.g., evaluating potential pathogen-carrying threats), which differ from threats that do not directly alter sensory-motor experiences. One further consideration is whether the effects are related to the types of pathogens. For instance, one may reason that respiratory pathogens may particularly distort senses and related semantic structures such as smell and taste, whereas sexually transmitted diseases can have a substantially different psychological effect that is more socially oriented. We carried out an exploratory analysis using historical data (Study 2) from the US, which contained disease data by disease type. We extracted the raw time series of mortality rates caused by respiratory diseases (e.g., pneumonia and influenza) and sexually transmitted diseases (e.g., syphilis and AIDS) over the past century. Positive correlations between pathogen severity and *sensory-motor* dimensions were found for both types of diseases ($r_\tau = 0.49$ for respiratory diseases; $r_\tau = 0.29$ for sexually transmitted diseases). This preliminary result, although not to the same level of rigor as the main analyses, hints that the relationship between pathogens and *sensory-motor* semantic enhancement might be general across pathogen types. Nonetheless, detailed mechanisms, including whether different pathogens affect other semantic dimensions differently, or whether they affect different *sensory-motor* aspects (e.g., smell and taste vs. touch) differently, remain to be further examined with richer data and analyses. Importantly here, the two key findings—semantic organization varies with pathogens in a systematic manner (heightened *sensory-motor* dimensions), and that modulation pattern is similar across four countries with very different languages and cultures—support a dynamic universality hypothesis of semantics. Starting from a universal semantic brain structure and a universal mechanism of responding to environmental stress variables such as pathogens, the current diversity can be explained, as shown in Study 1.

What are the broader implications of our findings in relation to the post-COVID-19 age? These results highlight neurocognitive pathways through which pathogens modulate human behavior. Given that the semantic space is central to broad cognitive processes and behaviors[13,14], our findings may provide a potential cognitive mechanism underlying the previously reported associations between pathogens and sociocultural changes in cultural evolution research. That is, pathogens may affect the internal world model (i.e., semantic space), as we demonstrated here, which in turn leads to sociocultural behavioral changes. Consistent with this explanation, our historical study showed that the effects of pathogens on

