

1 **Supplementary Information**
2 **Pathogen stress shapes universal human collective conceptual changes in history**
3
4 This file includes:
5 Supplementary Texts;
6 Figure S1-S5;
7 Table S1-S7.
8

Supplementary Texts:

Validation analyses using finer-grained (monthly) linguistic corpus: Monthly word frequency data in the US during the last century were collected from large-scale US newspaper archives (<https://www.newspapers.com>). The semantic dimensional words selection was based on the NEL-related concept list (for neuro-cognitive dimensions) and the manually-selected cultural concept lists (for social-cultural dimensions). For neuro-cognitive dimensions, the word frequency was aggregated and scaled in the proportion of all six cognitive dimensions (one vs all), while the pathogen changes were estimated using the pathogen-related word frequency data with 'a' as the scalar word since the official mortality rates have no monthly data available. We applied autoregressive integrated moving average (ARIMA) models to test the relationships between pathogens and semantic dimensional word frequency changes, with adjustments made to the error terms that address the temporal dependencies (following Choi et al., PNAS, 2021). Briefly here, three main parameters (p , d , q) were specified for the ARIMA errors, with p being the number of lags that account for the autoregressive structure of the model, d the order of differencing needed to stabilize the variance in the time series, and q the moving average value that explains the model's lagged random errors. The model parameter identification utilized the Akaike information criterion (AIC).

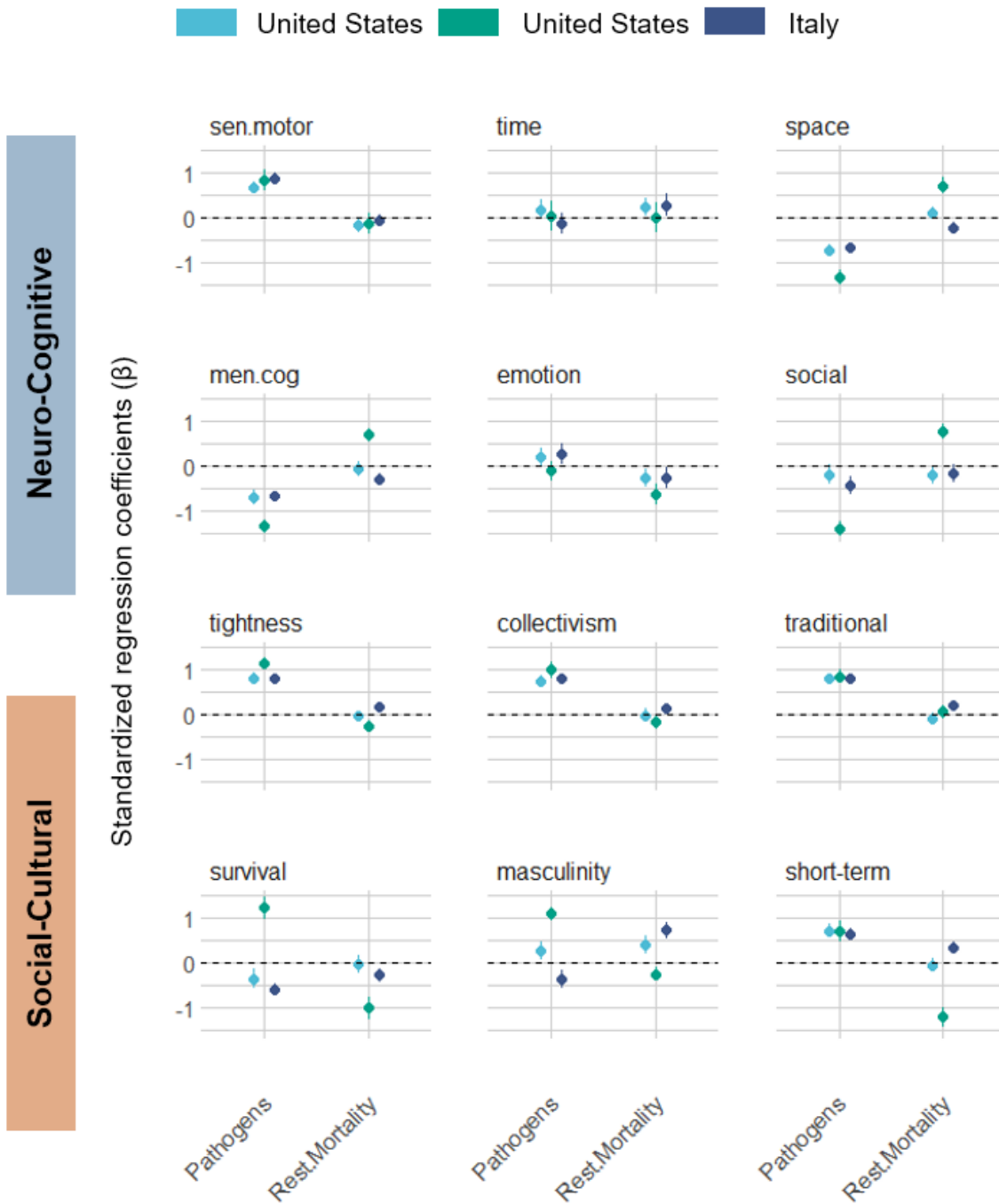
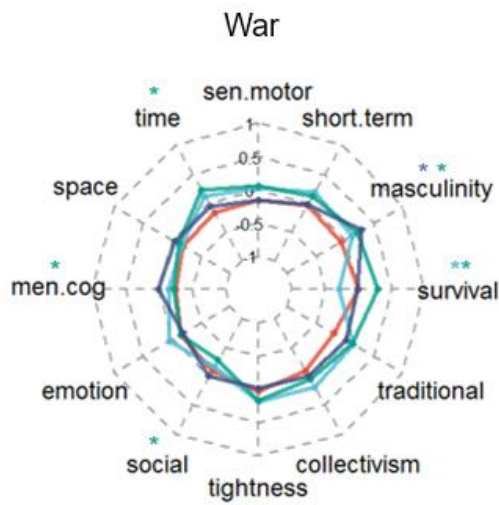
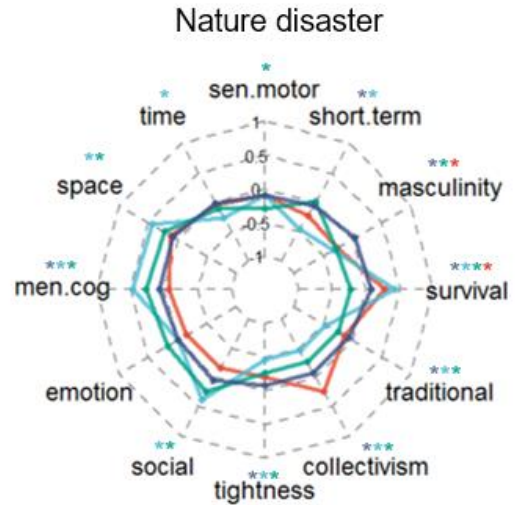


Figure S1. Pathogen-semantic associations when considering non-major-pathogen-related mortality rates as the controlling variable. Linear regressions were performed putting both major-pathogen-related mortality and the remaining mortality rates (estimated by total mortality – major-pathogen-related mortality) into models simultaneously. Standardized β values were extracted for each variable to show the effects on explaining the shifts of collective conceptual space. The error bars represent the 95% confidence intervals.

A

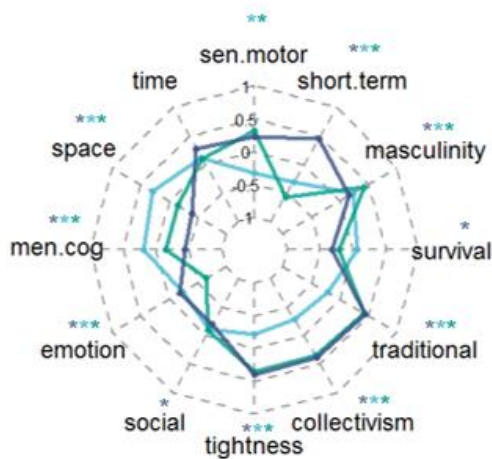


B



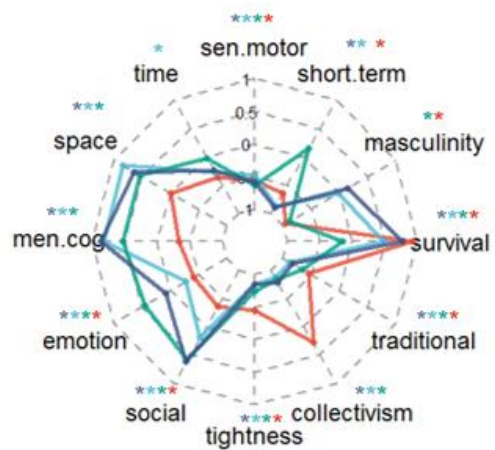
C

Non-major-pathogen-mortality-rates



D

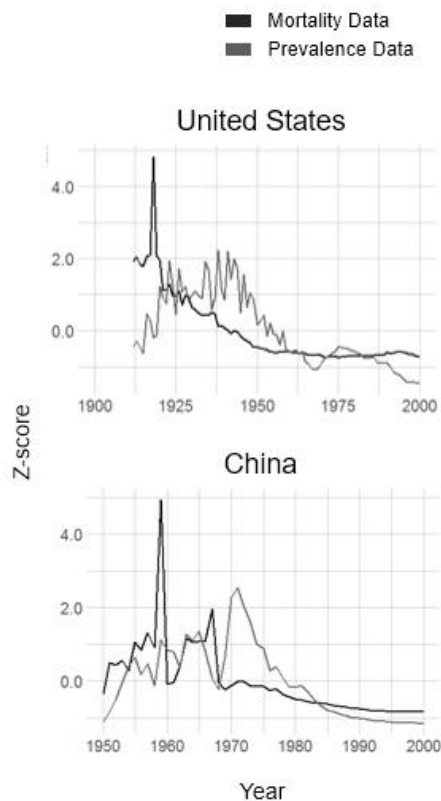
GDP per capita



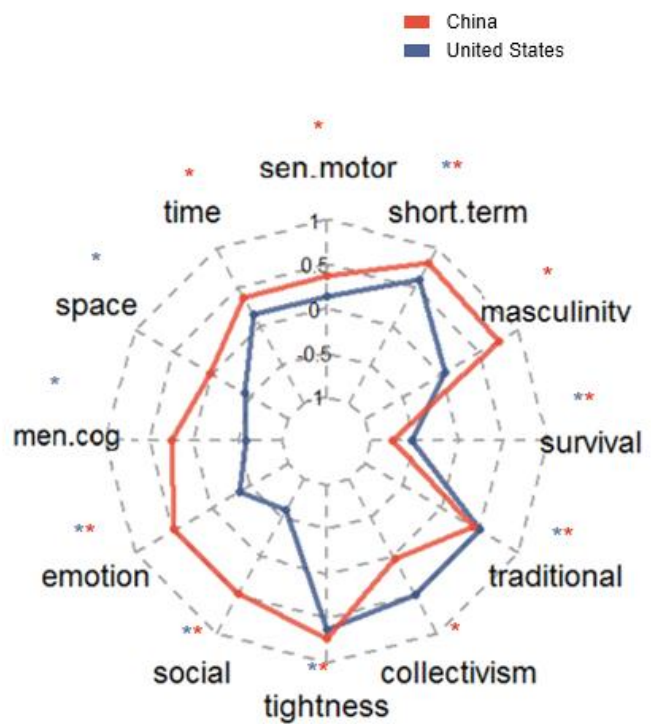
32

33 **Figure S2. Correlational plots of associative patterns between other ecological variables and**
 34 **collective conceptual spaces.** Correlations with (a) war, (b) nature disaster, (c) non-major-pathogen-
 35 mortality-rates, and (d) GDP per capita. Time series of word frequency data were smoothed using a
 36 rolling window (width = ± 3 years). The Kendall correlation coefficients between each ecological
 37 variable and each semantic dimension strength are depicted for each country (*, uncorrected $p <$
 38 0.05).

A



B



39

40 **Figure S3. Pathogen-semantic associations using pathogen prevalence as a complementary**
 41 **measure in the United States and China.** (a) Visualization of correlations between major pathogen-
 42 related mortality rates and prevalence rates. (b) Correlational patterns between pathogen prevalence
 43 rates and changes in collective conceptual spaces. Kendall correlations were used in estimating
 44 association patterns between time series, which is robust to nonlinear data. (*, $p < 0.05$). Time series
 45 of word frequency data were smoothed using a rolling window size = ± 3 before performing correlation
 46 tests.