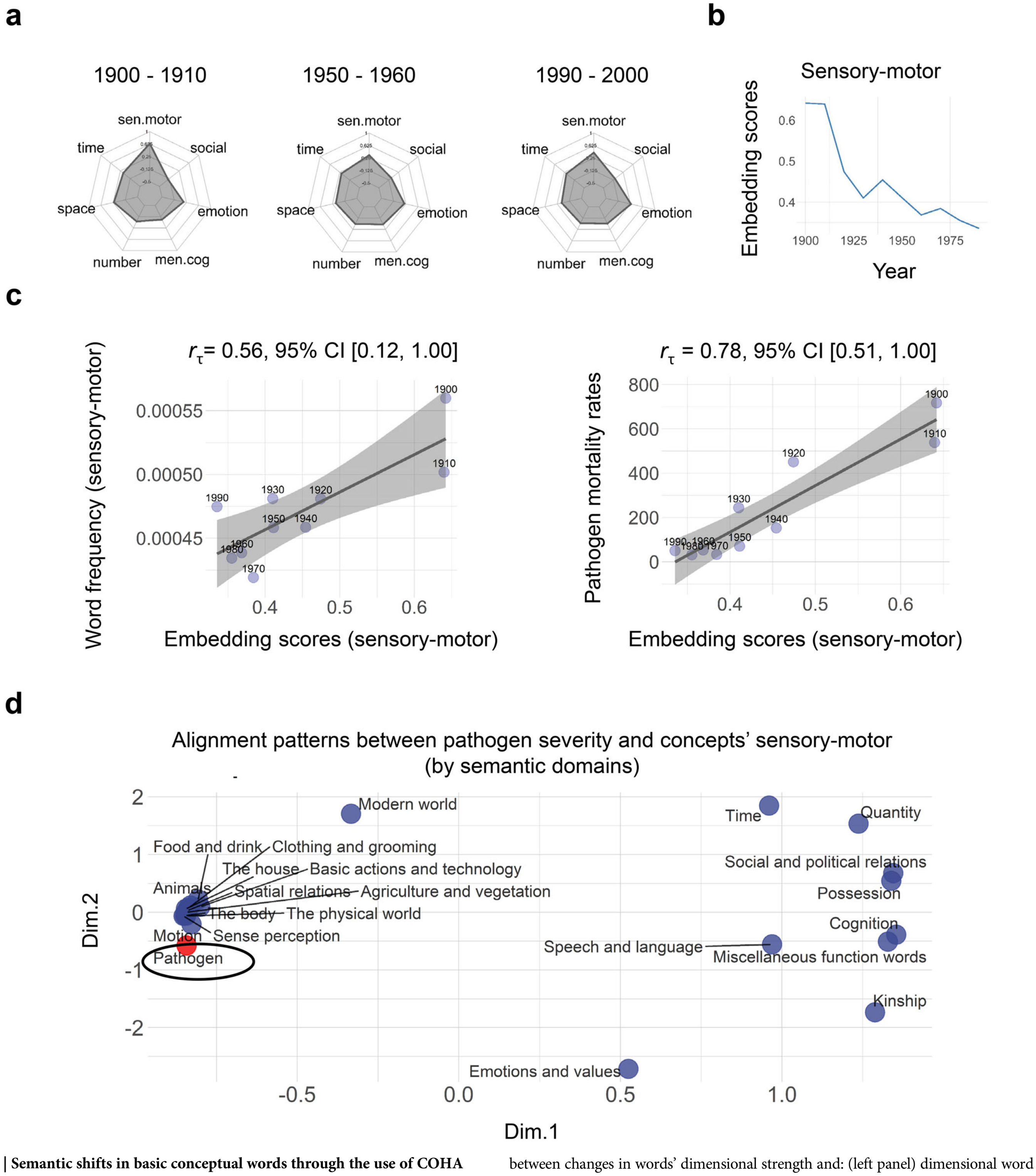

Fig. 6 | Semantic shifts in basic conceptual words through the use of COHA embedding models. **a** Radar plots showing the relative strength of seven semantic dimensions for 947 basic conceptual words in the US, comparing three time periods: 1900–1910, 1950–1960, and 1990–2000. The most noticeable changes observed were a decrease in sensory-motor strength, from 0.64 to 0.33. **b** Changes in average sensory-motor dimension strength for concepts (i.e., their embedding scores) from the 1900s to the 1990s ($N = 10$), with data points for each decade. **c** Correlations between changes in words' dimensional strength and: (left panel) dimensional word frequency, and (right panel) pathogen mortality rates in the US. **d** Alignment patterns in the changes observed in sensory-motor strength and pathogen severity across 21 semantic domains. The covariate patterns were visualized using isoMDS algorithms on the Kendall correlation matrix (converting into a distance matrix) among the time series of the pathogens and the sensory-motor strength of 21 semantic domains. The time series of pathogen severity is marked with black circles.

sensory-motor semantic changes were robust after controlling for the social tightness variable. The mediation analyses in the cross-cultural study showed a tendency that the heightened *sensory-motor* semantic space shift explained the effects of pathogens on social tightness, not vice versa. The increased usage of words denoting bodily sense awareness (corresponding to our *sensory-motor* words), beliefs, and personal pronouns has been taken as evidence of a shift away from the "rationality" towards "emotional expression", which may relate to salient sociocultural changes[49]. In addition to potentially mediating societal changes to become tighter and less rational, the mechanistic relationships between semantics and other real-world behaviors warrant systematic investigation[50,69–72]. This is particularly important for better understanding human behavioral changes after the COVID-19 pandemic and developing potential interventions. Future studies could benefit from incorporating ecological variables into semantic and social experimental settings (e.g., priming with threat cues[73,74]) to establish the causal mechanisms between pathogens, semantics, and behavioral changes.