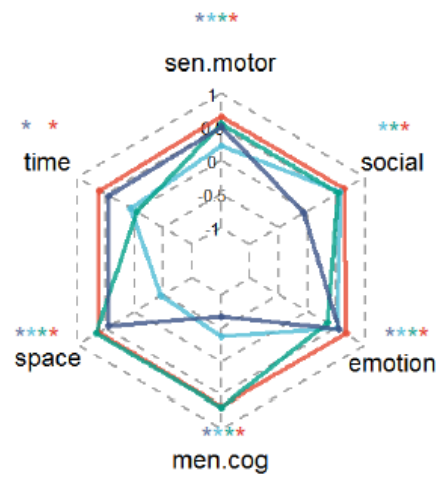
A

LIWC-Dictionary (vs. baseline)



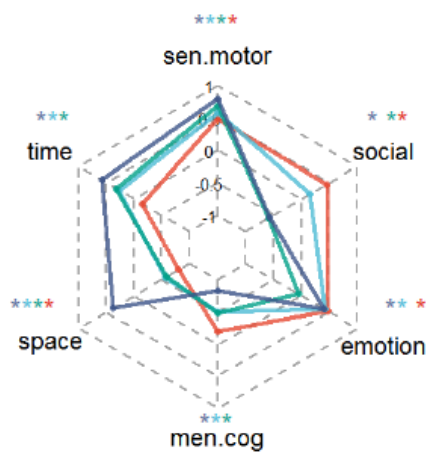
B

NEL-Dictionary (vs. baseline)



C

LIWC-Dictionary (vs. all six)



D

NEL-Dictionary (vs. all six)

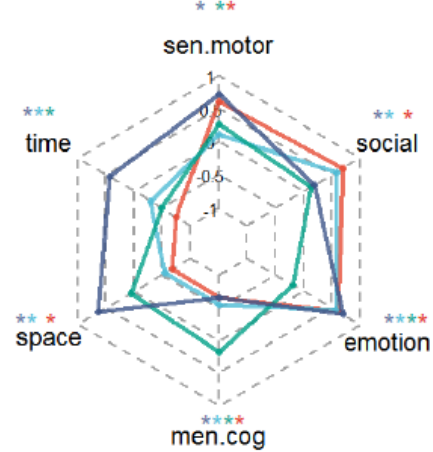
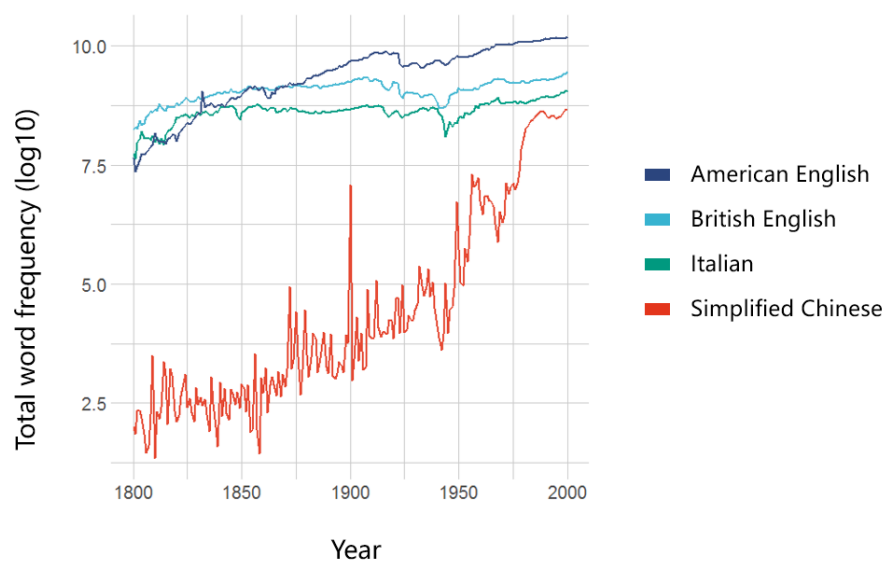


Figure S4. Correlational plots between pathogen mortality rates and six neurocognitive dimensions using different scaling methods and different word lists. (a) Correlations in the main results using the LIWC dictionary and the scaling method of contrasting baseline words (e.g., 'a', 'an', 'the' in English). (b) Correlation results using the Northeuralex (NEL) dictionary and the scaling method of contrast baseline words. (c) Correlation results using the LIWC dictionary and the scaling method of contrasting all six dimensions. (d) Correlation results using the NEL dictionary and the scaling method of contrasting all six dimensions. Time series of word frequency data were smoothed using a rolling window (width = ± 3 years). The Kendall correlation coefficients between the pathogen mortality rate and each semantic dimension strength are depicted for each country (*, uncorrected $p < 0.05$). Detailed statistical values (Kendall τ and p values) are shown in Table 3.



58
59 **Figure S5. Total word frequency (log10) of Google Book Unigrams**

60 **Table S1 Major infectious diseases included in our major pathogen severity estimations across four**
61 **countries.**

Country	Periods	Major Infectious Diseases
United States	1900-1932 (death registration states); 1933-1949 (all states)	pneumonia, influenza, tuberculosis, diarrhea and enteritis, diphtheria, pertussis (whooping cough), syphilis, measles, malaria, poliomyelitis, meningitis, tetanus, dysentery, typhoid and paratyphoid fever, scarlet fever, erysipelas;
	1950-2000	lower respiratory infections, upper respiratory infections, tuberculosis, diarrheal diseases, diphtheria, pertussis, syphilis, measles, malaria, poliomyelitis, meningitis, tetanus, AIDS;
United Kingdom	1901-1949 (England and Wales)	pneumonia, influenza, tuberculosis, diarrhea and enteritis, diphtheria, pertussis (whooping cough), syphilis, measles, malaria, poliomyelitis, meningitis, tetanus, dysentery, typhoid and paratyphoid fever, scarlet fever, erysipelas;
	1950-2000	lower respiratory infections, upper respiratory infections, tuberculosis, diarrheal diseases, diphtheria, pertussis, syphilis, measles, malaria, poliomyelitis, meningitis, tetanus, AIDS;
Italy	1887-1950	pneumonia, influenza, tuberculosis, diarrhea and enteritis, diphtheria, pertussis (whooping cough), syphilis, measles, malaria, poliomyelitis, meningitis, tetanus, dysentery, typhoid and paratyphoid fever, scarlet fever, erysipelas;
	1951-2000	lower respiratory infections, upper respiratory infections, tuberculosis, diarrheal diseases, diphtheria, pertussis, syphilis, measles, malaria, poliomyelitis, meningitis, tetanus, AIDS;
China	1960-2000	measles, epidemic meningitis, epidemic encephalitis B, poliomyelitis, dysentery, pertussis (whooping cough), diphtheria, malaria, typhoid and paratyphoid fever, typhus fever, scarlet fever, hemorrhagic fever, hepatitis, plague, cholera, rabies, leptospirosis, black fever, anthrax, and AIDS;

62

63 **Table S2 ICD-10 codes adopted in the WHO mortality database.**

Infectious diseases	Codes
Lower respiratory infections	J09-J22, P23, U04
Upper respiratory infections	J00-J06
Tuberculosis	A15-A19, B90
Diarrheal diseases	A00, A01, A03, A04, A06-A09
Diphtheria	A36
Pertussis (whooping cough)	A37
Syphilis	A50-A53
Measles	B05
Malaria	B50-B54, P37.3, P37.4
Poliomyelitis	A80, B91, G14
Meningitis	A39, G00, G03
Tetanus	A33-A35
AIDS	B20-B24

64

Table S3 ICD codes (1-6) adopted in the historical tabulations from the vital statistics agency in each country.

Infectious diseases	Versions	Codes
Pneumonia	ICD-R1-US	92, 93
	ICD-R1-UK	36, 37, 38, 39
	ICD-R2	91; 92 (A, B)
	ICD-R3	100; 101 (a, b)
	ICD-R4	107; 108, 109
	ICD-R5	107 (1, 2); 108 (1, 2); 109 (1, 2)
	ICD-R6	490-493
Influenza	ICD-R1-US	10
	ICD-R1-UK	12
	ICD-R2	10
	ICD-R3	11 (a1, a2, b1, b2)
	ICD-R4	11 (a1, a2, b1, b2)
	ICD-R5	33 (a1, a2, b1, b2)
	ICD-R6	480-483
Tuberculosis	ICD-R1-US	26-34
	ICD-R1-UK	46-53
	ICD-R2	28 (A, B); 29 (A, B); 30; 31 (A, B); 32; 33; 34 (A, B, C); 35
	ICD-R3	31; 32; 33; 34; 35; 36 (a, b, c, d, e); 37 (a, b)
	ICD-R4	23; 24; 25; 26; 27; 28; 29; 30; 31 (1, 2); 32 (a, b, c)
	ICD-R5	13 (a, b, c); 14; 15; 16; 17; 18; 19; 20; 21 (a, b); 22 (a, b, c)
	ICD-R6	001-019
Diarrhea and Enteritis	ICD-R1-US	105, 106
	ICD-R1-UK	20-22, 128-130
	ICD-R2	104 (A-H); 105 (A-H)
	ICD-R3	113 (1, 2, 3); 114 (1, 2, 3)