Limitations

Our study is inspired and motivated by the theoretical development of multiple disciplines (cultural evolution, historical psychology, and neurocognitive research), and made possible only with recent methodological advances of large-scale data construction and analyses across disciplines, including language text data, medical data, and social data. By incorporating the semantic neurocognitive disciplines into the analyses of cross-cultural and temporal historical data, we offer an initial bridging pathway between ecological stress, cognitive semantic structure, and potential real-life behaviors. With this initial bridge, a few caveats and important future avenues are worth noting. Language data obtained from prints and social platforms are subject to interventions and biases, including censorship and selection bias. For instance, Google Books' sampling policy can lead to unrepresentative language data, influenced by socioeconomic, educational, and editorial biases[48]. Dictionary-based analyses in Studies 1 and 2 rely on the assumption that lexical semantics remain relatively static over time. Word frequency changes sometimes do not directly reflect semantic/psychological changes[10]. Corpus sizes and quality vary among countries, which could potentially confound important socioeconomic factors[75]. The consistency of multiple analyses, including historical and cross-cultural analyses from various language sources, as well as analyses of word frequency and embedding data, raises confidence in the obtained positive results. More generally, establishing causal relations based on historical data is challenging. Now that specific associative patterns are established based on historical data, and that the results of Granger causality analysis and single-case DID analysis are suggestive of a potential causal effect, hypotheses for causal mechanisms can be motivated and tested in various settings in future studies. A theoretical caveat is that written language represents only part of the human language reservoir, affected by writer/readership development, and may only partly reflect conceptual thoughts. Semantic knowledge derived from sensory experiences that are not coded in language may not be revealed in the current analyses[76]. Finally, future studies may also apply context-dependent NLP models (e.g., the Fill-Mask Association Test using BERT model[77]) to reveal dimensional effects that are more sensitive to top-down contextual variations, and also delve into the finer variations within these dimensions, such as emotional positivity within the *emotion* domain and specific social relationships (e.g., morality) within the *social* domain, to predict the specific long-term effects of pathogens.

Conclusion

To conclude, pathogens have shown broad effects on human collective semantic space throughout history. Their rise has been associated with an increase in *sensory-motor* oriented semantic processing, indicating that a more sensorial style of thinking evolved. These findings mark the beginning of understanding human semantic memory from the response-to-stress perspective and lead to new questions about the neural mechanisms and the consequences for a broad range of behaviors, as well as how such changes may accumulate and lead to the cultural evolution of human cognition.

Data availability

The data to reproduce the analyses in the manuscript and supplementary materials are available on the Open Science Framework (https://osf.io/byvts[78]).

Code availability

The analysis code is available on the Open Science Framework (https://osf.io/byvts[78]).

Received: 23 February 2024; Accepted: 19 December 2024;
Published online: 05 January 2025

References

1. Armstrong, G. L., Conn, L. A. & Pinner, R. W. Trends in infectious disease mortality in the United States during the 20th century. *J. Am. Med. Assoc.* **281**, 61–66 (1999).
2. Fumagalli, M. et al. Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution. *PLoS Genet.* **7**, e1002355 (2011).
3. Creanza, N., Kolodny, O. & Feldman, M. W. Cultural evolutionary theory: how culture evolves and why it matters. *Proc. Natl Acad. Sci. USA* **114**, 7782–7789 (2017).
4. Gelfand, M. J. et al. Differences between tight and loose cultures: a 33-nation study. *Science* **332**, 1100–1104 (2011).
5. Fincher, C. L. & Thornhill, R. Assortative sociality, limited dispersal, infectious disease and the genesis of the global pattern of religion diversity. *Proc. R. Soc. B* **275**, 2587–2594 (2008).
6. Tybur, J. M. et al. Parasite stress and pathogen avoidance relate to distinct dimensions of political ideology across 30 nations. *Proc. Natl Acad. Sci. USA* **113**, 12408–12413 (2016).
7. Atari, M. et al. Pathogens are linked to human moral systems across time and space. *Curr. Res. Ecol. Soc. Psychol.* **3**, 100060 (2022).
8. Malik, M., Hopp, F. R., Chen, Y. & Weber, R. Does regional variation in pathogen prevalence predict the moralization of language in COVID-19 news? *J. Lang. Soc. Psychol.* **40**, 653–676 (2021).
9. Muthukrishna, M., Henrich, J. & Slingerland, E. Psychology as a historical science. *Annu. Rev. Psychol.* **72**, 717–749 (2021).
10. Atari, M. & Henrich, J. Historical psychology. *Curr. Dir. Psychol. Sci.* **32**, 176–183 (2023).
11. Iliev, R. I. & Ojalehto, B. L. Bringing history back to culture: on the missing diachronic component in the research on culture and cognition. *Front. Psychol.* **6**, 716 (2015).
12. Henrich, J. et al. A cultural species and its cognitive phenotypes: implications for philosophy. *Rev. Philos. Psychol.* **14**, 349–386 (2023).
13. Murphy, G. *The Big Book of Concepts*. (MIT Press, 2004).
14. Fodor, J. A. *Concepts: Where Cognitive Science Went Wrong*. (Oxford University Press, 1998).
15. Laurence, S. & Margolis, E. in *Concepts: Core Readings* (eds Margolis, E. & Laurence, S.) 3–81 (MIT Press, 1999).
16. Martin, A. The representation of object concepts in the brain. *Annu. Rev. Psychol.* **58**, 25–45 (2007).
17. Binder, J. R. et al. Toward a brain-based componential semantic representation. *Cogn. Neuropsychol.* **33**, 130–174 (2016).
18. Spelke, E. S. & Kinzler, K. D. Core knowledge. *Dev. Sci.* **10**, 89–96 (2007).
19. He, C. et al. Selectivity for large nonmanipulable objects in scene-selective visual cortex does not require visual experience. *NeuroImage* **79**, 1–9 (2013).
20. Fernandino, L. et al. Concept representation reflects multimodal abstraction: a framework for embodied semantics. *Cereb. Cortex* **26**, 2018–2034 (2016).
21. Wang, X., Men, W., Gao, J., Caramazza, A. & Bi, Y. Two forms of knowledge representations in the human brain. *Neuron* **107**, 383–393. e385 (2020).
22. Bi, Y. Dual coding of knowledge in the human brain. *Trends Cogn. Sci.* **25**, 883–895 (2021).
23. Kemmerer, D. Grounded cognition entails linguistic relativity: a neglected implication of a major semantic theory. *Top. Cogn. Sci.* **15**, 615–647 (2023).