	ICD-R4	119 (a, b, c); 120 (a1, a2, b)
	ICD-R5	119 (a, b); 120 (a, b)
Diphtheria	ICD-R1-US	9
	ICD-R1-UK	15
	ICD-R2	9 (A, B, C)
	ICD-R3	10
	ICD-R4	10
	ICD-R5	10
	ICD-R6	055
Pertussis (whooping cough)	ICD-R1-US	8
	ICD-R1-UK	13
	ICD-R2	8
	ICD-R3	9
	ICD-R4	9
	ICD-R5	9
	ICD-R6	056 (0, 1)
Syphilis	ICD-R1-US	36
	ICD-R1-UK	29
	ICD-R2	37
	ICD-R3	38
	ICD-R4	34 (a, b, c)
	ICD-R5	30 (a, b, c, d, da, db, dc, dd)
	ICD-R6	020-029
Measles	ICD-R1-US	6
	ICD-R1-UK	6
	ICD-R2	6
	ICD-R3	7
	ICD-R4	7

Malaria	ICD-R5	35
	ICD-R6	085 (0, 1)
	ICD-R1-US	4
	ICD-R1-UK	25
	ICD-R2	4
	ICD-R3	5
	ICD-R4	38
	ICD-R5	28
	ICD-R6	110-117
Poliomyelitis	ICD-R1-US	63
	ICD-R1-UK	NA
	ICD-R2	63 (B)
	ICD-R3	22 (1, 2)
	ICD-R4	16 (1, 2)
	ICD-R5	36 (1, 2)
	ICD-R6	080, 081
Meningitis	ICD-R1-US	61
	ICD-R1-UK	83
	ICD-R2	61 (A, B, C)
	ICD-R3	24; 71
	ICD-R4	18; 79
	ICD-R5	6; 81
	ICD-R6	057; 340
Tetanus	ICD-R1-US	72
	ICD-R1-UK	24
	ICD-R2	24
	ICD-R3	29
	ICD-R4	22
	ICD-R5	12

	ICD-R6	061
	ICD-R1-US	14
	ICD-R1-UK	23
	ICD-R2	14
Dysentery	ICD-R3	16 (a, b, c)
	ICD-R4	13 (a, b, c)
	ICD-R5	27 (a, b, c, d)
	ICD-R6	045-048
	ICD-R1-US	1
	ICD-R1-UK	18
	ICD-R2	1
Typhoid and paratyphoid fever	ICD-R3	1a, 1b
	ICD-R4	1, 2
	ICD-R5	1, 2
	ICD-R6	040, 041
	ICD-R1-US	7
	ICD-R1-UK	8
	ICD-R2	7
Scarlet fever	ICD-R3	8
	ICD-R4	8
	ICD-R5	8
	ICD-R6	050
	ICD-R1-US	18
	ICD-R1-UK	40
	ICD-R2	18
Erysipelas	ICD-R3	21
	ICD-R4	15
	ICD-R5	11

Table S4 Wordlists of pathogen-related words in English, Italian, and Chinese.

Language	Words
English	pneumonia, influenza (grippe), tuberculosis (phthisis), enteritis, diarrhea, diphtheria, pertussis, syphilis (lues), measles (rubeola), malaria (ague), poliomyelitis, meningitis, tetanus, dysentery, typhoid, paratyphoid, scarlatina, erysipelas, AIDs;
Italian	polmonite, polmoniti, influenza, tubercolosi, diarrea, enterite, difterite, pertosse, sifilide, morbillo, malaria, poliomielite, meningite, tetano, dissenteria, febbre, paratifo, scarlattina, erisipela, AIDs;
Chinese	麻疹, 风疹, 脑膜, 炎, 乙脑, 小儿麻痹, 脊髓, 痢疾, 腹泻, 百日咳, 白喉, 疟疾, 伤寒, 副伤寒, 斑疹, 猩红热, 出血热, 肝炎, 瘟疫, 霍乱, 狂犬病, 钩虫, 黑热病, 炭疽, 艾滋病;

Table S5 Word examples of neurocognitive and sociocultural dimensions (in English).

Systems	Dimensions	Sources	Word examples	Sample sizes (N)
Neurocognitive	Sensory-motor	LIWC-percept	look, hear, feel	436
		LIWC-body	hand, leg, foot	215
	Time	LIWC-time	end, until, season	310
	Space	LIWC-space	down, in, thin	360
	Mental Cognition	LIWC-cogproc	cause, know, think	797
	Emotion	LIWC-affective	happy, cried, fearful	1393
	Social	LIWC-social	mate, talk, they	756
Sociocultural	Tightness	Jackson, et al. ⁴⁴	restrain, prevent, comply/allow, autonomy, openness	40
	Collectivism	Grossmann and Varnum ⁷	belong, duty, give/able, achieve, differ	14
	Masculinity	Lewis and Lupyan ⁶⁹	man, he, son/woman, she, daughter	14
	Traditional Values	Inglehart ^{38*}	religion, god, family/information, science, atheistic	32
	Survival Values	Inglehart ^{38*}	safe, food, money/humanity, protest, election	20
	Short-term Orients	Strathman, et al. ^{40*}	soon, present, now/future, century, investment	20

* indicates adaptations from surveys, rather than wordlists established by the previous study were directly used. Two-dimensional social-development-stages-related World Value Survey questions ³⁸ and Consideration of Future Consequences Scales (CFC) ⁴⁰ were used to generate keyword concepts, which were further extended to formalize final wordlists.

Table S6 Wordlists of social-cultural dimensional words in English, Italian, and Chinese

Dimension	English	Italian	Chinese
tight	restrain	trattenere	抑制
	prevent	prevenire	预防
	comply	ottemperare	遵守
	constrain	costringere	约束
	uniformity	uniformità	一致
	adhere	aderire	遵循
	enforce	imporre	执行
	proscribe	proscrivere	禁止
	abide	rispettare	遵从
	dictate	dettare	指示
	circumscribe	limitare	限制
	impose	imporre	强加
	uphold	difendere	维护
	discourage	scoraggiare	打消
	compel	costringere	迫使
	forbid	proibire	不准
	confine	confinare	限定
	govern	governare	统治
	prohibit	vietare	阻止
	preclude	precludere	排除
loose	allow	permettere	允许
	freedom	libertà	自由
	create	creare	创造
	variability	variabilità	变化
	autonomy	autonomia	自主
	openness	franchezza	开放
	leeway	margin	余地
	flexibility	flessibilità	灵活
	broadmindedness	aperto	豁达
	transformatory	trasformazione	改革
	customize	personalizzare	定制
	modify	modificare	修改
	subjectivities	soggettività	主体
	limitless	illimitato	无限制
	empower	potenziare	赋权
	adaptiveness	adattabilità	适应
	pluralism	pluralismo	多元化
	personalize	personalizzare	私人
	encourage	incoraggiare	鼓励

	diverse	diverso	多样
collective	belong	appartenere	属于
	duty	dovere	职责
	give	dare	给予
	harmony	armonia	和谐
	obey	ubbidire	服从
	share	condividere	分享
	together	insieme	共同
individual	able	capace	能够
	achieve	raggiungere	实现
	differ	differire	差异
	own	proprio	拥有
	personal	individuale	个体
	prefer	preferire	喜欢
	special	speciale	特别
male	man	maschio	男人
	boy	ragazzo	男孩
	brother	fratello	兄弟
	he	egli	他
	him	lui	他
	his	il suo	他的
	son	figlio	儿子
female	woman	donna	女人
	girl	ragazza	女孩
	sister	sorella	姐妹
	she	lei	她
	her	lei	她
	hers	il suo	她的
	daughter	figlia	女儿
tradition	god	dio	神
	religion	religione	宗教
	pray	pregare	祈祷
	worship	culto	崇拜
	heaven	paradiso	天堂
	parent	genitore	家长
	hell	inferno	地狱
	church	chiesa	教堂
	work	lavoro	工作
	spirit	spirito	灵魂
	respect	rispetto	尊敬
	family	famiglia	家庭
	army	esercito	军队
	tradition	tradizione	传统

	household	domestico	家务
	euthanasia	eutanasia	安乐死
	divorce	divorzio	离婚
	politics	politica	政治
	physical	fisico	客体
	materiality	materialità	物质
	atheistic	ateo	无神论
	concrete	calcestruzzo	具象
	objective	obiettivo	客观
secularity	civilization	civiltà	文明
	substantiality	sostanzialità	实体
	science	scienza	科学
	technique	tecnica	技术
	education	istruzione	教育
	information	informazione	信息
	system	sistema	系统
	discipline	disciplina	学科
	realism	realismo	现实
	safe	sicuro	安全
	food	cibo	食物
	finance	finanza	财务
	money	denaro	金钱
survive	security	sicurezza	保障
	poor	povero	贫穷
	income	reddito	收入
	ownership	proprietà	所有权
	government	governo	政府
	possession	possesso	财产
	democracy	democrazia	民主
	petition	petizione	请愿
	protest	protesta	抗议
	leisure	svago	闲暇
self-expression	humanity	umanità	人道
	movement	movimento	运动
	election	elezione	选举
	appeal	appello	诉求
	boycott	boicottare	抵制
	accomplishment	realizzazione	成就
	soon	presto	很快
short-term	instant	istantaneo	立刻
	immediate	immediato	立即
	moment	momento	目前
	shortly	presto	不久

	quickly	rapidamente	迅速
	suddenly	improvvisamente	突然
	now	adesso	当下
	present	attuale	现在
	improvident	imprevidente	短视
	future	futuro	未来
	predict	predire	预测
	consequence	conseguenza	后果
	anticipate	prevedere	预料
long-term	visionary	visionario	远见
	investment	investimento	投资
	long	lungo	长期
	impact	impatto	影响
	century	secolo	世纪
	plan	progetto	计划