24. Lazer, D., Kennedy, R., King, G. & Vespignani, A. The parable of Google Flu: traps in big data analysis. *Science* **343**, 1203–1205 (2014).
25. Michel, J.-B. et al. Quantitative analysis of culture using millions of digitized books. *Science* **331**, 176–182 (2011).
26. Cohn, M. A., Mehl, M. R. & Pennebaker, J. W. Linguistic markers of psychological change surrounding September 11, 2001. *Psychol. Sci.* **15**, 687–693 (2004).
27. Neppalli, V. K., Caragea, C., Squicciarini, A., Tapia, A. & Stehle, S. Sentiment analysis during Hurricane Sandy in emergency response. *Int. J. Disaster Risk Reduct.* **21**, 213–222 (2017).
28. Wang, J. et al. Global evidence of expressed sentiment alterations during the COVID-19 pandemic. *Nat. Hum. Behav.* **6**, 349–358 (2022).
29. Thornhill, R. & Fincher, C. L. The parasite-stress theory of sociality, the behavioral immune system, and human social and cognitive uniqueness. *Evolut. Behav. Sci.* **8**, 257 (2014).
30. Murray, D. R. & Schaller, M. The behavioral immune system: Implications for social cognition, social interaction, and social influence. *Adv. Exp. Soc. Psychol.* **53**, 75–129 (2016).
31. Gelman, S. A. & Roberts, S. O. How language shapes the cultural inheritance of categories. *Proc. Natl Acad. Sci.* **114**, 7900–7907 (2017).
32. Fernandino, L., Tong, J.-Q., Conant, L. L., Humphries, C. J. & Binder, J. R. Decoding the information structure underlying the neural representation of concepts. *Proc. Natl Acad. Sci. USA* **119**, e2108091119 (2022).
33. Caramazza, A. How many levels of processing are there in lexical access? *Cogn. Neuropsychol.* **14**, 177–208 (1997).
34. Levelt, W. J. Spoken word production: a theory of lexical access. *Proc. Natl Acad. Sci. USA* **98**, 13464–13471 (2001).
35. Jackson, J. C. et al. From text to thought: how analyzing language can advance psychological science. *Perspect. Psychol. Sci.* **17**, 805–826 (2022).
36. Gimenes, M. & New, B. Worldlex: Twitter and blog word frequencies for 66 languages. *Behav. Res. Methods* **48**, 963–972 (2016).
37. Thompson, B., Roberts, S. G. & Lupyan, G. Cultural influences on word meanings revealed through large-scale semantic alignment. *Nat. Hum. Behav.* **4**, 1029–1038 (2020).
38. Gelfand, M. J. et al. The relationship between cultural tightness–looseness and COVID-19 cases and deaths: a global analysis. *Lancet Planet. Health* **5**, e135–e144 (2021).
39. Murray, D. R. & Schaller, M. Historical prevalence of infectious diseases within 230 geopolitical regions: a tool for investigating origins of culture. *J. Cross-Cult. Psychol.* **41**, 99–108 (2010).
40. Lambon Ralph, M. A., Jefferies, E., Patterson, K. & Rogers, T. T. The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* **18**, 42–55 (2017).
41. Dehaene-Lambertz, G. & Spelke, E. S. The infancy of the human brain. *Neuron* **88**, 93–109 (2015).
42. Dellert, J. et al. NorthEuraLex: a wide-coverage lexical database of Northern Eurasia. *Lang. Resour. Eval.* **54**, 273–301 (2020).
43. Pennebaker, J. W., Boyd, R. L., Jordan, K. & Blackburn, K. The Development and Psychometric Properties of LIWC2015 (The University of Texas at Austin, 2015).
44. Uz, I. The index of cultural tightness and looseness among 68 countries. *J. Cross-Cult. Psychol.* **46**, 319–335 (2015).
45. Bolt, J. & Van Zanden, J. L. Maddison-style estimates of the evolution of the world economy: a new 2023 update. *J. Econ. Surv.* https://doi.org/10.1111/joes.12618 (2020).
46. Bromham, L., Hua, X., Cardillo, M., Schneemann, H. & Greenhill, S. J. Parasites and politics: why cross-cultural studies must control for relatedness, proximity and covariation. *R. Soc. Open Sci.* **5**, 181100 (2018).
47. Van de Vliert, E. & Van Lange, P. A. Latitudinal psychology: an ecological perspective on creativity, aggression, happiness, and beyond. *Perspect. Psychol. Sci.* **14**, 860–884 (2019).
48. Younes, N. & Reips, U.-D. Guideline for improving the reliability of Google Ngram studies: evidence from religious terms. *PloS ONE* **14**, e0213554 (2019).
49. Scheffer, M., van de Leemput, I., Weinans, E. & Bollen, J. The rise and fall of rationality in language. *Proc. Natl Acad. Sci. USA* **118**, e2107848118 (2021).
50. Jackson, J. C., Gelfand, M., De, S. & Fox, A. The loosening of American culture over 200 years is associated with a creativity–order trade-off. *Nat. Hum. Behav.* **3**, 244–250 (2019).
51. Jackson, J. C., Gelfand, M. & Fox, A. Replicating the creativity-order tradeoff with more conservative time series models. *PsyArXiv* https://doi.org/10.31234/osf.io/7zs5t (2023).
52. Wing, C., Simon, K. & Bello-Gomez, R. A. Designing difference in difference studies: best practices for public health policy research. *Annu. Rev. Public Health* **39**, 453–469 (2018).
53. Crawford, J. R., Garthwaite, P. H. & Ryan, K. Comparing a single case to a control sample: testing for neuropsychological deficits and dissociations in the presence of covariates. *Cortex* **47**, 1166–1178 (2011).
54. Grand, G., Blank, I. A., Pereira, F. & Fedorenko, E. Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nat. Hum. Behav.* **6**, 975–987 (2022).
55. Lewis, M., Zettersten, M. & Lupyan, G. Distributional semantics as a source of visual knowledge. *Proc. Natl Acad. Sci. USA* **116**, 19237–19238 (2019).
56. Charlesworth, T. E., Caliskan, A. & Banaji, M. R. Historical representations of social groups across 200 years of word embeddings from Google Books. *Proc. Natl Acad. Sci. USA* **119**, e2121798119 (2022).
57. Garg, N., Schiebinger, L., Jurafsky, D. & Zou, J. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proc. Natl Acad. Sci. USA* **115**, E3635–E3644 (2018).
58. Davies, M. Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. *Corpora* **7**, 121–157 (2012).
59. Hamilton, W. L., Leskovec, J. & Jurafsky, D. *Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change*. (Association for Computational Linguistics, 2016).
60. List, J.-M., Cysouw, M. & Forkel, R. in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. 2393–2400.
61. Cohen, M. L. Changing patterns of infectious disease. *Nature* **406**, 762–767 (2000).
62. Bebber, D. P., Holmes, T., Smith, D. & Gurr, S. J. Economic and physical determinants of the global distributions of crop pests and pathogens. *N. Phytol.* **202**, 901–910 (2014).
63. Guernier, V., Hochberg, M. E. & Guégan, J.-F. Ecology drives the worldwide distribution of human diseases. *PLoS Biol.* **2**, e141 (2004).
64. Carlson, C. J. et al. Climate change increases cross-species viral transmission risk. *Nature* https://doi.org/10.1038/s41586-022-04788-w (2022).
65. Inglehart, R. *Christian Welzel Modernization, Cultural Change, and Democracy The Human Development Sequence*. (Cambridge University Press, Cambridge, 2005).
66. Greven, C. U. et al. Sensory processing sensitivity in the context of environmental sensitivity: a critical review and development of research agenda. *Neurosci. Biobehav. Rev.* **98**, 287–305 (2019).
67. Kelley, J. L., Chapuis, L., Davies, W. I. & Collin, S. P. Sensory system responses to human-induced environmental change. *Front. Ecol. Evolut.* **6**, 95 (2018).
68. Whitcroft, K. L. & Hummel, T. Olfactory dysfunction in COVID-19: diagnosis and management. *J. Am. Med. Assoc.* **323**, 2512–2514 (2020).