Table S7 F statistic for Granger causality tests in lags of 1-year and 5-year

Dimensions	US	UK	Italy	China
F statistic for Granger causality tests in lags of 1 year				
Sensory-motor	F(97, 98) = 4.81*	F(96, 97) = 7.26**	F(110, 111) = 8.60**	F(37, 38) = 3.65†
Time	F(97, 98) = 0.11	F(96, 97) = 5.36*	F(110, 111) = 1.31	F(37, 38) = 7.60**
Space	F(97, 98) = 2.96†	F(96, 97) = 8.59**	F(110, 111) = 6.60*	F(37, 38) = 1.73
Mental cognition	F(97, 98) = 7.71**	F(96, 97) = 1.60	F(110, 111) = 7.35**	F(37, 38) = 4.55*
Emotion	F(97, 98) = 12.74***	F(96, 97) = 0.99	F(110, 111) = 45.15***	F(37, 38) = 3.93†
Social	F(97, 98) = 0.83	F(96, 97) = 7.53**	F(110, 111) = 9.02**	F(37, 38) = 5.05*
Tightness	F(97, 98) = 8.24**	F(96, 97) = 15.45***	F(110, 111) = 54.16***	F(37, 38) = 2.62
Collectivism	F(97, 98) = 6.24*	F(96, 97) = 13.08***	F(110, 111) = 5.82*	F(37, 38) = 0.13
Traditional	F(97, 98) = 2.16	F(96, 97) = 7.69**	F(110, 111) = 19.68***	F(37, 38) = 6.50*
Survival	F(97, 98) = 9.16**	F(96, 97) = 6.55*	F(110, 111) = 0.33	F(37, 38) = 0.07
Masculinity	F(97, 98) = 4.94*	F(96, 97) = 3.52†	F(110, 111) = 15.00***	F(37, 38) = 3.22†
Short-term	F(97, 98) = 3.81†	F(96, 97) = 14.96***	F(110, 111) = 14.53***	F(37, 38) = 3.22†
F statistic for Granger causality tests in lags of 5 year				
Sensory-motor	F(85,90) = 3.04*	F(84, 89) = 3.06*	F(98, 103) = 3.05*	F(25, 30) = 2.38*
Time	F(85,90) = 2.42*	F(84, 89) = 2.08†	F(98, 103) = 1.88	F(25, 30) = 2.21†
Space	F(85,90) = 2.98*	F(84, 89) = 1.67	F(98, 103) = 0.62	F(25, 30) = 3.80*
Mental cognition	F(85,90) = 1.55	F(84, 89) = 0.81	F(98, 103) = 2.08†	F(25, 30) = 4.10**
Emotion	F(85,90) = 3.35**	F(84, 89) = 1.91	F(98, 103) = 5.02***	F(25, 30) = 3.99**
Social	F(85,90) = 0.83	F(84, 89) = 3.81**	F(98, 103) = 2.85*	F(25, 30) = 3.46*
Tightness	F(85,90) = 3.52**	F(84, 89) = 3.24*	F(98, 103) = 7.16***	F(25, 30) = 2.85*
Collectivism	F(85,90) = 2.51*	F(84, 89) = 1.70	F(98, 103) = 0.87	F(25, 30) = 0.27
Traditional	F(85,90) = 3.84**	F(84, 89) = 2.67*	F(98, 103) = 2.50*	F(25, 30) = 1.87
Survival	F(85,90) = 2.04†	F(84, 89) = 0.52	F(98, 103) = 0.33	F(25, 30) = 2.24†
Masculinity	F(85,90) = 2.38*	F(84, 89) = 1.83	F(98, 103) = 1.39	F(25, 30) = 4.05**
Short-term	F(85,90) = 2.53*	F(84, 89) = 1.66	F(98, 103) = 3.08*	F(25, 30) = 1.72

†, P < 0.10; *, P < 0.05; **, P < 0.01; ***, P < 0.001