69. Choi, V. K., Shrestha, S., Pan, X. & Gelfand, M. J. When danger strikes: a linguistic tool for tracking America's collective response to threats. *Proc. Natl Acad. Sci. USA* **119**, e2113891119 (2022).
70. Ash, E., Stammbach, D. & Tobia, K. What is (and was) a person? Evidence on historical mind perceptions from natural language. *Cognition* **239**, 105501 (2023).
71. Seraj, S., Blackburn, K. G. & Pennebaker, J. W. Language left behind on social media exposes the emotional and cognitive costs of a romantic breakup. *Proc. Natl Acad. Sci. USA* **118**, e2017154118 (2021).
72. Varnum, M. E. & Grossmann, I. Pathogen prevalence is associated with cultural changes in gender equality. *Nat. Hum. Behav.* **1**, 1–4 (2016).
73. Jackson, J. C. et al. Tight cultures and vengeful gods: how culture shapes religious belief. *J. Exp. Psychol.* **150**, 2057 (2021).
74. Little, A. C., DeBruine, L. M. & Jones, B. C. Exposure to visual cues of pathogen contagion changes preferences for masculinity and symmetry in opposite-sex faces. *Proc. R. Soc. B* **278**, 2032–2039 (2011).
75. Blasi, D., Anastasopoulos, A. & Neubig, G. Systematic inequalities in language technology performance across the world's languages. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics* Vol. 1: Long Papers 5486–5505 (2022).
76. Majid, A. et al. Differential coding of perception in the world's languages. *Proc. Natl Acad. Sci. USA* **115**, 11369–11376 (2018).
77. Bao, H. The Fill-mask association test (FMAT): measuring propositions in natural language. *J. Personal. Soc. Psychol.* https://doi.org/10.1037/pspa0000396 (2024).
78. Fu, Z. et al. Pathogen stress heightens sensorimotor dimensions in the human collective semantic space. *OSF* https://doi.org/10.17605/OSF.IO/BYVTS (2024).

Acknowledgements
We thank Xiaosha Wang, Huichao Yang, Xi Yu, Haojie Wen, and Ziyi Xiong for helpful comments on earlier drafts of the manuscript. This work was supported by STI2030-Major Project 2021ZD0204100 (2021ZD0204104 awarded to Y.B.), Changjiang Scholar Professorship Award (T2016031 awarded to Y.B.), the National Natural Science Foundation of China (31925020, 82021004 awarded to Y.B.; 62106126 awarded to H.C.), Beijing Academy of Artificial Intelligence (BAAI) (to Z.Y.L.), the Institute Guo Qiang at Tsinghua University (to M.S.), and "the Fundamental Research Funds for the Central Universities" (to Y.B.). The funders had no role in the conceptualization, design, data collection, analysis, decision to publish, or preparation of the paper.

Author contributions
Y.B. designed and supervised research; Z.F. and H.C. performed research; Z.F. and Z.L. analyzed the data; H.C., ZY.L., and M.S. contributed valuable discussions; Y.B. and Z.F. wrote the paper.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44271-024-00183-5.

Correspondence and requests for materials should be addressed to Zhiyuan Liu or Yanchao Bi.

Peer review information *Communications psychology* thanks Frederic R. Hopp and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Yafeng Pan and Jennifer Bellingtier. A peer review file is available.